TECHNISCHE UNIVERSITÄT WIEN

D I S S E R T A T I O N

Empirical assessment of minimum dimensions
for cartographic symbology on smartphone displays

Ausgeführt zum Zwecke der Erlangung des akademischen Grades
eines Doktors der technischen Wissenschaften unter der Leitung von

Univ.Prof. Mag.rer.nat. Dr.rer.nat. Georg Gartner

Forschungsbereich Kartographie, E 120-6

Department für Geodäsie und Geoinformation

Eingereicht an der Technischen Universität Wien

Fakultät für Mathematik und Geoinformation

von

Dipl. Ing. Florian Ledermann

Mat. Nr. 09426416

Wien, am 20. 10. 2022

Empirical assessment of minimum dimensions
for cartographic symbology on smartphone displays

Florian Ledermann

Abstract

Map design guidelines proposed in the cartographic literature often include advice on minimum dimensions for cartographic symbology. For paper maps, such guidelines were based on the properties of the printed medium facilitating the reproduction of fine details, as well as assessments of the capability of map users to recognize and discriminate the graphical elements of the map. Due to the comparably low resolution of conventional computer screens, published advice on designing maps for digital presentation usually mandates the use of much larger symbology, in order to ensure reliable reproduction and legibility of map symbology.

In recent years, displays with greatly increased pixel densities have become available for computers and mobile devices. The goal of this thesis is to empirically verify the validity of existing advice regarding the minimum dimensions of cartographic symbology for maps targeting modern, high-resolution smartphone screens, and develop updated guidelines reflecting the current state of technology. For this purpose, three user studies are conducted, in which the legibility of graphical elements related to cartographic symbology is tested on different mobile phone displays, representing the range of display resolutions that is currently available in the smartphone market.

Study 1 aims to assess the perceptual thresholds of participants for discriminating graphical elements related to a wide range of cartographic symbology (point icons, line features, and text labels), at decreasing size levels. The results and limitations of this initial study inform the methodology adopted for two subsequent studies. Study 2 investigates in detail the legibility of icons potentially used as point symbols on maps. For this purpose, a method to analyse sets of real-world cartographic icons is proposed and implemented, that facilitates the identification of subsets of "most similar" icons from such map icon collections. The identified subsets are then used to empirically assess the minimum size at which icons can be reliably discriminated, for four icons sets and two types of tasks – discriminating between the icons of the set when presented in isolation, and counting icons among similar icons on a map. Study 3 investigates the legibility of cartographic line symbology in more detail, in particular the capability of modern smartphone displays to accurately reproduce line symbols containing internally differentiated geometry to potentially communicate additional information, such as directional arrows or partial hachures. As in the second study, the legibility of line symbology is tested for two types of tasks: discriminating between line symbols presented in isolation, and counting lines of various kinds on a pseudo-map.

Studies 2 and 3 also test some initial ideas for exploiting the pixel structure of modern smartphone displays to optimize the legibility of map symbols. For Study 2, the effect of optimizing icon designs for ideal alignment with the pixel grid is tested, as well as an algorithmic approach to "shape difference amplification" of small map icons. For Study 3, the effect of placing highlights on arrows embedded in a line is tested, with the hypothesis that this may improve legibility at small sizes.

One objective of the work undertaken for this thesis is to propose practical advice for creators of maps, informed by the empirical assessment done in the three studies. Such advice is presented on three levels: General advice for map designers on the design and deployment of maps presented on screens; detailed recommendations for minimum dimensions of cartographic symbology, differentiated by display pixel density and visual acuity of the map user; and advice for practitioners who want to empirically verify the legibility of particular cartographic designs themselves.

The thesis concludes with a proposed definition of the term "minimum dimension" in the context of cartographic design, the discussions of practical implications and application ideas, a reflection on the limitations of the methodology and implementation of the presented studies, and an outlook on potential future research opportunities that may be informed or motivated by this thesis.

Kurzfassung

Etablierte Richtlinien zur Kartengestaltung beinhalten oftmals Empfehlungen für Minimaldimensionen kartographischer Symbole. Empfehlungen für gedruckte Karten berücksichtigten dabei die Eigenschaften des Herstellungsverfahrens, sowie die Fähigkeiten der Kartennutzerin, kleinste Details zu erkennen und korrekt zu deuten. Aufgrund der vergleichsweise niedrigen Auflösung konventioneller Bildschirme verlangen etablierte Richtlinien für die Gestaltung von Karten für die digitale Wiedergabe üblicherweise die Verwendung weit größerer grafischer Elemente, um die Lesbarkeit der Karte zu gewährleisten.

Die technische Entwicklung der vergangenen Jahre ermöglicht mittlerweile jedoch die Herstellung digitaler Anzeigegeräte mit sehr hoher Auflösung, und solche hochauflösenden Displays sind, sowohl als Smartphones als auch als Desktop-Bildschirme, heute weithin verfügbar. Ziel der vorliegenden Arbeit ist, basierend auf empirischen Untersuchungen, aktualisierte Richtlinien für kartographische Minimaldimensionen zu erarbeiten, die die Verfügbarkeit von hochauflösenden Displays berücksichtigen. Zu diesem Zweck werden drei empirische Studien vorgestellt, die die Lesbarkeit von graphischen Elementen auf Smartphone-Displays unterschiedlicher Auflösung untersuchen.

Studie 1 ermittelt die Erkennbarkeitsschwellen von graphischen Elementen mit Bezug zu kartographischen Signaturen (Punktsymbole, Liniensymbole, sowie Schrift) bei kontinuierlich reduzierter Größe. Die Erkenntnisse aus dieser ersten Studie fließen in die Gestaltung zweier weiterer Studien ein. Studie 2 untersucht weitere Aspekte betreffend die Erkennbarkeit von Punktsymbolen. Zu diesem Zweck wird eine Methode vorgestellt, um aus verfügbaren Sammlungen von ikonographischen Symbolen jene Symbole mit größter Ähnlichkeit zu identifizieren. Solche Gruppen ähnlicher Symbole finden als Stimuli in der zweiten Studie Verwendung, um die Mindestgröße für die zuverlässige Unterscheidung von aus realen Anwendungskontexten entnommenen, graphisch ähnlichen Kartensymbolen zu ermitteln. Die Unterscheidbarkeit wird dabei in zwei Arten von Aufgaben getestet: bei isoliert stehender Betrachtung, sowie beim Abzählen von Symbolen auf einer Karte. Studie 3 widmet sich der Erkennbarkeit von Liniensymbolen, wobei hier die zuverlässige Reproduktion von intern ausdifferenzierten Liniensymbolen, beispielsweise durch die Anordnung von Pfeilen oder Schraffuren im inneren der Linie, getestet wird. Auch hier wird die Unterscheidbarkeit bei isolierter Betrachtung, sowie für das Abzählen auf einer „Pseudo-Karte", untersucht.

Zusätzlich zur Untersuchung etablierter kartographischer Symbole werden in Studien 2 und 3 auch mögliche Verfahren zur aktiven Verbesserung der Unterscheidbarkeit getestet. Für Studie 2 wird die Verbesserung der Lesbarkeit durch Optimierung der Ausrichtung der Symbolgeometrie mit dem Pixelraster untersucht, sowie eine algorithmische Methode zur möglichen Verstärkung der graphischen Unterschiede ähnlicher Symbole vorgestellt und getestet. In Studie 3 wird der Effekt der graphischen Hervorhebung der Eckpunkte von in Linien eingebetteten Richtungspfeilen untersucht.

Ein Ziel dieser Arbeit ist die Ausarbeitung praktisch anwendbarer Richtlinien für Minimaldimensionen für digitale Karten. Ein Vorschlag für solche Leitlinien erfolgt auf drei Ebenen, basierend auf den vorgestellten Studien: Generelle Empfehlungen für die Gestaltung von Karten für digitale Ausgabegeräte auf aktuellem technischen Stand; Empfehlungen für Minimaldimensionen kartographischer Elemente, differenziert nach Display-Auflösung und Sehschärfe der Nutzerin; sowie praktische Hinweise für das empirische Testen der Lesbarkeit digitaler Karten.

Die Arbeit schließt mit dem Versuch einer Definition des Begriffs „Kartographische Minimaldimension", der Erörterung von praktischen Anwendungsmöglichkeiten der gewonnenen Erkenntnisse, einer Diskussion von Beschränkungen der vorgestellten Methodik und Umsetzung, sowie einem Ausblick auf mögliche weiterführende Forschung.

Aspects of this research have been presented in the following publications:

F. Ledermann, G. Gartner (2021): *Towards Conducting Reproducible Distributed Experiments in the Geosciences*. In: Open-access proceedings of the Association of Geographic Information Laboratories in Europe (AGILE 2021). AGILE: GIScience Series, 2 (2021), p. 33–39. https://doi.org/10.5194/agile-giss-2-33-2021

F. Ledermann (2021): *Small differences: Limits of Legibility of Cartographic Symbols on High- and Ultra-High-Resolution Mobile Displays*. In LBS 2021: Proceedings of the 16th International Conference on Location Based Services, p. 177-182. https://doi.org/10.34726/1785

F. Ledermann (2021): *Investigating the Effect of Display Pixel Density on the Minimal Legible Size of Cartographic Symbols*. In: Abstracts of the ICA, Vol. 3 (ICC 2021: Proceedings of the 30th International Cartographic Conference, Firenze, Italy), p. 174. https://doi.org/10.5194/ica-abs-3-174-2021

F. Ledermann, G. Gartner (2021): *Towards Reproducible User Studies and Behavioral Experiments for Cartography With the stimsrv Framework*. In: Abstracts of the ICA, Vol. 3 (ICC 2021: Proceedings of the 30th International Cartographic Conference, Firenze, Italy), p. 175. https://doi.org/10.5194/ica-abs-3-175-2021

F. Ledermann (2022): *Minimum Dimensions for Cartographic Point Symbols on Mobile Phone Screens: Theoretical Considerations and Empirical Verification*. In: Abstracts of the ICA, Vol. 5 (EuroCarto 2022: Proceedings of the European Cartographic Conference, Vienna, Austria) p. 110. https://doi.org/10.5194/ica-abs-5-110-2022

F. Ledermann (2022): *The Effect of Display Pixel Density on the Minimum Legible Size of Fundamental Cartographic Symbols*. The Cartographic Journal (in press). https://doi.org/10.1080/00087041.2022.2055938

F. Ledermann (forthcoming): *Minimum Dimensions for Cartographic Symbology – History, Rationale and Relevance in the Digital Age*. (Accepted for publication in the International Journal of Cartography, forthcoming special focus issue on the State of the Art & Science in Cartography)

Editorial note:

This thesis uses British spelling throughout the text. Program code fragments and citations from other works are reproduced using their original spelling.

Acknowledgements

Table of Contents

List of Abbreviations

AFC	Alternative-Forced-Choice
AIGA	American Institute of Graphic Arts
API	Application Programming Interface
CC	Creative Commons
CPU	Central Processing Unit
CRT	Cathode Ray Tube
CSS	Cascading Style Sheets
CSV	Comma Separated Values
DOM	Document Object Model
DPI	Dots Per Inch
FFP	Filtering Face Piece
FFS	Fringe Field Switching
FrACT	Freiburg Visual Acuity Test
GIMP	GNU Image Manipulation Program
GIS	Geographic Information System
GNU	GNU is not Unix
GPL	General Public License
GPS	Global Positioning System
GPU	Graphical Processing Unit
HDMI	High Density Media Interface
HTML	HyperText Markup Language
HTTP	HyperText Transfer Protocol
ICA	International Cartographic Association
ICC	International Cartographic Conference
ISO	International Standard Organisation
JSON	JavaScript Object Notation
LAN	Local Area Network
LCD	Liquid Crystal Display
LGN	Lateral Geniculate Nucleus
LED	Light Emitting Diode
logMAR	Logarithm (base 10) of the Minimum Angle of Resolution (MAR)
MAR	Minimum Angle of Resolution
n-AFC	n-Alternatives Forced-Choice
NPS	National Park Service
OCHA	United Nations Office for the Coordination of Humanitarian Affairs
OLED	Organic Light-Emitting Diode
OSM	OpenStreetMap
PDF	Portable Document Format
ppi	Pixels Per Inch
PVC	Polyvinyl Chloride
RAM	Random Access Memory

RGB	Red / Green / Blue
RO	Research Objective
RQ	Research Question
SD	Standard Deviation
SLR	Single-Lens Reflex (camera)
SVG	Scalable Vector Graphics
TAO	The Auckland Optotypes
TFT	Thin-Film Transistor
TV	Television
USB	Universal Serial Bus
VESA	Video Electronics Standards Association
XGA	Extended Graphics Array

1. Introduction and motivation

Cartography is a discipline concerned with depicting a space of virtually unlimited information density – the real world. The challenges of managing, transforming and selecting information from the range of available data and conveying it to the map user in an understandable and attractive form are as old as the discipline itself. At every step of the cartographic pipeline, cartographers have to negotiate the conflicting demands of wanting to convey as much information as possible to the map user, while at the same time needing to keep the amount of information manageable on a technical, perceptual and semantic level.

When the real world is depicted on a cartographic medium, simply scaling down all features in view, or selecting only a subset of features and omitting all others, will not lead to desirable results. One of the main goals of cartographic generalisation is thus to convey as much relevant information as possible, despite the fact that the chosen map scale would prevent some features from being represented in a legible way. Spiess (1990b) articulates this desire to "delay the loss of information" when proceeding to a smaller map scale:

> Trotz des Übergangs auf stets kleinere Kartenmaßstäbe und damit auf immer weniger Fläche soll **mit Hilfe der Generalisierung der Informationsverlust hinausgezögert** werden. (Spiess, 1990b, p. 50, emphasis original)

In their desire to delay the loss of information, cartographers are confronted with several obstacles, even if the information is available in the required detail and format. Firstly, any particular output medium has inherent limitations regarding the density of information that can be accurately reproduced. Generalisation techniques developed for printed maps took into account not only the limits of resolution of the reproduction processes involved, but also the behaviour of ink flowing on the page in the printing process and other factors. Minimal distances between map features were adjusted accordingly, to allow for a clear representation in the face of these media-specific limitations (Schweizerische Gesellschaft für Kartografie, 1980, p. 30).

Secondly, even if a hypothetical "perfect" reproduction medium that allowed for infinite detail would be available, map users would still be limited by their perceptual and cognitive abilities in the density of information that can be reliably picked up from the mediated representation, which demands a certain minimum size of visual marks for them to be perceived and distinguished accurately.

Thirdly, map designers need the conceptual and practical tools to create maps suitable maps for conveying all information considered relevant. If the limitations of the medium or the map user are not known, or suitable techniques for generalisation of geographic features or implementation of an envisioned map design are not available, the cartographer will be limited in their ability to produce a legible and easy to use map.

Those foundational challenges of cartography – wanting to convey as much relevant information and detail as possible to delay the loss of information, in the face of limitations

of the cartographic medium, the cognitive abilities of the map user, and the mapmaking process – are as relevant as ever in the digital age. However, the appeal of computers has largely been seen to lie in other areas, not mainly in the accurate reproduction of static maps on a screen. Screens are still considered inferior to paper as an output medium in terms of pure visual quality (Lobben & Patton, 2003; Jenny et al., 2008; Muehlenhaus, 2014; Buchroithner, 2016), and the advantages of the digital medium for the map producer and user are usually seen in other areas than their graphical fidelity. In particular, the possibility to interact with the cartographic display for zooming in on or interactively selecting aspects of interest (Lobben & Patton, 2003) has been seen as potentially compensating some of the shortcomings regarding the accuracy of the display. Detailed guidelines for map design for screens have been developed, taking into account the lower spatial density of the output medium (Malić, 1998; Neudeck, 2001; Schweizerische Gesellschaft für Kartografie, 2002), and are widely considered canonical knowledge for cartographers. For many map usage scenarios, the loss of graphical detail of a given static map view was considered a price well worth paying for the gained benefits of interactive zooming and selection, animated map displays and, last but not least, worldwide instant publishing of maps in the age of the internet.

However, developments of recent years require us to fundamentally challenge the assumption that screens are inferior to paper-based maps in terms of visual fidelity and contrast, and that, in addition to the advantages mentioned above, digital displays may today have surpassed high-quality printed reproduction, even regarding the properties that have so far been seen as major strengths of the printed medium. This may in turn open up new possibilities for solving the dilemma of information loss, which is amplified on a low-resolution display. The introduction of touchscreen-based smartphones as affordable consumer devices, often attributed in origin to the introduction of the Apple iPhone in 2007 (although phones with visual interfaces were available earlier), laid not only the technological but, equally important, also the *economic* basis for heavy investments in the development of display technology in the last one and a half decades. Due to the typically shorter viewing distance, smartphones require a higher pixel density than conventional desktop computer displays to produce an image of equal graphical fidelity. As a consequence, resolutions of smartphone displays available to consumers increased dramatically over the past years, driven by the demand for greater graphical clarity, setting in motion technological developments that also benefitted desktop monitors (Chen et al., 2018). While until the early 2000's, a computer monitor could be conveniently assumed to have a pixel density in the range of 80-100 pixels per inch (Malić, 1998), as of 2020, the pixel density of recently released smartphones ranges from 268 to 808 pixels per inch, while desktop and laptop monitors can be acquired today with pixel densities of approximately 100 to 350 pixels per inch. The historical evolution of pixel densities of smartphone screens is shown in Figure 1.1.

The dramatic increase in pixel densities of digital displays in our pockets and, potentially, on our desks means that well-established design guidelines for digital maps, based on assumptions of display resolutions typical at the turn of the millennium may need to be reconsidered. High- and ultra-high resolutions are now widely available to users, and not only in settings where the display hardware is under control of the map creator (e.g. public

displays, museum exhibits etc.). It is currently not known whether, as a consequence of the digital medium actually approximating or, in some cases, potentially surpassing the printed medium in visual fidelity, we can simply go back to cartographic guidelines established for paper-based maps, or if those guidelines need to be adapted and refined to reflect the quality of visual reproduction that can be reliably accomplished on modern digital displays.

Pixel Densities of Mobile Phone Screens 1996−2020
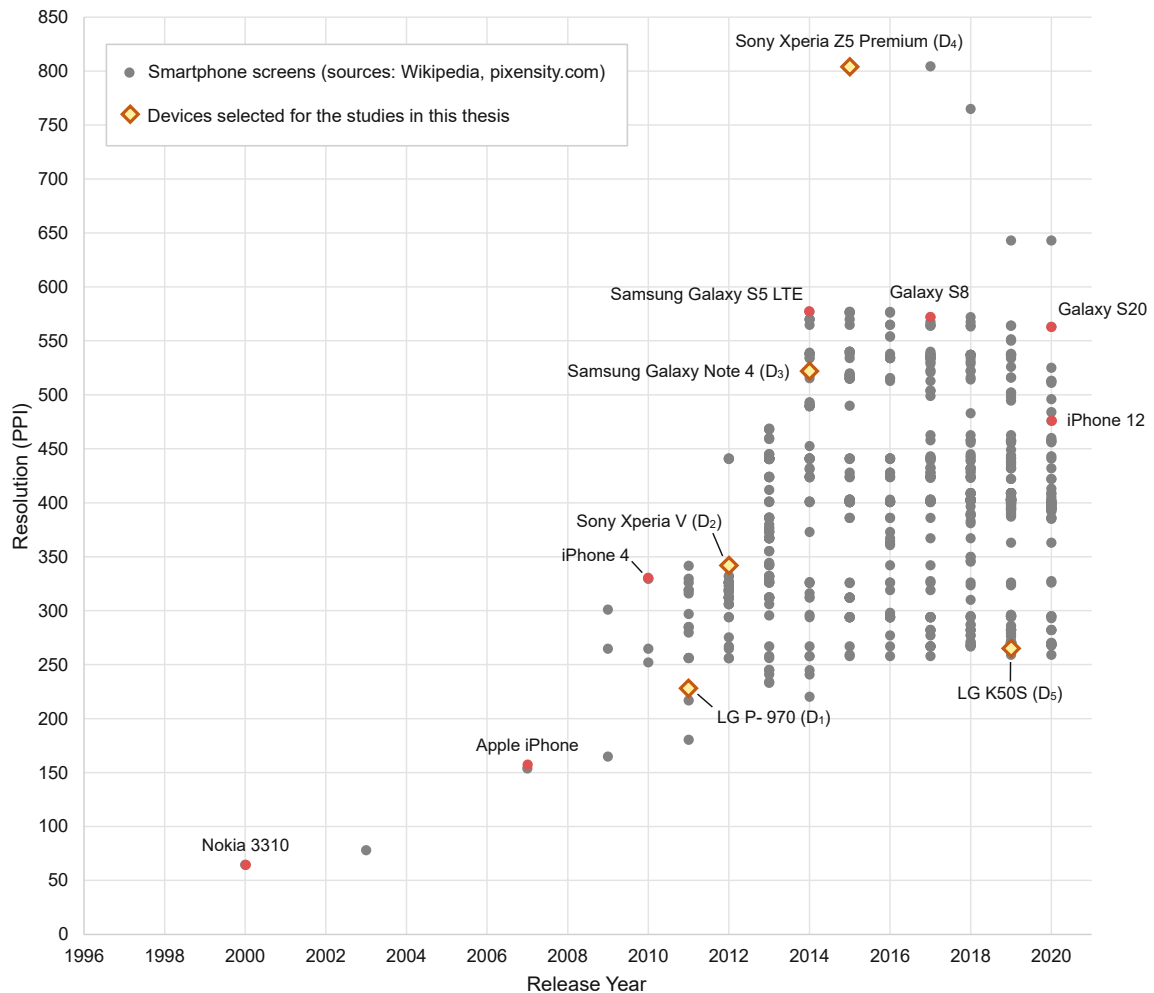
Figure 1.1 Evolution of mobile phone screen resolution, 1996-2020. The phone models numbered D_1–D_5 were selected for use in the studies presented in this thesis. Sources: pixensity.com, accessed 2021-02-03; https://en.wikipedia.org/wiki/Comparison_of_high-definition_smartphone_displays, accessed 2021-02-03.

While significantly higher pixel densities may be the most noticeable improvement of display quality in recent decades, it is not the only area in which significant improvements to display quality have been made. In recent years, OLED technology has made it possible to produce displays with unique properties compared to conventional LCD displays. Because in OLED displays each pixel acts as a controllable emissive light source (in contrast to LCDs, which subtract light from a permanently active background light), it is possible for the screen to display "true black" by switching off the corresponding pixels, therefore not emitting any light. The theoretical contrast ratio (the ratio of lowest to highest intensity of light emitted) of such a display is infinite, because the lowest light level is truly zero. While such infinite contrast ratio cannot be accomplished in reality, due to ambient light reflected from the

display surface, contrast ratios of modern displays are generally much improved in indoor settings (Chen et al., 2018). Furthermore, improvements in display manufacturing processes have made novel subpixel arrangements, deviating from the conventional "RGB" subpixel arrangements known from desktop LCDs, possible. The significance of these developments for cartography has not yet been thoroughly investigated.

In the face of continuous and rapid technological development, a question that needs to be asked before embarking on a scientific project is whether it is timely to start an investigation at the present moment in order to derive somewhat stable knowledge, or whether such an endeavour would only produce a "snapshot" of the present state of technology, quickly made obsolete by further developments. Although the author of this thesis is not in possession of a proverbial "crystal ball" to foresee future developments, a few arguments can be made to support the claim that now is indeed the appropriate time to attempt an update of cartographic knowledge and best practices with respect to screen-based output: Firstly, comprehensive investigations of cartographic guidelines for screen-based output (Malić, 1998; Neudeck, 2001) are now two decades old, and, as has been shown, took place before the significant technological developments of the recent 15 years. If anything, it can be argued that a thorough re-evaluation of those findings has been overdue for a while. Secondly, although a clear upwards trend in pixel density can be observed over the last 15 years for both for mobile phone screens and desktop monitors, some indications for a "plateau" or even a recent decline in maximum pixel densities in both domains can be observed: the two outliers in the realm of mobile phones, the Sony Xperia Z5 Premium and its predecessor the Xperia XZ Premium, both surpassing a pixel density of 800 pixels per inch, did not have their full physical resolution enabled by default by the manufacturer, indicating perhaps some doubt regarding the question of whether such extreme resolutions could still be made productive by normal applications. The outlier in the realm of desktop monitors, the Dell UltraSharp UP3218K, costs over 4.200€ as of July 2022[1], despite having been released over 4 years earlier, and has not been matched in display resolution by products from competing manufacturers. For both types of devices, no new consumer devices surpassing the pixel densities of those mentioned above have been announced by any manufacturer, which could potentially indicate the end of a phase of continuously increasing pixel densities in which increased display resolution has been used as a major selling point for new devices, as manufacturers find it harder to find a compelling "story" to sell those devices to end users.

The third argument that can be made in favour of investigating the topic at the present moment is that devices of a wide range of pixel densities and other properties are available and in production right now, making the comparison between a wide range of different devices feasible. It is important to note that the pixel density of devices currently on the market is not in fact limited by technological factors any more. Indeed, current technology allows the production of displays with pixel densities of thousands of pixels per inch – but currently the main driver for the development and production of such displays with extremely high resolutions is their use in headsets for immersive virtual reality applications, which requires smaller absolute display sizes and, therefore, higher pixel densities (Chen et

---

[1] Source: https://www.dell.com/de-de/shop/dell-ultrasharp-32-premiercolor-ultrahd-8k-monitor-up3218k/apd/210-amfd/monitore-und-monitorzubeh%C3%B6r , accessed 2022-07-22.

al., 2018). Technological progress is, in fact, still ongoing, but development now focuses at pixel densities beyond of what can be made useful for application in desktop monitors or smartphones, and on form factors that cannot be easily repurposed as displays for mobile phones.

Looking at the graph in Figure 1.1, a dense cluster of devices released in recent years within a certain range of pixel densities becomes apparent; most smartphones manufactured recently have a pixel density between 250–580 ppi. While a general tendency for pixel densities to go upwards can still be observed, it seems reasonable to expect devices with pixel densities within those boundaries to be available for the foreseeable future. For desktop monitors, this range can be expected to lie around 100-280ppi. While it is interesting to investigate the possibilities opened up for the deployment of maps in controlled settings in which a high-resolution display device can reliably made available, it will also be important to develop strategies for how to deal with the full range of resolutions used by the general population, and develop design guidelines for maps which are widely deployed and will therefore be potentially viewed on the full range of heterogeneous devices.

In this thesis, it is proposed to assess the potential impact of recent developments in display technology on the design of cartographic visualisations through the lens of a simple question: How small can graphical elements of a map be made, so that they can still be reliably read? Guidance on the smallest details that can be reliably reproduced on a map and read by the map user have historically been included in map design guidelines. At the same time, it is commonly recommended that maps should communicate the relevant information on them clearly and unambiguously. The use of graphical elements scaled down to the limits of the capabilities of map medium and user would certainly be in contradiction to such calls for clear and reliable communication. Consequently, some may ask why the issue of minimum dimensions should be investigated in the first place. Several reasons can be given for the relevance of minimum dimensions of cartographic symbology that do not contradict aforementioned demands:

- The smallest graphical elements will establish the lowest levels of the *visual hierarchy* of the map. While for simple thematic maps, a hierarchy of two levels can be sufficient, more complex maps may need to have five or more levels at which symbology and graphical elements can be clearly distinguished (Dent et al., 2008; Muehlenhaus, 2014). Dent (1972) demands that "objects should become less intense and their edges more fuzzy" (p. 84) with each level. Giving designers clear guidance on the smallest elements that can be reliably used to convey information will increase their options for establishing the visual hierarchy of the map, and will therefore potentially influence symbology at higher levels of importance, even if at those levels symbology of significantly larger size will be used.
- The smallest size of details that can be reliably reproduced by the map medium, and reliably be read by the map user, are an important *parameter for map generalisation* and other techniques of algorithmic symbolization. As has been stated above, if the resolution capabilities of map medium and user would be infinite, generalisation would not be necessary. Therefore, even theoretical considerations of generalisation techniques introduce properties of the map medium as a pragmatic aspect. Töpfer (1974) states that "the limits of scaled-down depiction are not only depending on map scale and size of an

object's footprint, but also on the means of visualization dedicated to the task"[2] (p. 20). McMaster and Shea (1992) conceptualize digital generalisation as an algorithmic problem, listing "graphical clarity" as one of its main objectives. Minimum dimensions are one way to formalize the concept of graphical clarity, and therefore would be a relevant parameter for the configuration of generalisation algorithms.

- Even though the information of main importance on a map should be sized significantly larger, minimum dimensions could be of relevance for the *display of non-essential additional information*. On digital maps, it is often expected that additional information can be made available on demand, upon user interaction, and thus the design of the map can be reduced to include only a few layers of primary importance. However, it may not always be obvious to the user that additional information is available in the first place, particularly when concerning geographic features that are not depicted on the map at all. The inclusion of additional layers on the map, providing contextual or auxiliary information which can then be retrieved in more detail upon interaction, may therefore be found helpful and have the potential to increase user interaction. Furthermore, information that is displayed repeatedly or for which the meaning can be inferred from its spatial context, reduced symbology sizes may be used.

- While using symbology close to the minimum size may prevent some users with reduced visual capabilities to accurately interpret the information, modern digital systems can offer interactive means or configuration options that *allow users to adapt the representation to their individual needs and preferences*. Thus, in contrast to printed maps which are fixed in their appearance, digital maps may not always need to incorporate a "safety margin" in their design to ensure reliable communication in adverse circumstances of for users with special needs.

Some have claimed that printed maps still possess a certain "aura" that is not present in currently produced digital cartographic artefacts (Buchroithner, 2016). Figure 1.2 illustrates what this could refer to, by juxtaposing a map produced 100 years ago by means of copper engraving with a contemporary digital map of the same geographical region. Despite the digital map containing quite a lot of information, it appears at the same time somewhat bland, but also hard to read. A perceived "aura" of printed maps may simply be related to the fact that they, particularly when intended for "general reference" or outdoor navigation purposes, had to pack as much information as possible in a given area, depicted by symbology of high contrast, since later changes or additional queries were not supported by the medium, and the symbology was limited to elements that could be created with manual procedures. Applications in novel and continuously developing fields, such as maps for multimodal transport (Fairbairn, 2005), for hiking and cycling (Wessel & Widener, 2015), or for navigating dense urban environments (Carlino, 2022), which are facilitated by new data and symbolization techniques becoming available, may require map users to be able to process multiple aspects of spatial information in an ad-hoc integrated fashion, and therefore require cartographers to again consider designs of higher graphical density.

---

[2] Own translation; German original: "daß die Grenzen der maßstäblichen Darstellung nicht nur vom Kartenmaßstab und der Größe des Objektgrundrisses, sondern auch von den vorgesehenen Darstellungsmitteln abhängig sind."
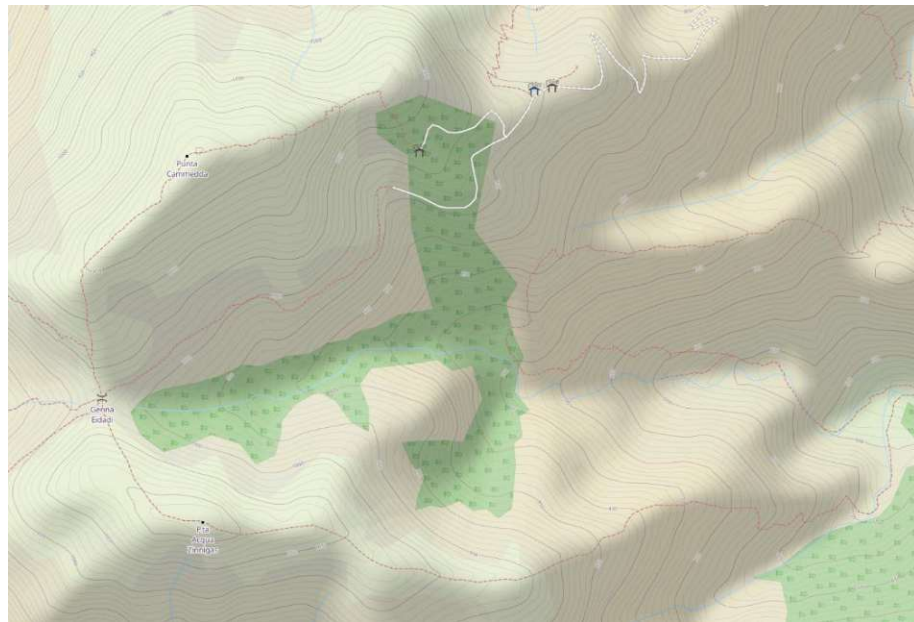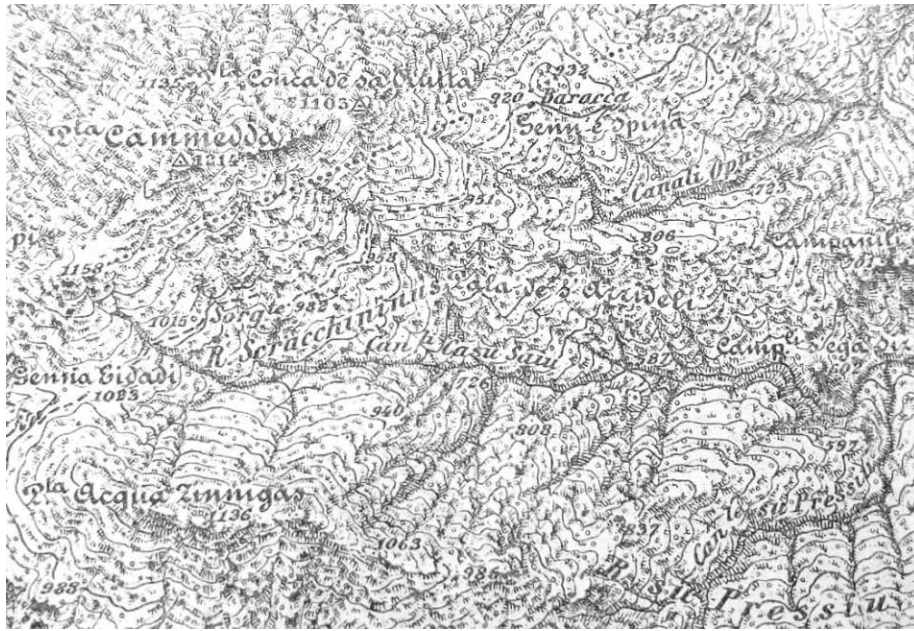
Figure 1.2: Top: Detail of a monochrome map produced in 1922 by Italian cartographer Olinto Marinelli, on display at ICC 2021 in Firenze, Italy. Bottom: Screenshot from OpenCycleMap³ from the same region, which provides the most detailed cartographic depiction of the area out of several tested online map providers.

At present, maps can not only be published instantly to a world-wide audience on the web, but have the potential to "go viral" by being reciprocally shared by online audiences (A. C. Robinson, 2019). Viral media is often geared towards provoking user reaction in order to entice users to re-share the content or assert its algorithmically determined relevance by way of "user engagement". Paradoxically, any requirement to engage with the content in deeper ways, for example by using complex interaction techniques, inspecting optional layers of information not visible on first glance, or critically assessing a map's sources or visualisation techniques, could be seen as detrimental to the success of viral content, which thrives on knee-jerk social reactions of sharing, re-sharing and remixing. Showing more details

³ https://www.opencyclemap.org/?zoom=16&lat=39.43618&lon=8.64971

represented in the static image of a map by adding layers of symbology with secondary information could potentially help to preserve more information even when shared as a screenshot or "remixed". The increased demand of audiences with short attention spans to process information "at a glance" should not be lamented by cartographers, but taken as motivation to use currently available technology to its full potential and develop visualisation techniques and cartographic designs that "can tell you a lot of things quickly" (Tjukanov, 2021) without requiring explicit interaction. Highly detailed cartographic designs, informed by a thorough understanding of the limitations of the medium and the map reading process, may be one strategy to pursue such a goal. Reliably establishing the minimum dimensions of graphical elements to be used in cartographic designs can be a first step towards realising such a strategy.

2. Problem statement, research objectives & overview

Chapter 1 gave an overview of some recent technological developments and ideas that motivate the proposed investigation. These can be briefly summarized as follows:

- Currently available smartphone displays potentially surpass printed paper in terms of visual fidelity and contrast ratio (at least in indoor settings).
- These technological developments make necessary a reconsideration of established cartographic guidelines regarding the design of maps for screen-based presentation, which were generally based on assumptions of relatively homogeneous resolutions below 100 pixels per inch. It is unclear whether, as a consequence, established guidelines for printed maps can simply come into effect unchanged for screen-based maps (as has been hypothesized by Neudeck (2001, p. 62) in his discussion anticipating the development of higher resolution screens), or whether some limitations, rooted in perceptual or technological aspects, remain that mandate to maintain the recommendation to use larger symbology sizes for presentation on digital devices. It may even be the case that modern smartphone displays are in some aspects superior to printed paper, and allow map designers to go beyond the limitations established for the latter medium.
- Even if it turns out that the guidelines established for printed maps can now be applied to digital maps in their original form, the technological developments that have taken place in the area of digital map production now facilitate a larger choice of cartographic symbology than what may have been previously considered. It may therefore be necessary to consider new types of symbology or novel strategies for optimizing symbology for screen-based presentation, and therefore to revise and extend established guidelines.
- The introduction of devices with ever higher pixel densities on the market has plateaued a few years ago, probably because a "killer app" – an application scenario in which those high resolutions could be put into practice and made a noticeable difference for the user – has not emerged for such displays. Could digital maps be such a "killer app" that can put the visual fidelity that is now technologically achievable to practical use?
- For some of the guidelines on minimum dimensions proposed by cartographers in the past, it is not entirely clear on what basis these recommendations have been made. It is therefore difficult to reproduce the empirical investigations or theoretical considerations that may have informed these guidelines. In order to facilitate reproducible research (Giraud & Lambert, 2017; Nüst & Pebesma, 2021), the goal should be to propose, and fully document, a set of symbol geometries, methods of presentation, procedures for analysis and guidelines derived from the empirical work, for other researchers to reproduce, extend, and critically reflect upon in the future. This includes making transparent any limitations, errors and suboptimal decisions that occurred in the course of conducting the research.

These aspects inform the research objectives and research questions that have been identified to guide the work undertaken for the project presented in this thesis.

Research Objective <u>RO 1</u>: Establish the *current state of the art by compiling an overview of existing guidelines* **for minimum dimensions of cartographic symbology, both for printed and screen-based maps.**

Recommendations for minimum dimensions, both for paper and screen-based maps, can be found throughout the cartographic literature. As a foundation for this project, existing guidelines given by various authors, based on empirical evidence or theoretical considerations, shall be reviewed and compiled into an overview.

Chapter 3 will present the state of the art and knowledge of three aspects of relevance for this research project: the capabilities of the human visual system, the technological state of the art for digital displays, and the related work from the cartographic literature. While a review of the related literature is customary for any research project and does not usually warrant to be listed as a distinct research objective, it is hoped that a concise overview of recommended minimum dimensions in tabular form, extracted from the reviewed literature, will be of use for future work by map designers and academics alike. The compiled tables can be found in Section 3.3.2.1 (page 59) for printed maps, and Section 3.3.3.3 (page 69) for screen-based maps.

Research Objective <u>RO 2</u>: Empirically *establish minimum dimensions for fundamental types of cartographic symbology* **for presentation on smartphone screens of various resolutions.**

To fulfil this research objective, three empirical studies will be undertaken: Study 1, presented in Chapter 5, will perform an initial empirical examination of participants' limits to discriminate a wide range of graphical stimuli related to cartographic symbology; Study 2, presented in Chapter 6, will investigate the legibility of point symbols in more detail; Study 3, presented in Chapter 7, will investigate the legibility of line symbols in more detail. For each study, detailed hypotheses guiding the research and assessment of results will be proposed at the outset. Overall, the empirical research is guided by the following research objectives (RO), research questions (RQ), and corresponding general hypotheses (HG):

<u>RO 2.1</u>: Propose, develop and refine a practical framework for empirical investigation of minimum dimensions in cartography.

In the age of open science, the sharing of ideas and tools for the implementation of research projects is an important factor facilitating other researchers reproducing and expanding published research results, or conceiving new research ideas without having to start from a blank slate. Thus the software, apparatus, procedures and concepts developed for this research project are seen as an integral part of its contribution to science. The practical framework initially developed to conduct the research is presented in Chapter 4, and modifications to the original framework are discussed for each of the three studies. Section 8.3 compiles practical advice for map designers and academics who want to conduct their own empirical investigations of legibility of cartographic symbology. All software and data created throughout the course of the project will be made available under open licenses.

<u>RO 2.2</u>: Propose a set of stimuli with well-documented geometry to use for investigating minimum dimensions of cartographic symbology.

As will be discussed in Section 3.3, various authors have used different sets of symbology to discuss or investigate minimum dimensions. While more work than a single thesis would need to be done before a universally agreed standard for such investigations could be claimed, the goal is nevertheless to propose stimuli that can be used for similar projects, to produce results for which a valid comparison can be made across studies. Particularly, stimuli should not be designed only to the needs of theoretical investigation, but reflect cartographic symbology used in real world projects.

RQ 2.2.1: How can sets of map icons, suitable for establishing minimum dimensions, be identified from collections of map icons used in real-world cartographic projects?

In the context of RO 2.2, Sections 6.1.1 and 6.1.2 will propose a method for identifying map icons to use for testing legibility, from real-world icon collections comprising hundreds of icons.

RO 2.3: Investigate the impact of display pixel density on the legibility of cartographic symbology at small sizes.

This is the main objective informing the three studies presented in this thesis. The research questions and hypothesis listed in the following will be picked up in the context of these studies.

RQ 2.3.1: Can cartographic applications potentially benefit from the increased visual fidelity of high-resolution and ultra-high-resolution displays?

Hypothesis HG2.3.1.1: An increase in resolution, from devices at the low end of the spectrum of available resolutions to devices with higher pixel density, results in improved recognition accuracy at small sizes for some cartographic symbology.

Hypothesis HG2.3.1.2: An increase in resolution is never detrimental to the legibility of cartographic symbology at small sizes.

Hypothesis HG2.3.1.3: Participants with high visual acuity will benefit more from an increase in resolution for some cartographic symbology, showing in better recognition accuracy at small sizes, than participants of lower visual acuity.

Hypothesis HG2.3.1.4: Mobile devices of the highest pixel density (>800 ppi) have surpassed the capabilities of humans to resolve small details; legibility of small cartographic symbology will not be improved, compared to presentation on a display with a pixel density around 520 ppi.

RO 2.4: Attempt to find methods that exploit the potential of high-resolution displays and possibly enhance the legibility of small cartographic symbology on such devices.

The goal of this research project should not only be to passively test the legibility of cartographic symbology on a range of devices, but also to actively conceive methods and interventions that may have the potential to exploit the capabilities of modern high-resolution displays to improve the legibility of such graphical elements. Potential interventions for improving the legibility will be discussed in sections 6.1.3 and 6.1.4, and

applied for tasks 5–8 of the second study for point symbols, and for tasks 2 and 3 of the third study for line symbols. Specific hypotheses related to these interventions will be discussed in the corresponding chapters.

Research Objective <u>RO 3</u>: Propose a set of *guidelines, concepts and tools* of practical applicability for the design of screen-based maps.

The author believes that academic cartography should not only produce sound research, but also attempt to translate the results of such research into advice that is actionable in practical map production contexts. Thus, the results and conclusions of the empirical studies should be presented in a form that is of practical use to people who create maps for screen-based consumption. In order to produce guidelines that can be communicated and applied in such contexts, the results of the empirical investigation will potentially have to be simplified and generalised. Chapter 8 will discuss some ethical aspects that have been considered when generalising the empirical results into actionable guidelines, and will present practical guidelines structured into three sections: Section 8.1 will present general advice that can be derived from the empirical investigations; Section 8.2 will give a detailed table of minimum dimensions for cartographic symbology derived from the results, differentiated by display resolution and visual acuity of the map user; Section 8.3 presents advice for practitioners and academics who want to conduct their own empirical investigations of map symbol legibility. Such investigations will hopefully also be facilitated by the framework presented in Chapter 4, by the software developed in the course of this research project, which is made available under open source licenses, and by potential practical software tools that may be developed in the future, as will be discussed in Section 9.4.

The thesis will conclude by reflecting on the work in Chapter 9, which will present an updated comprehensive definition of the term "minimum dimension" in the context of cartographic symbology, reflect about the practical applicability of the presented findings, and discuss limitations of the chosen approach and potential future research opportunities.

3. Foundations and related work

This chapter presents currently available knowledge relevant for the question of the legibility of finest details on a map. Section 3.1 will discuss relevant aspects of human visual perception, with a focus on the eye's capability to resolve and process fine graphical detail. Section 3.2 will present the current state of digital display technology, which includes aspects of display hardware as well as concepts and methods for representing and synthesising graphical content for such displays. Finally, Section 3.3 will compile previous work in the field of cartography related to the question of minimum dimensions. Some authors have hypothesised that once digital displays have reached a sufficient resolution, the guidelines established for the design of printed maps would become applicable to the design of maps presented on digital media. Therefore, existing guidelines for the printed medium will be discussed as well as those proposed for digital displays.

3.1. Capability of the human visual system to resolve fine graphical detail

"There is good evidence that the eyes are important for vision."
(Snowden et al., 2012, p. 20)

Any processing of visual stimuli by humans has to begin with light entering the eye through the cornea, the pupil and the lens, passing through the anterior chamber, and arriving at the fovea, where it can be detected by light-sensitive cells. The visual system is equipped to operate in a wide range of light intensities. The ratio of lowest detectable light, to light intensities so high that may start to damage the eye, is approximately 1:10 000 000 000 (Snowden et al., 2012). The pupil – a dynamic aperture – is the main mechanism to physically regulate the amount of light entering the eye. In an optical system with a dynamic aperture (such as the human eye or a photographic camera), a smaller aperture setting will increase the depth of focus, resulting in a larger range of objects in the outside scene to be in focus. This means that in brightly lit scenes (in which the pupils constrict), more of the scene can be seen in focus without adjusting the lens than in dimly lit scenes (in which the pupils will expand to let more light in).

Adjusting the focus of the eye is done with the flexible lens. The lens is surrounded by the ciliary muscles, which upon contraction compress the lens to a more rounded form. A flatter lens will cause the eye to focus in the distance, whereas a rounder lens (with contracted muscles) will provide focus for closer objects, up to the near point, which is the closest distance the optical system of the eye can focus on. If the focal point of distant objects seen through the lens in its relaxed state lies before the retina and therefore the lens cannot provide

focus for distant objects, the person is short-sighted or myopic (there is no way for the human eye to stretch out the lens beyond the relaxed state, the ciliary muscles can only contract the lens). If the lens cannot be bent enough to focus on close objects, the person suffers from hypermetropia or far-sightedness. With age, the lens in the human eye loses its elasticity and therefore tends to become more far-sighted, a condition called presbyopia. (Snowden et al., 2012).

The *retina* covers the back of the eye where the incoming light is projected onto, and hosts the photoreceptor cells responsible for converting light into neural impulses. There are two types of photoreceptors – rods, which are most sensitive to dim light, and cones, which are less light sensitive but are available in three subtypes, each responsive to a different profile of wavelengths (see Figure 3.1). The retina also contains neural cells for processing and relaying the information provided by the photoreceptors, blood vessels to supply the photoreceptors and neural cells with energy, and the bundle of neurons that make up the optic nerve (composed of approximately 1 million nerve fibres) which delivers the information to the brain. In this arrangement, the density of photoreceptors varies across the retina, and several zones of the overall retina can be identified that vary widely in the perceptual performance they can deliver.
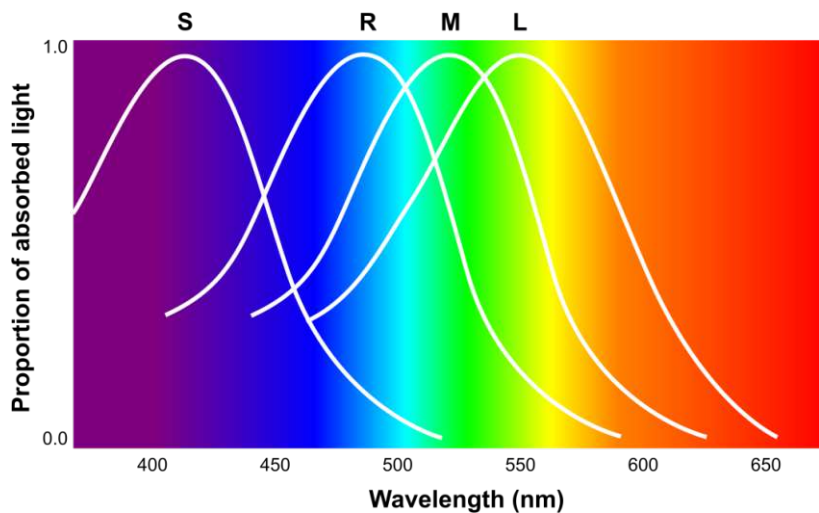
Figure 3.1 Absorption spectra of rod cells (R) (responsible for low-light vision) and short- (S), medium- (M) and long-wavelength (L) cone cells. Recreated after Snowden (2012, p. 154).

At the centre of the retina, an area called *fovea centralis* hosts the densest array of photoreceptors. This area, about 1.5 mm in diameter, is largely devoid of blood vessels, neural cells, rod cells and short-wavelength cone cells, and packs a dense array of medium- and long-wavelength cone cells, which have a more compact form than in other areas of the retina. This area coincides with the innermost ~5 degrees of human's field of vision. In the most compact area, the cones are spaced approximately 2μm apart, resulting in a angular resolution of 0.6 minutes of arc per cone (Westheimer, 2009; Riordan-Eva, 2017). The spatial arrangement of photoreceptors in the retina, though not perfectly regular, can be approximated by a hexagonal lattice in which each photoreceptor integrates the light energy that falls on them (Westheimer, 2009, p. 181).
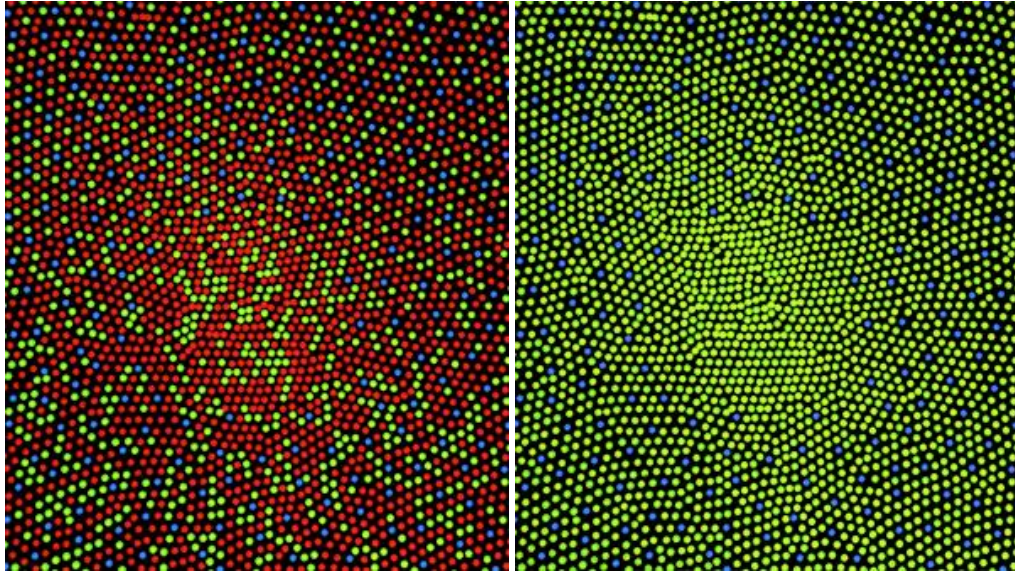
Figure 3.2 Distribution of cone cells in the fovea centralis, with cones sensitive to long-, medium- and short wavelengths coloured red, green, and blue. Left: individual with normal colour vision. Right: colour blind individual. Image Source: Mark Fairchild, CC BY-SA 3.0, via Wikimedia Commons. https://en.wikipedia.org/wiki/Fovea_centralis

The nervous system transmits information emitted from the photoreceptors through electrical impulses, with each impulse having roughly the same electrical potential. Higher intensities of a signal are encoded as higher frequency of such impulses, with an upper bound to how rapidly the neural cell can regenerate its electrical potential limiting the maximum intensity that can be encoded as a signal transported by the nervous system (Bruce et al., 2003, p. 25f). The light falling onto a photoreceptor cell is converted to an impulse that is roughly proportional to the logarithm of light intensity. Such logarithmic coding of intensities is common in the sensory system and helps to compress a wider range of stimulus intensities into a limited range of impulse rates that can be transmitted through the nervous system (ibid.). Another mechanism commonly found in our perceptual system is adaption – a stimulus of constantly high intensity is encoded by impulses that decrease in frequency after the initial onset. In photoreceptors, the duration after which the frequency of impulses falls back from the peak at the onset to a steadily sustained level, which continues to be proportional to the logarithm of intensity of the stimulus, is under 1 second (Bruce et al., 2003, p. 26).

3.1.1. Visual acuity

> *"much of what your visual system is required to do*
> *has little to do with the fine detail"*
> *(Snowden et al., 2012, p. 115)*

The previous section has shown that the human visual system does not act as a regular sampling device, but that even in the early stages of perception, complex mechanisms are at work that transform, filter and emphasize parts of the visual image that is projected onto the retina. However, particularly in the fovea, which corresponds with the centre of our vision, the size, density and sensitivity of photoreceptors, and the low-level neural processing that

takes place in the retina, set a limit to the finest details a person can recognize in the environment.

Due to the fact that a larger object at a farther distance and a smaller object at a proportionally reduced distance may result in the same proportion of the retina covered by their projected image, it is common to specify the *apparent size* of objects as projected onto the retina, independent of their distance from the viewer, in angular units. The apparent size of an object can be calculated using the equation

$$\delta = 2\arctan(\frac{S}{2D})$$

in which *S* is the real-world size of the object and *D* is the distance from the focal point to the object. This relationship, and the fact that two objects of different size may have an identical apparent size, is illustrated in Figure 3.3.
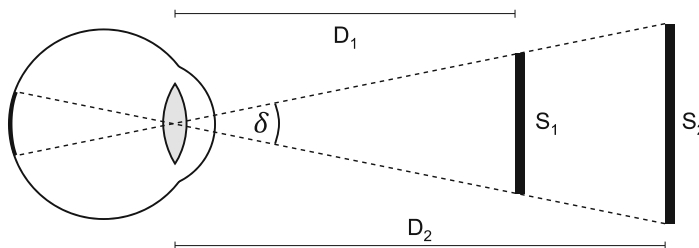
Figure 3.3 The apparent size of two objects, expressed as the angle δ. The size of object S_1 at distance D_1 may be identical to the apparent size of a larger object of size S_2 at a proportionally larger distance D_2.

The concept of visual acuity refers to the smallest detail of sufficient contrast an observer can detect in the environment. Because this is mainly dependent on the extent of the retinal image of the feature, the relevant dimension is the apparent size (measured as an angle), not the real-world size of the feature. A fundamental concept to assess visual acuity is thus the *minimum angle of resolution* (MAR) as the apparent size of the smallest detail that can be detected in the retinal image.

The earliest definition of visual acuity as the capability to separate two points a certain distance apart was given by Robert Hooke in 1674

> "'tis hardly possible for any unarmed eye well to distinguish any angle much smaller then that of a minute: and where two objects are not farther distant then a minute, […] they coalesce and appear [as] one." (Hooke, 1674, p. 9)

It is notable that already in this earliest definition, an estimate of one minute of arc is used to quantify human visual acuity. Hooke also stated that the "rarest of individuals" have visual acuity allowing them to detect separations of half a minute of arc in ideal conditions (Buchwald & Feingold, 2013, p. 50).

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

TU Bibliothek
Your knowledge hub
WIEN

Current conceptualizations of visual acuity examine the eye's capability to detect high spatial frequencies instead of separating individual dots or lines (where there is no obvious choice, for example, for how large or wide these graphical elements should be). Modern psychophysical experiments often use *sinusoidal gratings* instead of discrete bars of maximum/minimum intensity, since the sharp edge of a bar is in itself a high-frequency phenomenon (a large change in intensity occurring over a narrow spatial extent) and could therefore interfere with the examination of lower-frequency detection capabilities (Snowden et al., 2012, p. 114).

3.1.1.1. Theoretical limits of visual acuity

"the question of resolution contains some ambiguities that remain challenging to this day."
(Westheimer, 2009, p. 179)

Whenever a continuous, highly detailed signal (such as a visual image of the real world) is sampled with an arrangement of discrete sensors (such as the photoreceptors in the retina), the Nyquist-Shannon sampling theorem is of relevance, which says that to reconstruct a signal containing a maximum frequency component of *B* hertz (having a wavelength of λ), a sampling frequency greater than *2B* (with a wavelength smaller than λ/2) is needed[4]. For application in the spatial domain (such as sampling an image), the theorem can be applied to spatial instead of temporal frequencies. A black line of width *d* against a white background, or two points separated by a distance *d*, represent a signal with a (local) wavelength of *2d*, since one cycle contains the entire transition from white (maximum intensity) to black (minimum intensity) and back to white, and can thus be sampled with a sampling interval of slightly smaller than *d*.

For signals of a shorter wavelength than λ, the signal cannot be reconstructed accurately from samples spaced at a distance larger than λ/2. In such a situation of sampling at too low frequency, *aliasing* effects can distort the reconstruction of the original signal. This may lead to the sampling of a constant intensity or of a signal of apparently lower frequency from a high-frequency signal.

Applying the sampling theorem to the fovea centralis, it has been mentioned above that the cone cells located there are spaced about 0.6 minutes of arc apart, which, given an optimal and regular distribution, should allow humans to sample spatial signals slightly larger than that in the centre of their visual field, according to above consideration that an object of size *d* corresponds to a spatial wavelength of *2d* (Westheimer, 2009).

Even if the retina could sample spatial signals at infinite resolution, other factors would limit the capability of the human eye to resolve fine details. The resolution that can be accomplished by an optical instrument is not unlimited, even under ideal circumstances. A main factor limiting the theoretically achievable resolution is *diffraction*, caused by the

---

[4] While strictly speaking the Nyquist-Shannon theorem applies only to the sampling of mathematical functions which are composed of a finite number of periodic functions, it is routinely applied to the discussion of sampling of real-world signals (which could be thought of having infinite frequency components and therefore cannot be reproduced with infinite fidelity by discrete sampling).

interferences of light waves passing through a circular aperture. Due to the different distances that light waves have to travel from opposing edges of the aperture to the projection surface, a phase shift occurs that causes the projected image to blur. For monochrome light, the minimum distance for which two points can still be clearly distinguished can be calculated using the Rayleigh criterion:

$$\theta = 1.22 \frac{\lambda}{D}$$

Where θ is the minimum angular distance of two points that will result in two distinct intensity maxima of high enough amplitude in the projected image, λ is the wavelength of incoming light, and D is aperture diameter. For the human eye, approximate values can be substituted as such: The aperture diameter of the pupil is between 2 and 9mm, and the wavelength of visible light is in the range of 390 to 750nm. An estimate of the resolution capability of this optical system can therefore be given to lie between a minimum of 0.2 minutes of arc (violet light, fully open pupil) and a maximum of 1.5 minutes of arc (red light, pupil closed to 2mm) (Schor & Miller, 2011). For common viewing conditions, this theoretical limit is comparable to the diameter of a single cone cell in the fovea, which is indication that the eye has evolved to match those physical limitations (Brodie, 2019, p. 23).

It is important to note that the limit proposed by Rayleigh is a "rule of thumb" and not an absolute criterion (Westheimer, 2009, p. 179). For two point stimuli separated by the Rayleigh limit, there will be a drop of intensity between the two intensity maxima at the centres of the points of about 20%, which can be used to accurately detect the presence of two separate objects either by exploiting optical interference (Rueckner & Papaliolios, 2002) or by sampling intensity with high enough fidelity to reconstruct the separation (Westheimer, 2009). Even when the points are so close that there is no drop in intensity between the two maxima, the spatial distribution of intensity will always be different for two points than for a single one, if a hypothesis about the shape of the stimulus can be assumed, which would allow for a probabilistic estimate about the true nature of the stimulus (ibid.).

Further factors which limit the maximum acuity that can be accomplished by an optical system are *scattering* (the random dispersion of a portion of light in an imperfect medium) and *chromatic aberration* (the fact that light of different wavelengths/colour has a different focal length when projected through the eye's lens). Scattering mainly affects contrast sensitivity and has only a strong detrimental effect when some part of the optical system of the eye is damaged. In a healthy human eye, scattering is estimated to account for a loss of 10% of light on the way to the retina (Schor & Miller, 2011, p. 19f). However, the eye has evolved some mechanisms to reduce the effect of scattering, namely the orientation of photoreceptors to accept light mainly from the direction of the centre of the lens, and the embedding of yellow pigments in the fovea to absorb scattered blue light. This can be compared to the effect of yellow-tinted skiing glasses, which increase sharpness by reducing the spectrum of light that passes through (ibid.). Chromatic aberration is compensated in the eye by including only "red" and "green" cones in the fovea centralis (recall Figure 3.1, which shows that the wavelengths of maximum intensity of these cone types are not very far apart),

thus also significantly narrowing the spectrum of light that is processed in the area of highest visual acuity.

This section has shown that the density of cones in the fovea is not the sole limiting factor of human visual acuity, but that the eye is an optical system that cannot be easily improved by "upgrading" a single one of its components. The retina's visual resolution has evolved to match the overall optical properties of the eye. As Westheimer writes: "in a striking evolutionary convergence of optical and anatomical imperatives, the optimal sampling and actual binning intervals coincide in the human fovea" (Westheimer, 2009, p. 182). Of course, humans have developed optical visual aids, from corrective glasses and contact lenses to magnifying instruments and telescopes, to improve our effective visual capabilities, but these modify the light outside our eyeballs to produce a modified retinal image. Within the constraints of biology and physics, the human eye approximates an optimal visual sensing device, at least in the small area of the fovea centralis of healthy individuals.

Outside of the fovea centralis, acuity drops rapidly and decreases with distance from the centre of vision, corresponding with a reduced density of both photoreceptors and nerve fibres transporting the information to the brain, and a reduced number of neurons in the cortex dedicated to the processing of information received from peripheral vision. At a distance of 10 degrees from the centre, acuity is already reduced to 1/10th of the maximum achieved in the fovea (Ware, 2020).

3.1.1.2. Hyperacuity

> "It would be rash, however, to draw the conclusion
> that the subject of visual acuity is closed."
> (Westheimer, 1981, p. 3)

There is a class of visual tasks that can be successfully performed by humans that would require visual acuity beyond the theoretical limitations outlined above and therefore cannot be explained by the eye's *spatial* sampling capabilities alone. Such phenomena that exceed the apparent physical capabilities of the eye are referred to as *hyperacuity*. As has been argued, the photoreceptors in the retina are not binary sensors (light on/light off), but also sense the intensity of light falling onto the receptor. As such, photoreceptors perform a binning operation across the visual field, each collecting light intensities across an area defined by the diameter of the photoreceptor, and reporting the overall intensity of light in that area (Westheimer, 2009, p. 181). It is believed that for hyperacuity phenomena, the sampling resolution of binned intensity values is substituted for (lack of) spatial resolution, and later neural processing allows the original (spatial) signal to be reconstructed from the overall sampled field of intensities (Westheimer, 1981; Schor & Miller, 2011).

It has been discussed above in Section 3.1.1.1 that the retinal image of two fine lines separated by a very small distance is subtly different than that of a single wider line, even when the two lines are closer than the distance mandated by the Rayleigh criterion. This fact can be exploited by human observers to discriminate double lines from single lines closer than the Rayleigh limit in psychophysical experiments (Westheimer, 2009). Telegraph cables against a bright sky can be seen by the naked eye in situations that translate to much less than the 1

minute of arc established above as a coarse estimate for foveal resolution. Whether such a line can be detected depends only on whether the local decrease of light intensity in the fovea is within the receptors' luminance difference threshold (Westheimer, 1981, p. 4). But there is a fundamental difference between tasks that merely require the detection of presence or absence of some target, and tasks that require making spatial judgements about the arrangement of multiple objects (ibid.).

Tasks for which human observers demonstrate to successfully make spatial judgements beyond theoretical foveal resolution are discussed in the psychophysical literature. These include tasks of bisection (juding whether the central of three lines is offset to the left or to the right), lateral displacement in so called "Vernier" tasks (judging whether a line segment is laterally displaced to the left or to the right from a collinear second segment), and the detection of a slight angle in the continuation of two line segments (see Figure 3.4). Hyperacuity can be improved with practice and is independent of stimulus orientation (McKee & Westheimer, 1978). Human subjects are typically able to classify hyperacuity stimuli with a success rate of 75% for displacements of 6 seconds of arc or less (Westheimer, 1981).
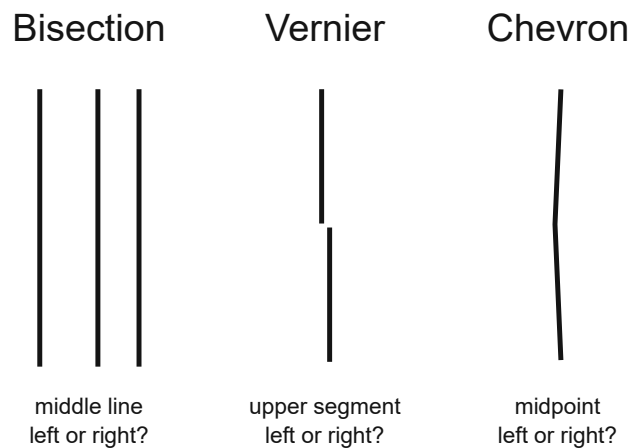
Figure 3.4 Hyperacuity stimuli used to demonstrate (a) bisection (b) Vernier (c) chevron acuity. The questions at the bottom can be answered by human observers at smaller displacement levels than what would be expected based on the density of photoreceptors in the retina.

For hyperacuity stimuli, additional components in the image such as additional parallel "flanking" lines reduce the threshold of detection or introduce bias (Ruda, 2013, p. 16). This raises the question whether hyperacuity can be made use of in real-world application scenarios, where stimuli will usually contain complex graphical arrangements instead of isolated stimuli. Hyperacuity phenomena have only been found to work for very specific stimulus designs, and cannot be generalized into a model of general higher visual acuity.

It is notable that Vernier acuity has found a real-world application in the "Vernier scale" used in callipers and other measuring devices. To determine the sub-millimetre component of a measurement, the human observer has to identify the two lines on the scale that match up exactly with the millimetre grid – a task that can be performed much more reliably thanks to

Vernier acuity than trying to directly read the measurement from a finer measuring grid (see Figure 3.5).

Figure 3.5 Vernier scale on a calliper. Because the human eye is very good at detecting the alignment of collinear lines, the line in the lower part that lines up precisely with one of the lines on the millimetre scale can be easily identified (the line numbered "6" at the bottom), and conveys the fractional component of the measured distance (in this case: 2.6 mm).

3.1.1.3. Effect of colour and contrast ratio on visual acuity

Photoreceptors in the retina are not equally sensitive to light of all wavelengths, with maximum sensitivity at a wavelength around 540 nm, in the green-yellow part of the spectrum, and sensitivity reduced by a factor of over 100 at a wavelength of 450 nm, in the blue part of the spectrum (Ware, 2020). The term luminance refers to the amount of light energy reaching the eye, weighted by the spectral sensitivity profile of the photoreceptors. The luminance ratio between two stimuli, or different parts of a stimulus, is referred to as contrast ratio. ISO standard 9241 recommends a contrast ratio of 1:10 for the display of text (ibid.). While this restricts the combination of colours that can be used for foreground and background of fine details (e.g. it is not advisable to use blue text on a black background, due to the low sensitivity of photoreceptors to blue light mentioned above), such contrast ratios are several orders of magnitude below the contrast ratios achieved with modern displays.

One way of expressing the contrast of different parts of a stimulus as a single number is Michelson contrast, calculated with the formula

$$C = \frac{I_{max} - I_{min}}{I_{max} + I_{min}}$$

where I_{max} and I_{min} are the highest and lowest luminance values (Ware, 2020). Michelson contrast results in values between 0 and 1, and is usually expressed as percentage value.

Decreasing the contrast of the stimulus affects visual acuity. However, contrast sensitivity does not decrease uniformly, but remains highest for spatial frequencies in the range of 1-5 cycles per degree (Campbell & Maffei, 1974), which translates to line widths of 6-30 minutes of arc, and decreases for frequencies higher than that. At contrast ratios above 75%, changes in contrast have a negligible effect on acuity (Sloan, 1951, p. 721; Colenbrander, 2016).

Specifications for the production and illumination of charts used for testing visual acuity do not mandate exact contrast and luminance values, but specify a contrast ratio between 1:20 (~90%) and 1:33 (~94%) for printed charts, and as low as 1:5 (~67%) for projected charts, and recommend a luminance between 80 and 320 cd/m² (Schor & Miller, 2011, p. 10). The standard for visual acuity testing ISO 8596 mandates a contrast value of 85% or higher (Bach, 2007).

3.1.1.4. Higher-level processing of visual information

"How human vision works is, of course,
incompletely understood"
(MacEachren, 1995, p. 53)

The light intensities projected onto the photoreceptors in the retina are not transferred directly as a "pixel image" to further processing in the brain. Already in the retina, a layer of ganglion cells process the signals coming from photoreceptors, and apply mechanisms of grouping, amplification or inhibition to the signals. In addition to the temporal non-linearity of the nervous system, this compresses the information emitted from photoreceptors in the spatial domain, amplifying edges and changes of intensity, and dampening spatially constant stimuli (Snowden et al., 2012, pp. 35f, 49f). The signal processed in this way is then delivered through the optic nerve to the lateral geniculate nucleus (LGN), where information from both eyes is integrated and further low-level processing of the signal – such as the detection motion, colour and linear segments of specific angles – is done (ibid., p. 37ff). Importantly, the LGN receives most input not from the retina, but from other regions of the cortex. This is indicative that already on this low level, processes of filtering and selection of the visual input take place, which are controlled by higher level cognitive processes, to "direct the spotlight of attention" (ibid., p. 40f). Also, of all nerve fibres that connect the eyes to the visual cortex, a disproportionate amount is dedicated to the fovea (approximately 10% of all neural connections, despite the fovea covering only about 0,005% of the visual field) (Snowden et al., 2012, p. 76). Photoreceptors in the fovea also have "private lines" of direct nerve connections to higher levels of neural processing, bypassing the neural preprocessing applied to photoreceptor signals outside the fovea. (Westheimer, 2009, p. 185).

Visual information is thus already heavily processed, selected and of non-uniform resolution even before it reaches the cortex and is there subjected to any kind of conscious interpretation. Further processing in the cortex happens in multiple stages, for which at lower levels, orientations of linear components of the image, length of line segments, spatial frequencies or objects at a particular speed are identified, while higher levels seem to react to specific trigger features matching multiple visual attributes (Snowden et al., 2012, p. 77ff; Kalat, 2018, p. 170). It is notable that also the processing speed of information varies for different visual stimuli: low spatial frequencies, used for detecting motion or spatial location are processed more rapidly, while information of high spatial frequency, which may be used to infer the exact shape of objects, is transmitted via slower channels. This means that coarse – and often unconscious – processing of the information in the visual field (potentially also in the peripheral vision, where photoreceptors are more sparse) is happening faster than the detailed processing that allows for the conscious recognition of shapes and objects in the scene (Eysenck & Keane, 2020, p. 104).

Information from the lower stages of the visual cortex eventually reaches the inferotemporal cortex, where neurons respond to specific semantic categories recognized in the stimulus, such as body parts, faces or different types of animals (Eysenck & Keane, 2020, p. 50). Up to which semantic differentiation the visual system directly triggers neural responses is presently not precisely known – while it could be shown that there are specific neurons that trigger whenever a human face (or even the caricature of a face) or, for other neurons, the shape of an insect is visible in the visual field (Snowden et al., 2012, p. 86), there is much debate whether specific neurons are always responsible for recognizing specific objects in the world, or whether recognizing such objects is done through a combination of perceived properties (Chlupac, 1982, p. 12). The proverbial "grandmother neuron", which would trigger whenever an image of an observer's grandmother is received by their eyes (but not in the presence of other elderly people etc.), has not been demonstrated to exist, and there is some evidence against higher-level object recognition actually working by mapping highly specific objects in the world like concrete persons to individual neurons (Snowden et al., 2012, p. 85ff)

Certainly, adult humans are able to recognize well-known elements of their environment, but also integrate it with new information, and also recognize new elements they have recently encountered when presented again without first having to "learn" them.

Conceptually, the extraction of information from visual sensation is thought to be structured by *visual channels*, including aspects of form, for example orientation of edges and spatial frequencies, motion, such as the direction of movement, and colour. These aspects are evaluated across the entire visual field, and can direct the attention – and, by eye movement, the centre of vision – of the observer to relevant parts of the scene. It is a well-known technique in information visualization to differentiate symbols by multiple channels (e.g. by shape, orientation and colour) in order to make them more readily distinguishable (Ware, 2020).

Many types of visual information can only be picked up from the scene by an observer by directing their attention, and the centre of the field of vision, to individual elements – for example, reading the sequence of digits of a long number. However, differentiating elements in the scene by particular channels may facilitate a "popout" effect, allowing the observer to register their presence at a glance, without directing their attention towards each individual element. Examples for channels facilitating such preattentive processing of information are orientation, colour, size, blur and pronounced differences in shape, curvature or lightness (Ware, 2020, p. 158f). Because such preattentive features are picked up without centring the visual field on each feature, it is implied that the features must be large enough to be sensed by the resolution capabilities of peripheral vision, which, as discussed above, is several orders of magnitude lower than that of the central fovea. Sophisticated processing and extracting of information from a scene, like the perception of fine geometric details of a feature, has to be performed after directing the focal attention to, and centring the visual field on, the feature (Snowden et al., 2012, p. 284).

In the context of cartography, authors have advocated for an "ecological" approach towards understanding visual perception and called for studying how maps direct the user's

attention, potentially before any conscious analysis of minute graphical detail takes place (see MacEachren, 1995 for an introduction). However, the understanding of the physiological and neurological properties of human vision presented above leads to the conclusion that the finest details of graphical features can only be perceived and interpreted by directing the observer's attention, and centre of vision, towards the features in sequence. A comprehensive model of visual perception and, in particular, visual acuity that integrates aspects of preattentive processing with the visual system's capabilities for conscious detailed analysis of finest graphical detail, and could be used to analytically derive the dimensions of stimuli required for supporting various tasks related to visual information processing, remains presently elusive (Healey & Enns, 2012).

3.1.2. Establishing visual acuity in ophthalmology and psychophysics

Experimental testing of human visual acuity has a long history, going back to Hooke's works in astronomy in the 17th century. For ophthalmological purposes, Herman Snellen (1873) proposed a set of letters that allowed the standardized testing of the visual acuity of human subjects. Snellen's original publication of test charts includes individual letters and paragraphs of text in the then widely used "Fraktur" (calligraphic) style, but also the "tumbling E" stimulus which continues to be widely used to assess visual acuity to this day (Sloan, 1951; Snowden et al., 2012). Also the basic idea to construct letters on a 5×5 grid, with strokes and gaps being one grid cell wide, was introduced in Snellen's publication.

Edmund Landolt recognized that the letters in Snellen's charts were not equally difficult to discriminate. As a more neutral test, he proposed a circular shape of outer diameter d and stroke width $d/5$, with an opening in one of the four cardinal directions of width $d/5$, this keeping with the 5×5 grid system established by Snellen. This had the advantage of equal difficulty for identifying each of the 4 stimuli, and being usable for test subjects not familiar with the western alphabet. Despite these advantages, the Landolt ring "gained only limited acceptance in clinical use" (Colenbrander, 2016), potentially because verifying the correctness of the subject's response requires "more attention and concentration […] of the examiner" to correctly keep track of the sequence of answers (Sloan, 1951).

Louise Sloan proposed a set of 10 letters with revised design, maintaining Snellen's basic 5×5 grid and using the Landolt ring as the letter "C" (Sloan, 1959). Sloan also introduced an updated notation for measured visual acuity: $V = m/M$, where m is the test distance (in meters) and M is the letter size in M-units, with 1 M-unit being equal to 5 minutes of arc at 1 meter distance (≈1,454 mm) (Colenbrander, 2016). The fraction is usually left intact when specifying test results, so that the information about the distance at which the test was performed at is preserved. For example, a visual acuity of **5/5** would indicate that the test was done at 5 meters distance (first number), and letters of a height of 5 minutes of arc (5 M-units × 5 minutes of arc / 5 meters) could be successfully discriminated by the tested individual. Since strokes and gaps of the letters are 1/5 of the overall height, each stroke would cover 1 minute of arc in this example. The inverse value of the fraction corresponds with the angular resolution (MAR) in minutes of arc, with a value of 1 (as in the example just given) corresponding to a visual acuity of 1 minute of arc, a value of 5/2.5 to a resolved detail of ½ minute, et cetera.

Testing for visual acuity is the single most widely applied medical examination of the eyes (Westheimer, 2009). Basic medical acuity tests have changed little since their introduction by Snellen (Westheimer, 2009). Testing is usually done a medium to far distances, with 5 meters being typical on the European continent, 6 meters in Britain, and 20 feet (6.096 m) in the US (Colenbrander, 2016). Results for 1 arcminute of visual acuity are accordingly specified as 5/5, 6/6 and 20/20 in these regions, reflecting the different testing distances but giving an equal numerical result of 1. For near vision, Sloan (1959) recommends 40cm (or 16 inches) distance as it "corresponds more closely to normal reading and work habits" (p. 811), requires less adaption of focus, and because such measurements can more easily be converted to the corresponding values at 20ft distance. The International Council of Ophthalmology recommends a testing distance at 40cm, mainly because of simple correlation to measurements taken at 4m distance, but "testing at other distances is accepted" (International Council of Ophthalmology, 1984). A value of 14 inches (33 cm) can also be found in the medical literature as the recommended distance for near-distance acuity testing (Chang, 2017, p. 71), while others specify the subject should hold the test chart at "comfortable reading distance" (Salmon, 2019, p. 4).

Modern assessment methods specify the accomplished visual acuity as the base-10 logarithm of the minimum angle of resolution (logMAR), with a lower logMAR score corresponding to better visual acutiy. A visual acuity score of "20/20", corresponding to a resolved detail of 1 minute of arc (see discussion above), thus results in a logMAR score of zero. A score of "20/10" on the conventional scale, marking the limit of human visual acuity, would result in a logMAR score of -0.30. Different Methods for testing visual acuity have been shown to yield results that differ up to 0.10 logMAR units (Pointer, 2008). It is therefore recommended to specify which test was used in order to get consistent results on subsequent testing.

Multiple proposals have been made for symbols to be used for assessing the visual acuity of individuals unfamiliar with roman letters, including preschool children. These include subsets of letters or single letters in different orientation, with the "rotating E" being used for illiterate subjects (Chang, 2017, p. 72), to symbols specifically designed to provide more choices without the need to recognize roman letters. A recent example of a symbol set that has been designed with the requirements for objective assessment of recognition acuity in mind is the "Auckland Optotypes" set of symbols (Hamm et al., 2018). These symbols have a 1:1 aspect ratio and consistent stroke width, require no internal lines to draw and minimize ambiguity between symbols, both geometrically and semantically. Figure 3.6 shows various stimulus types which are used to assess the visual acuity of human subjects.

Ophthalmological examinations are mostly concerned with detecting visual acuity worse than logMAR 0 (1 minute of arc of resolved detail), which has been traditionally defined "normal" or "good" vision, although the average acuity of healthy young adults is better, as will be discussed in the next section. Determining the limits of visual acuity in healthy individuals for different stimuli, and establishing the precise relationship between stimulus size and correctness of responses, has mainly been done in the context of psychophysical studies.

Gustav Fechner introduced the idea to estimate thresholds of perception and to measure the relationship of physical stimuli and a person's subjective experience (Goldstein & Brockmole, 2016, p. 13). One of the methods introduced by Fechner is the *method of limits*, in which stimuli are presented to an observer in ascending or descending order of intensity in a series of trials, in order to determine that person's absolute threshold for detecting the presence of a stimulus. However, this threshold will usually not be a hard limit, above which the recognition rate is 100% and below which recognition immediately drops to zero, but can be modelled by a continuous *psychometric function* that plots stimulus size or intensity on the x-axis and probability of correct response on the y-axis. The plot of the psychometric function will be a continuous, monotonically increasing curve, with a value corresponding to the random chance of the participant guessing the correct answer for stimulus sizes well below the threshold (e.g. 0.25 if four choices are offered among which the answer has to be selected, or 0.1 for the chance of guessing one of the ten letters of the Sloan chart), and a value of 1-λ for sizes well above the threshold, representing a certain residual chance λ that stimuli may be incorrectly identified even when presented at large enough size, with the estimate for λ depending on the difficulty of the task (Cunningham & Wallraven, 2011).
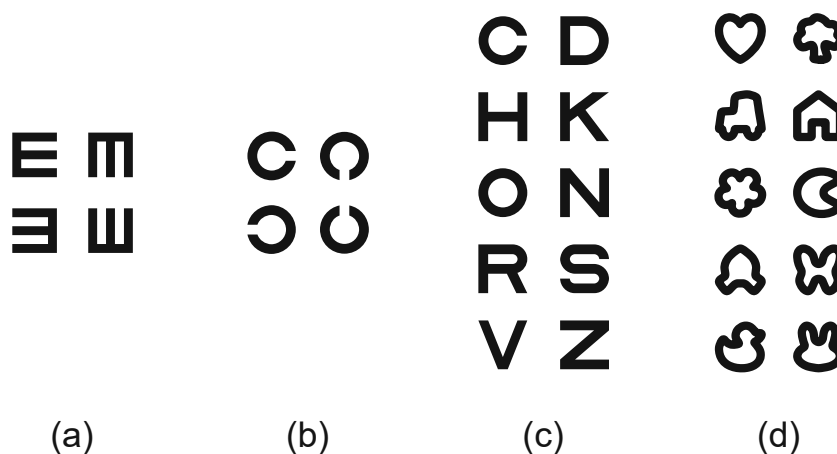
Figure 3.6 Stimulus geometries used for assessing visual acuity. (a) Snellen-E's (b) Landolt-C's (c) Sloan Letters (d) Auckland Optotypes (Hamm et al., 2018), CC-BY-NC-SA Dakin Lab.

Modelling acuity as a probability distribution consequently does not yield a binary separation of "legible" and "illegible" stimuli, but a continuous psychometric function, with the probability of simply guessing the correct answer at the one end, and a probability slightly lower than 100% on the other end, taking into account the (small) chance that even if the stimulus is clearly legible, a wrong answer may be given (e.g. by accident or equipment malfunction). Therefore the psychophysical literature generally deals with success rates below 100% as their focus of interest. Ware (2020) claims that a success rate of 90% is typically used as a threshold for success in perceptual studies and ophthalmological examinations[5]. Bach (2007) notes that, given a continuous relation between stimulus size and success rate,

---

[5] Quote: "20/20 vision means that a 5-minute letter can be seen 90% of the time." (Ware, 2020, p. 50) I could not find further evidence for this in the ophthalmological literature, where authors usually demand successful reading of a whole line of 5 Sloan letters as a criterion for passing the test at a given level (D. F. Chang, 2017, p. 72; Biousse & Newman, 2019, p. 2)

the definition of a single value as a threshold is "not obvious". Bruce et al. (2003, p. 153) specify a success rate of 75% in a task with two choices (halfway between random chance and complete success) as "a reasonable definition of 'just detectable'". Other authors agree on such a definition of the threshold value as "the average of the chance level and the highest performance level" (Cunningham & Wallraven, 2011, p. 354), and it can be argued that reporting this point for the results of experiments is also the "optimal choice from a signal-detection point of view" (Bach, 2007, p. 3). Notably, a stimulus at this level can be assumed to be the "most uncomfortable one for the patient: here, they are most uncertain whether or not they can recognize the target" (ibid).

Psychophysical experiments for assessing the psychometric function often follow an "alternative forced choice" (AFC) design, in which participants are presented with a stimulus (e.g. one of the Sloan letters) and a set of options as responses (e.g. the 10 letters of the Sloan set), from which they are asked to select the correct one. Participants are instructed to make a best guess at which letter is presented, and that they could select one of the choices at random if they really cannot discriminate the stimulus – an option for "I don't know" is not provided. This will result in a success rate representing the chance of randomly guessing the correct answer at levels at which the stimulus could truly not be identified, and higher success levels above that level (Cunningham & Wallraven, 2011).

Computers and digital displays are widely used to assess visual acuity for medical and psychophysical purposes. Computer-based experiments offer more flexibility in stimulus presentation (Metha et al., 1993), but also are the basis of more sophisticated methods to determine the stimulus size for each trial for maximally efficient testing, potentially taking the previous performance of the participant into account to determine the next level to show (Lieberman & Pentland, 1982). Traditionally, psychophysical experiments have used methods of constant stimuli (presenting a predefined set of stimulus sizes for a certain number of repetitions to each participant) or simple staircase procedures (adjusting the stimulus size up or down, depending on the correctness of the response) (Cornsweet, 1962) to adjust stimulus sizes. Using a computer to determine stimulus size and to display the stimulus on a monitor allowed for the implementation of more sophisticated methods that maximise the amount of information gathered in each trial for the purpose of estimating the psychometric function (Lieberman & Pentland, 1982; Watson & Pelli, 1983).

CRT monitors were widely used for administering acuity tests, and needed to be calibrated to provide a sharp and colour-corrected image (Metha et al., 1993). The Freiburg Visual Acuity Test (FrACT) initially provided a test for visual acuity based on Landolt rings (Bach, 1996) and provides today a web-based testing environment, offering a choice of different stimulus types and calibration options (Bach, 2007, 2020). The FrACT software is free to use and cited in over 1200 studies (Bach, 2020), indicating widespread use of such software in scientific contexts.

3.1.2.1. Demographics of visual acuity

> *"In spite of the fact that normal human visual acuity had already settled*
> *on the 1 arcmin value in Snellen and Landolt 's days […], modern data*
> *on typical resolution thresholds vary over a considerable range."*
> *(Westheimer, 2009, p. 182f)*

> *"if an eye, with good illumination, has only 20/20*
> *we can be almost sure that it has problems easily detectable."*
> *(Tscherning M. 1898: Le sens des forms, quoted after Velasco e Cruz, 1990)*

A minimum angular resolution of one minute of arc has been established as a rough estimate for human visual acuity ever since Robert Hooke's experiments and publications. However, Velasco e Cruz (1990) argues that the confirmation and continuous use of this value by later science may have been based on misunderstanding. Experimental verification of the minimum separable threshold undertaken by Helmholtz in the 19th century indeed reported a value of 63 seconds of arc (roughly equal one minute) for detecting two distinct bars, but Helmholtz reported the distance of the *centres* of two bars, not the width of the separation (Velasco e Cruz, 1990). This reported width of a complete cycle of black-white-black was later erroneously used as the width of a single bar (half a cycle) in Snellen's and later eye chart for measuring normal visual acuity. It is important to keep track of whether a specified value refers to a full cycle between two maxima (or minima) of intensity, or to a single "bar" or half-cycle. Ophthalmological literature seems to have settled on specifying the width of half-cycles as the relevant measurement, but arguably uses a too high value – based on initial experimental verification of the minimum width of a full cycle – as a threshold value for "normal" visual acuity.

As with any phenomenon rooted in biology, there is a certain variance to be expected across the population regarding the performance of the visual system. The many components of the optical system – the shape of the lens and the eyeball, the consistency and transparency of the lens and the body of the eye ball, the density and distribution of photoreceptors and the neural processing of nerve impulses, are all subject to individual differences, resulting in exceptional performance for some individuals while presenting limitations for others. Some of these limitations, such as diffraction errors caused by mismatching geometries of lens and eyeball (as discussed in Section 3.1.1.1), can be corrected by equipping the eye with glasses or contact lenses, which are generally widely accessible in developed countries. Other deficiencies cannot be readily overcome, and, depending on the magnitude of deviation from a median of "normal" performance, may be considered within the norm or, in more severe cases, as impairments that are acknowledged to limit an individual's capabilities (International Council of Ophthalmology, 2002).

In addition to individual differences, visual performance deteriorates with age. Many components of the visual system are subject to diminished performance as the body ages. The lens is continues to grow throughout an individual's lifespan (Glasser, 2011, p. 63) and increases in stiffness (ibid., p. 66), causing refraction errors and reducing the eye's capability to focus on near targets. The accommodation capability of the eye begins to deteriorate already in at young age and reaches a minimum plateau around the age of 50 years (Diniz et

al., 2019, p. 37). Contrast sensitivity is reduced in 60 year olds, particularly at high spatial frequencies, due to the increased scatter of the lens (Diniz et al., 2019, p. 26; Ware, 2020, p. 60).

In spite of the historical difficulties to find a stable definition, and the variance of visual performance in healthy adults, the value of one arcminute (for a half-cycle) is long established and has persisted as a threshold for "good" visual acuity (Westheimer, 2009, p. 182). However, as has been discussed above, it is widely acknowledged that typical adult vision can be expected to be better than logMAR 0, and does not drop below that value until the age of 60 or 70 (International Council of Ophthalmology, 2002).

What is, then, the visual performance that can be expected of the average human? It can be argued that due to the availability of corrective measures for diffraction errors, it can be expected from adults to have their vision corrected if their visual capability drops below a certain level. Furthermore, individuals with severe pathologies of the eye would not be considered to be "normal" viewers. A baseline expectation would therefore take into account the visual acuity that can be reached by individuals after applying reasonable corrective measure. Such best corrected visual acuity of healthy individuals has been shown to lie at -0.13 logMAR (standard deviation: 0.06) for 18-24 year olds, improve further to a maximum demographic acuity of -0.16 (SD: 0.06) in 25-29 year olds, and then slowly drop to an average value of -0.05 (SD: 0.05) in the age group of 65-69 year old subjects, and to -0.02 beyond that age (Elliott et al., 1995). This shows that visual acuity across the population, particularly at pre-retirement age, can be expected to be better than the commonly used benchmark of 20/20 vision, if optimal correction is available. Studies assessing habitual correction (the participants' correction that they are equipped with in real life) showed acuity scores worse by about 0.15 logMAR units compared to the numbers of Elliott et al. for 70-year olds, but did not exclude non-healthy subjects and may have suffered from methodological problems (Elliott et al., 1995, p. 189). Other studies undertaken to modern standards largely agree with the numbers established by Elliott et al. (Ohlsson & Villarreal, 2005, fig. 2; Radner & Benesch, 2019). Radner & Benesch (2019) note that visual acuity in 70-74-year-olds "is well above the level that would interfere with everyday life".

Frisén & Frisén (1981) present a model of 95% confidence intervals for expected visual acuity by age group, and argue to use a 90% success criterion in vision tests to avoid random errors influencing the results. While, consistent with other studies, average visual acuity based on a 90% success criterion does not drop below -0.05 logMAR up to the age of 79, the confidence interval stays fully below 0 logMAR only until the age of 59. This indicates that while average acuity may stay above the 20/20 mark for much of the older population, it should be very well expected that individuals with acuity performance below that mark will be encountered above the age of 60.

Regarding the highest visual acuity achievable by humans, there seems to be little reliable information available, as medical assessments of visual acuity often use charts that do not extend beyond a certain level (Elliott et al., 1995). Ghose (2015) claims that a score of 20/7 (logMAR -0.45, or a resolved detail of 0.35 minutes of arc) may be the highest achievable by humans with optimal predispositions and extensive training. Durrie et al. (2019) claims that

about 1% of healthy adults have 20/10 vision (logMAR -0.30, or a resolved detail of 0.5 minutes of arc), and such high acuity can be achieved by individuals who were treated with laser eye surgery for the purpose of correcting refractive aberration.

3.1.3. Conclusions from the review of anatomical and perceptual foundations

The following key conclusions can be drawn from the review of biological foundations presented above:

- The human visual system is not a sampling device of homogeneous resolution. The density and sensitivity of photoreceptors varies widely across the visual field, and facilitates highest acuity only in the fovea centralis at the centre of the visual field.
- A common estimate for the maximum resolution capability of humans is one minute of arc of apparent size, which is also commonly used as a threshold for good visual acuity, referred to as "20/20 vision". However, both the density of photoreceptors in the fovea, as well as other optical properties of the human eye, suggest a value of ½ minute as a better approximation of maximum resolution capability. For some specific visual stimuli, the displacement of graphical elements can be sensed with an accuracy that would correspond with much higher binary sampling resolutions, at about 1/10th of a minute.
- Visual acuity decreases with diminishing contrast. However, the contrast required for maximum performance is believed to lie well below the contrast ratios that can be generated by modern displays in indoor settings.
- Examination of visual acuity in medical settings often focusses on identifying reduced acuity, below the "20/20" threshold; historical medical examinations did not reliably assess or report the maximum acuity of each subject. Therefore, assessments of visual acuity across populations may not be reliable. Also, medical examinations often report either the uncorrected acuity of a subject, or the acuity accomplished with optimal correction. Reliable data on "habitual acuity" of human subjects, i.e. the acuity they are able to accomplish in their daily lives, is hard to find.
- Psychophysical studies with the goal to assess perceptual thresholds often use a threshold performance half way between perfect recognition and the chance of random guessing (e.g. 75% for a task with two possible answers). Such thresholds are well suited for estimating the psychometric function modelling the relationship between stimulus level and recognition rate, but are well below the reliability we would expect for cartographic communication.

3.2. Digital display technology

"We recommend screens with the highest possible resolution […]. This drastically improves readability, which is particularly important when working at a screen all day."
EIZO guide: The right monitor for your home office[6]

Over the past 20 years display technology has progressed in multiple areas, each of which could legitimately be qualified as a "revolution". Around the turn of the millennium, cathode ray tube (CRT) based displays were the dominating technology for desktop monitors and TVs. Laptops could already be equipped with liquid crystal displays (LCDs) capable of colour reproduction, but such screens were small and of low resolution[7]. Mobile phones had tiny, monochrome screens with coarse resolution and offered only rudimentary capabilities to display online content or run applications installed by the user[8].

Studies on cartographic symbology for screen-based maps undertaken around that time (Malić, 1998; Birsak, 2000; Neudeck, 2001) therefore exclusively focused on CRT monitors, which were usually used at resolutions roughly corresponding with display size, capable of pixel densities of around 58–127 ppi, with most common configurations at densities of 84–100 ppi (Malić, 1998, p. 63). Besides the limited pixel density, CRTs had a number of other disadvantages affecting the flexibility of use and clarity of image: they were bulky and heavy, required a warm-up period before being able to display colours accurately and uniformly, and had low contrast ratio and colour accuracy (Metha et al., 1993).

Shortly after the millennium, the technology and production capacity for LCDs were developed enough to allow for a rapid transition of TVs and desktop monitors to this newer technology. Not image quality, but mainly the aspects of decreased size and weight were the drivers for this transition, which for the most part happened over a period of a few years. A Swedish study shows that in only four years of 2003 to 2007, sales CRT-based domestic TVs in Sweden went from above 95% of market share to under 5% (Kalmykova et al., 2015). Detailed sales data for computer monitors was not available for their study, but can be assumed to have happened in a similarly dramatic fashion, potentially over a longer timespan due to longer lifecycles of computer monitors, and the specific requirements of computing applications (ibid.).

---

[6] Source: https://www.eizo.eu/knowledge/workstation-ergonomics/the-right-monitor-for-your-home-office/ accessed 2021-01-22

[7] Source: https://en.wikipedia.org/wiki/History_of_laptops, accessed 2022-04-26

[8] Source: https://en.wikipedia.org/wiki/History_of_mobile_phones, accessed 2022-04-26
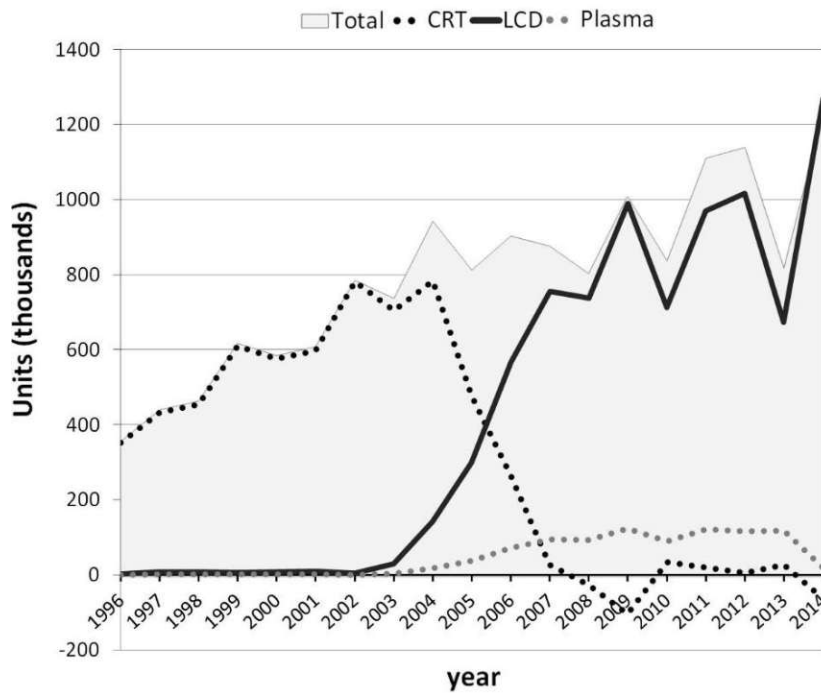
Figure 3.7 TV imports for Sweden, by technology. The years 2003-2007 saw a rapid transition from CRT to LCD technology. (Negative sales numbers for CRT monitors in 2008-9 indicate surplus inventory shipped back to other countries) Image source: CC BY-NC-ND 4.0 Kalmykova et al. (2015)

The second recent revolution in the realm of digital displays started appearing in public in January 2007, when Apple introduced the first iPhone. While mobile phones and "personal digital assistants" with data connectivity, keyboards and larger displays had been available before its release, the combination of a larger screen with twice the resolution of most competing devices at the time, a multi-touch capacitive touch display, and the by then general availability of mobile phone subscriptions providing mobile internet access and data transfer, created the conditions for widespread public adoption of a new paradigm in mobile media: the smartphone[9]. Other competitors released competing models soon after Apple's iPhone, and the large number of devices sold created the economical preconditions for larger investments in technological development and production capacity for smartphone displays (Nguyen, 2021).

In recent years another technological transition in the realm of digital displays could be witnessed, with the commercial availability of devices based on organic light-emitting diodes (OLED) instead of LCD technology. While not "revolutionary" in the sense of enabling completely new devices or usage scenario, OLEDs allow for more flexible form factors and subpixel geometries, and produces superior contrast ratios in some settings. Other display technologies, such as "electronic ink" displays, which can be read in bright sunlight and do not consume power while displaying a static image, had existed since before the introduction of the smartphone, but also benefitted from the changed market conditions and technological ecosystem. By the end of 2007, Amazon released its first Kindle e-book-reader device, which did not require a separate paid subscription for receiving updated content.

---

[9] Source: https://en.wikipedia.org/wiki/Smartphone, accessed 2022-05-02

3.2.1. Display hardware technologies

This section will discuss display technologies of current primary relevance for smartphone displays in more details. Outdated technologies such as CRT displays, or niche technologies such as e-Ink displays, are not covered in detail, since they were determined to be outside the scope of the studies presented in this thesis, which focuses on currently available smartphones.

Chen et al. (2018) list eight main metrics for assessing the performance of digital displays: resolution density, contrast ratio, peak brightness, colour gamut, viewing angle, response time, panel flexibility and lifetime. Of these properties, the first five affect the visual quality of (static) images and are therefore of primary interest for studies of graphical reproduction quality. To succeed in the marketplace, economic factors such as manufacturing cost and marketability also need to be taken into account – the recent history of digital displays outlined above illustrates that market forces can be a powerful driver of research and development, and do not merely passively follow technological progress.

3.2.1.1. Liquid crystal displays (LCDs)

A liquid crystal display (LCD) is based on the property of particular materials – liquid crystals – to change the phase of polarized light depending on the presence of electrical current in the material. When combined with polarizing filters, light can be blocked or passed through an element depending on the charge applied. For producing arbitrary colour output by additive mixing of primary colours, three such elements can be combined with red, green and blue colour filters to form a single pixel.

Liquid crystals do not emit light, therefore a homogeneous backlight is needed for LCDs when used as basis for pixel-based output. The backlight can be provided by a strip light source (such as a cathode fluorescent lamps or an LED strip) combined with a light guiding plate that distributes the light evenly across the area of the display, or a light source that covers the entire area of the display such as an array of white LEDs. If an array of controllable LEDs is not used as a backlight, the backlight will remain on at constant intensity for as long as the display is in use, resulting in continuous power consumption independent of the brightness of the actual content on the screen – even when showing a totally black screen, a conventional LED backlight will be fully switched on.

Because the liquid crystal layer is located in front of the backlight, any electronics controlling the liquid crystal elements will also need to be located in front of the backlight. For early LCDs, this resulted in a relatively large proportion of the overall screen area to be obstructed by the necessary electronics and the electrodes for the individual subpixels, resulting in low contrast ratio, reduced viewing angle and sensitivity to touch or pressure on the display (Chen et al., 2018). With thin-film transistors (TFT), the footprint of the electronics could be reduced, resulting in increased contrast and viewing angle. In-plane-switching (IPS) and fringe-field-switching (FFS) technology allows for the placement of both electrodes on the back side of the liquid crystal, resulting in better colour reproduction and robustness against force applied to the display surface (such as when touching the display in a touchscreen application) (ibid.).

3.2.1.2. Organic light-emitting diode (OLED) displays

In contrast to LCDs, which selectively block light provided by a backlight, in OLED devices the emission of light is directly regulated for each pixel of the display. In OLEDs, one or multiple layers of organic semiconductor material are located between two electrode layers, of which one is transparent. As voltage is applied to the electrodes, electrons and "electron holes" (chemical structures with a missing electron) move through the layers and recombine in the emissive layer, causing the emission of visible light (Chen et al., 2018). OLED displays also use thin film transistors (TFTs) for controlling individual pixels, thus the usage of this acronym to exclusively refer to LCD technology is misleading. Because OLEDs are themselves a light source, the necessary electronic components can be arranged behind or between the subpixels in such displays, affording high contrast ratios.

OLED displays have a much simpler layer structure than LCDs (Tsujimura, 2017), and the total thickness of its active components can be <1 μm, making it the perfect candidate for flexible displays (Chen et al., 2018). Because only as much light as needed is emitted, the power consumption of OLED displays varies with the overall brightness of the presented image, and a darker user interface design can be a viable power conservation strategy (Dash & Hu, 2021). However, power efficiency of OLEDs at full brightness is lower than of a comparable LCD-based display. As of 2018, the point of equal power consumption of state-of-the-art OLED displays and LCDs was at ~65% maximum brightness, with ongoing technological progress promising further improvements in OLED efficiency (Chen et al., 2018).

In contrast to LCDs which cannot fully block the light emitted from its backlight, OLED displays can produce a "true black" output image not emitting any light at all. In theory this results in a contrast ratio of "infinity", because the darkest pixels have a light emission of zero. However, this is only achievable in a fully dark environment in which no ambient light is reflected from the dark parts of the image. Chen et al. show that in practice LCDs and OLED-based displays have a very similar contrast ratio in any environment brighter than 100 lux ambient light (corresponding to a dimly lit living room) (Chen et al., 2017). For outdoor usage, the maximum brightness of the display is more relevant, where LCDs and OLEDs can also reach similar levels.

3.2.2. Display resolution, pixel density & subpixel arrangement

> *"Pixels are a surprisingly complex topic on the web"*
> *("Surma" on web.dev)[10]*

The term "resolution" has been used ambiguously in the context of digital displays. Historically, screen resolution often referred to the overall number of pixels that defined the full screen image – e.g. 1024×768 pixels for the standardized *XGA* resolution[11]. CRTs did not have a fixed resolution on the hardware level, but could display video signals within a range of resolutions, ultimately limited in one or both dimensions by the density of the dot or strip mask (Malić, 1998). Absolute pixel count was particularly relevant because, together with

---

[10] https://web.dev/device-pixel-content-box/
[11] https://en.wikipedia.org/wiki/Graphics_display_resolution, accessed 2022-05-02

colour depth (the number of bits per pixel used for storing colour information), it would determine the memory requirements for a given display configuration, as well as the CPU time required to process those pixels. Up until the early years of the new millennium the main limiting factor for pixel counts (and/or colour depth) was usually not the display device, but the memory capacity of the computer or graphics card supplying the video signal to the display. Therefore, absolute pixel count numbers of graphical resolution were much more important for estimating the system requirements for producing a particular graphical output than the density of physical pixels on the screen, which was generally limited within a range of about 80 to 100 ppi for most CRTs (ibid.). Neudeck (2001) commented that pixel densities of display devices available around the millennium "differed only insignificantly"[12]. However, in recent years, constraints of memory and computational power have become less relevant, and, as has been discussed above, devices with very diverse pixel densities have become available. Therefore, in this thesis, the term "resolution" will generally be used synonymous to "pixel density" – referring to the number of pixels within a given extent of the display, usually specified as pixels per inch (ppi). If the absolute number of pixels on the display is referred to (such as "1024×768 pixels"), the term "pixel count" will be used to avoid confusion.

The mobile devices that existed at the time of early studies on screen-based cartography were far away from reaching parity with desktop computers in terms of visual quality, because of limitations of display hardware, memory and processor speed. For applications requiring graphical quality, there was therefore mainly the desktop computer scenario to consider. For desktop viewing, a pixel density of 96 ppi was, for example, assumed in Microsoft Windows as a standard value (Hitchcock, 2005), with any significantly higher resolutions used for special applications being dealt with on an application level.

With the advent of the smartphone, a new viewing configuration was introduced into the mainstream of digital device usage. Smartphones are viewed from closer distance than desktop monitors (with a nominal viewing distance of 30cm instead of 60cm), requiring twice the pixel density to achieve an impression of equal fidelity. The first iPhone was released in 2007 with a pixel density of 163 ppi, thus corresponding to a desktop monitor with ~82 ppi, at the lower end of the range of common pixel densities. Absolute pixel count of the iPhone's display was 480×320 pixels, much lower than common resolutions of desktop or laptop monitors at the time[13]. But in a world with increasingly heterogeneous configurations of display sizes, pixel counts and viewing distances, absolute pixel counts alone were becoming less meaningful. Information about pixel density is a much more reliable indicator for image quality, if an assumption of typical viewing distance can be made.

At a pixel density of 163 ppi and a viewing distance of 30 cm, the pixels of an LCD screen and the gaps between them are still discernible with the naked eye. With the release of the iPhone 4, Apple introduced the "retina display", which doubled the pixel density to 326 ppi, while keeping the display size identical to its predecessor[14] (effectively quadrupling the number of

---

[12] Own translation. Original: "da sich die Bildpunktgrößen zur Zeit nur unwesentlich voneinander unterscheiden" (Neudeck, 2001, p. 46)

[13] Source: https://en.wikipedia.org/wiki/IPhone_(1st_generation) , accessed 2022-05-13

[14] Source: https://en.wikipedia.org/wiki/IPhone_4 , accessed 2022-05-13

pixels on the display). The marketing term "retina" suggested that such resolution, which at a distance of 12 inch / 305 mm corresponded to 57 seconds of arc per pixel (≈ 1 minute of arc), approached the limits of human visual acuity. It has been shown in Section 3.1 that there is no single number that unambiguously quantifies human visual acuity, and that the one-arcminute model is usually used as a lower bound for normal acuity rather than its absolute upper limit. Consequently, as technological progress allowed it, manufacturers released devices with even higher pixel densities after the iPhone 4. Also, screen sizes for mobile devices became more heterogeneous with the widespread availability of tablets and larger phones ("phablets").

Maximum pixel densities of smartphone displays steadily increased in the years 2012–2015. In that year, Sony released the Xperia Z5 Premium smartphone with an advertised pixel density of 806 ppi (corresponding to an apparent pixel size of 0.36 minutes of arc). The ultra-high resolution of this phone was not made available by default to regular applications, to which it appeared as a phone of half resolution at 403ppi; a setting that made available the full resolution to all applications had to be activated in "developer mode" (GSM Arena, 2015). This may be indicative of the manufacturer themselves doubting the utility of such extreme resolutions for normal applications. A predecessor with identical resolution, the Xperia XZ Premium, was released in 2017. After that, no manufacturer released a phone with matching or higher pixel density (as of 2022). With the exception of a few outliers, pixel densities of most smartphone screens have been in a range of ~250 to ~580 ppi from 2014 to today. A visualisation of the evolution of pixel densities of smartphones has already been provided in Figure 1.1 in the introduction.

A similar development can be observed for desktop and laptop monitors, at lower absolute pixel densities corresponding with a larger typical viewing distance. While pixel density for desktop LCD monitors was limited to about 100 ppi until about 2012 (with the exception of outliers for special applications), devices with double or nearly triple such pixel density have become available in recent years. As of today, the Dell UP3218K, introduced in 2017, is the monitor with the highest pixel density available commercially at 280 ppi, with an apparent pixel size of 0.5 minutes of arc at 60cm viewing distance. In terms of apparent pixel size at typical viewing distance, this is significantly larger than the pixels of the 806 ppi-display of the Xperia Z5 Premium, and more in line with the upper end of mainstream smartphones at 580 ppi. It is too early yet to tell whether development of desktop monitors has also plateaued at this resolution, but the correspondence with maximum resolutions the smartphone market has recently settled on would suggest so. Figure 3.8 shows the evolution of screen resolutions of desktop and laptop screens, for the time period of 1996–2018, for which data was found to be available.
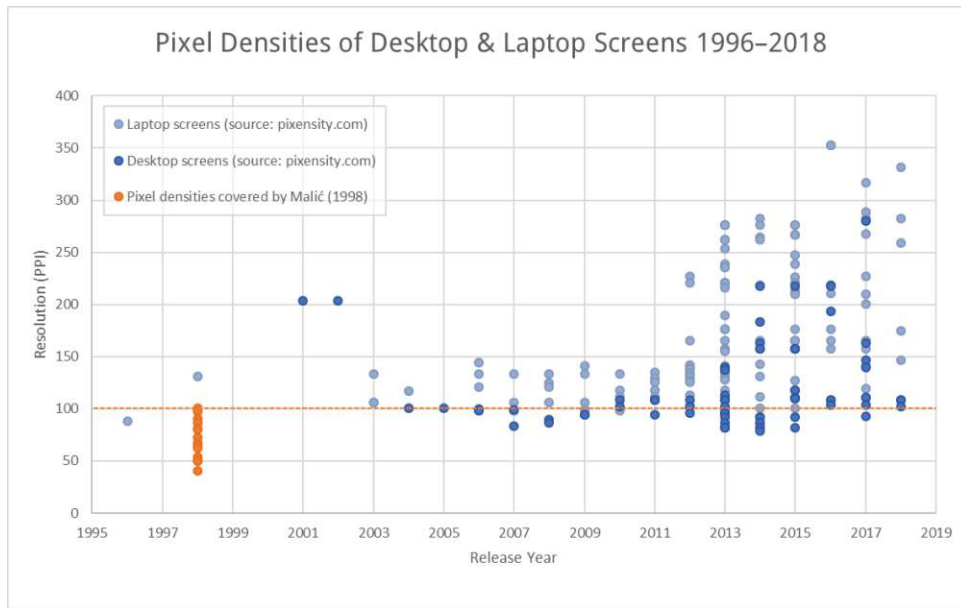
Figure 3.8 Evolution of desktop screen resolutions, 1996-2020. Data source: https://pixensity.com/list/desktop/ (accessed 2021-02-03)

As of today, researchers have demonstrated prototypes of both LCDs and OLED displays which surpass several *thousands* of pixels per inch (Chen et al., 2018; Fujii et al., 2018). However, the smartphone market has ceased to be the driving economic force behind further increase of pixel densities, as such resolutions are far beyond what can be sensibly made useful in typical viewing situations. Today, applications for VR and AR headsets are the drivers of display development to ever higher pixel densities, where higher resolutions and further miniaturization are advantageous to meet the optical and graphical requirements of these applications (Chen et al., 2018).

3.2.2.1. Subpixel arrangement

None of the display technologies discussed in the previous sections can directly produce light of arbitrary wavelength. The appearance of colour is created in digital displays by additive mixing of coloured, but monochromatic, subpixels. CRTs and many LCDs and OLED displays use one red, green and blue (RGB) subpixel for each display pixel. In conventional LCDs, these were arranged as vertical stripes, each covering about 1/3rd of the width of a square pixel (see Figure 3.9).
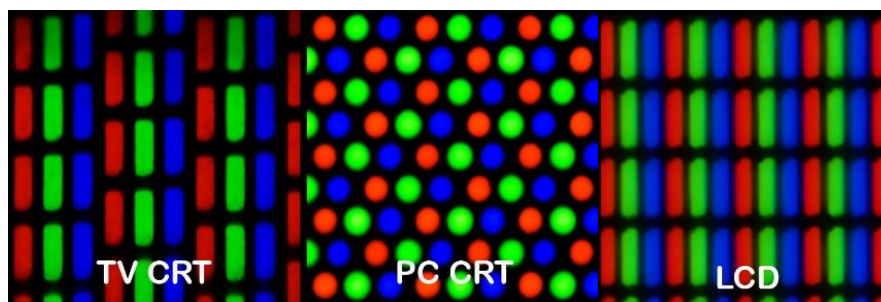

Figure 3.9 RGB Subpixel arrangements of conventional CRT and LCD displays. Image source: CC BY-SA 3.0 by Pengo on Wikimedia

Technological advancement in the production of LCDs and OLED displays allowed for different, more flexible arrangements of subpixels. Due to the asymmetrical sensitivity curves and uneven distribution of the different photoreceptors in the retina, the sensitivity of the fovea is highest in the green-yellow part of the spectrum. This is exploited by the subpixel arrangement of some modern displays, which may contain a larger number of green subpixels than red and/or blue subpixels, to make use of the higher visual acuity at green wavelengths (Credelle & Brown Elliott, 2005). Adding white subpixels for a "RGBW" arrangement can help reduce power consumption (Tsujimura, 2017, p. 138) and increase apparent horizontal resolution (Clairvoyante, 2010). Samsung markets displays with such subpixel arrangements under the PenTile™ trade name. Byford (2021) claims that "almost all OLED screens in portable consumer devices use some form of PenTile subpixel layout these days" and that the scheme has been used by the industry to claim higher pixel densities than actually present on the screen (by counting only the subpixels with the highest densities or using "overlapping" pixel boundaries).
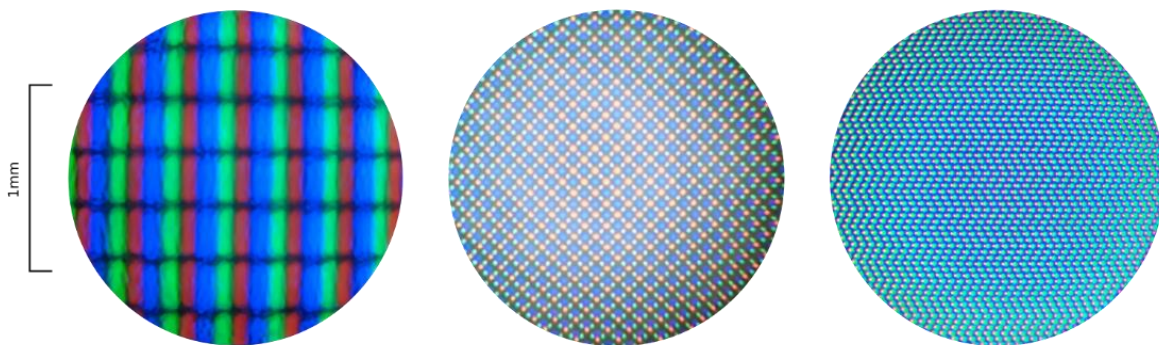

Figure 3.10 Subpixel arrangements of various displays, viewed through a microscope. Left: Desktop LCD monitor (86ppi) with conventional BGR subpixel arrangement; Centre: Display of Google Pixel2 smartphone (520ppi) with 1:2:2 BGR subpixel ratio, arranged in a diamond pattern; Right: Display of Sony Xperia Z5 Premium smartphone (801ppi) with BGR subpixels arranged in a zigzag pattern.

With unconventional subpixel layouts, the concept of display hardware resolution became somewhat fluid and disputed. To accommodate a wide range of display technologies, the Video Electronics Standards Association (VESA) defines resolution not by counting (sub)pixels and dividing by length, but by a display's capability to render line pairs with a minimum of 50% Michelson contrast (VESA Display Metrology Committee, 2001). According to such measurement, a resolution can be claimed for displays with non-RGB subpixel structures that does not result in each "pixel" having a red and/or blue subpixel, but is composed of only two subpixels (out of four in a RGBW display) (Clairvoyante, 2010).

3.2.2.2. Software representation of 2D screen graphics

In early raster-scan computer graphics systems, screen contents were represented by a continuous range of memory, the frame buffer. For each screen refresh, the video controller would read binary values stored in the frame buffer, and emit a video signal to produce the corresponding graphics on the screen. Since raster-scan CRT monitors projected the sequence

of pixels line by line, from top to bottom, this was also how graphics data was usually stored in memory. For monochrome bitmap graphics, in which pixels can have only one of two colours (black/white or a configurable foreground/background), only one bit of data was needed to represent each pixel. Systems with more colours used multiple bits per pixel and a colour lookup table to choose colours from, or one byte for each colour channel (red, green and blue) in "true colour" systems (Hearn & Baker, 1996).

With the advent of window-based operating systems, the primitive model of a single continuous memory area to act as the main interface to control the computer's graphical output was replaced by a partitioned model, in which each application could control only its own assigned part of the overall screen content, and the operating system would compose the overall screen image from these components. Specialized graphics cards started to provide hardware accelerated routines for rendering 2D and 3D graphics, and these cards were controlled by a software driver that offered an API to applications, which could control the card by calling functions of the API and passing graphics data as coordinates or bitmaps stored in memory. For 2-dimensional drawing, bitmap graphics, and windowing operations, the screen pixel remained the fundamental unit that structured screen coordinates and memory representation. Pixel counts of graphical displays increased gradually, from 640×480 to 800×600 and then to 1024×768 as typical resolution settings throughout the 1990s, but such improvements were met by larger monitors and users' continued demand for graphics of higher fidelity. As discussed above, the size of a typical pixel remained fairly stable within some boundaries in mainstream computing hardware until the second decade of the 21st century, and when using a CRT users could choose from a set of supported screen resolutions to match their preference or needs. Pixels were a convenient, well-known and computationally feasible unit to work with, despite their somewhat uncertain real-word size.

When Apple introduced the iPhone 4 with its "retina display", a sudden jump in resolution by a factor of two was introduced for the first time in the history of mainstream computing, without the display getting any bigger or the user being able to adjust the resolution downwards. To avoid the need for software to be adapted or rewritten, or existing user interfaces appearing too small on the higher resolution display, the concept of *logical resolution* was applied: on the software side, the pixel count would appear as identical to that of earlier models at 480 × 320 pixels, and only be converted to the higher resolution internally by the operating system. Each logical pixel was therefore mapped to 2×2 physical pixels, decoupling hardware resolution and software representation (Diaz, 2010). Application programmers could simply work with the lower resolution in their programs, and would benefit automatically from higher fidelity for text and vector graphics; bitmap graphics could be supplied in a separate "2x" version for the higher resolution devices, or were scaled automatically from the lower-resolution version (Apple Inc., 2012b).

The concept of a multiplying factor continued to be adopted for devices of higher pixel densities, with "3x" devices, corresponding to a pixel density of ~480 ppi being introduced in the market soon after the iPhone 4 (Android Open Source Project, 2021). Gradually, this resulted in a "big mess" (Muehlenhaus, 2014, see Section 3.3.3.1), with devices of widely diverging form factors and pixel densities being in active use, without giving content creators a reliable way of exactly querying the real-world size of the screen, and therefore scaling the

graphics to precise dimensions. Content creators had to rely on the pixel multiplier factors or device categories supplied by the manufacturer to have user interfaces and graphics roughly align with intended dimensions, without precise control (Android Open Source Project, 2021).

The Android operating system today differentiates 7 classes of mobile devices, with corresponding pixel density and logical-to-screen-pixel multiplier values (see Table 3.1). So called medium-DPI devices are still considered representing a baseline, at around 160 ppi – the pixel density of the original iPhone. It needs to be noted that mobile phones at such low resolutions are, as of 2022, not commercially available any more. The pixel multipliers of devices with higher resolution are not constrained to integer factors any more, with a multiplier of 0.75 for low-density devices and 1.5 for high-density devices (~240 ppi). Such non-integer pixel multipliers abandon the idea of a mapping of software coordinates or bitmaps to discrete physical pixels altogether. In such configuration, even integer coordinates are not guaranteed to fall precisely on the boundaries of a hardware pixel any more. Vendors have configured phones with non-integer pixel ratios by default – for example, the "Google Pixel" phone, released in 2016, had a 441 ppi display[15] and was configured with a default pixel multiplier[16] of 2.6, for an effective logical resolution of ~170 ppi[17].

Pixel density class	Pixel Multiplier	Description
ldpi	0.75	Low-density (ldpi) screens (~120 ppi)
mdpi	1.0	Medium-density (mdpi) screens (~160 ppi) (baseline density)
hdpi	1.5	High-density (hdpi) screens (~240 ppi)
xhdpi	2.0	Extra-high-density (xhdpi) screens (~320 ppi)
xxhdpi	3.0	Extra-extra-high-density (xxhdpi) screens (~480 ppi)
xxxhdpi	4.0	Extra-extra-extra-high-density (xxxhdpi) uses (~640 ppi)
tvdpi		"Somewhere between mdpi and hdpi; approximately 213dpi. This is not considered a 'primary' density group. It is mostly intended for televisions and most apps shouldn't need it"

Table 3.1 Android pixel density classes and pixel multipliers. Source: Android Open Source Project (2021).

The introduction of non-integer arbitrary pixel multipliers allows for devices of any size and physical resolution to present themselves as having a logical resolution at or near the reference resolution of 160 ppi on the software side. Furthermore, by making the pixel multiplier user-configurable, the mechanism could be used to let users adapt the size of content on their device to their own individual preferences. On Android phones, users can influence the pixel multiplier with the "Display Size" setting in the settings app, and can set their preferred font size using a second, independent setting (Google Inc., 2021). Apple has stuck with integer multipliers in its iOS devices[18] (Apple Inc., 2022) (which may reflect a more closely controlled range of hardware products) and does not allow users to override the pixel

[15] Source: https://www.gsmarena.com/google_pixel-8346.php , accessed 2022-06-01
[16] Source: https://blisk.io/devices/details/google-pixel , accessed 2022-06-01
[17] Interestingly, this deviates from the 160 ppi specified by Google as the reference resolution, for which a pixel multiplier of 2.75 would have been the more appropriate value.
[18] Also see https://www.ios-resolution.com/ for an overview of devices, resolutions and scale factors

multiplier, but allows adjusting the preferred text size and offers a zooming functionality for accessibility purposes[19]. Windows 10 allows adjusting the display in discrete steps in an overall range of 100%-500% and automatically rescales some application dialogs and controls to match the configured value, but requires application authors to explicitly query the configured pixel density for the monitor it is running on (Microsoft Inc., 2021).

Separating the pixel as a measurement unit from the physical organization of the display has allowed operating systems to run applications designed for a particular resolution on systems with a wide range of pixel densities, and let end users control the size of on-screen user interface elements according to their preferences and needs, without requiring application authors to change their code or explicitly deal with such a heterogeneous landscape of hardware configurations. However, as will be argued in Section 3.2.3, in some cases it is necessary for software to precisely control the content of physical pixels. Therefore, some means for controlling actual hardware pixels is still required, and provided on most systems. On Android devices, the unit `px` refers to *hardware pixels*, while the unit `dp` refers to *density-independent pixels* (corresponding to one pixel on an 160ppi device), and `sp` refers to *scalable pixels*, taking into account the user's preferred font size (Android Open Source Project, 2021). Apple's iOS differentiates between hardware *pixels* and user-space *points* (Apple Inc., 2012a). In Windows, application authors can query the pixel density set for a display using various API calls, but need to make the necessary calculations and adaptions themselves (Microsoft Inc., 2021). None of the common operating systems exposes information regarding the display's subpixel structure (RGB subpixel order, PenTile arrangements etc.) to applications.

The pixel, represented by the CSS unit `px`, has also long been a fundamental unit for specifying the dimension of content elements on the web. In the heterogeneous landscape of display devices and viewing situations used to access web content, not only the pixel density of the display, but also the viewing distance varies significantly across users. The difference in typical viewing distance between mobile (~30cm), laptop (~45cm) and desktop (~60cm) viewing configurations means that apparent sizes of graphical elements of identical physical extent vary by a factor of up to two. This is partly reflected by the higher nominal "standard" resolution of 160 ppi for mobile devices (compared to the 80-100 ppi common for conventional desktop monitors), but inconsistent device configuration and user configurability of display scaling factors aggravate the uncertainty of expected dimensions for pixel-based measurements for web content. Reflecting the relevance of the *apparent* size of graphical elements, the CSS specification, starting with version 2.1, has defined the unit `px` – the *CSS pixel* – as extending across a visual angle of 1/96th of an inch at a distance of 28 inches (~71cm), about 1.3 minutes of arc (World Wide Web Consortium, 2011) (see Figure 3.11). This models the size of a pixel on a conventional screen at 96 ppi at a distance of a typical arm's length, or a proportionally smaller pixel at closer viewing distances. Browser vendors are expected to configure the web browser in such a way that the CSS pixel unit is

---

[19] https://support.apple.com/guide/iphone/display-text-size-iph3e2e1fb0/ios ,
https://support.apple.com/guide/iphone/zoom-iph3e2e367e/ios

appropriately converted to physical pixels for the device and expected viewing situation (ibid.).
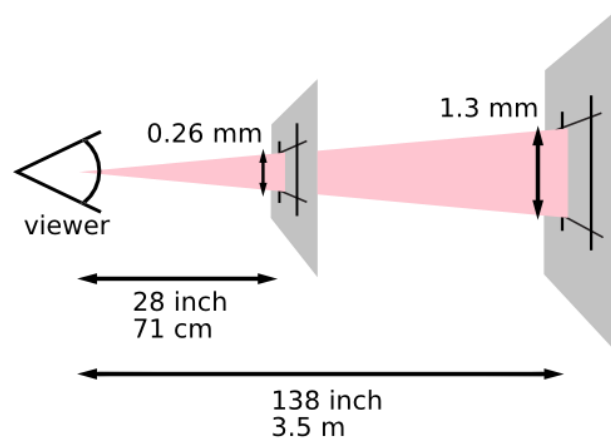

Figure 3.11 Geometrical model of the CSS pixel. Image source: World Wide Web Consortium (2019)

The CSS pixel provides a pragmatic solution for content creators to specify dimensions of graphical elements with a mental model of conventional desktop and mobile screen pixels, without the need to convert such dimensions to actual physical pixels. For graphical applications that require detailed control over physical pixels, information about the underlying display is provided by modern browsers. In CSS, media queries can be used to adapt content to devices matching certain pixel densities, specified either as pixels per inch (`dpi` unit), pixels per centimetre (`dpcm` unit) or as a CSS-pixel-to-screen-pixel multiplier (`dppx` or `x` unit) (World Wide Web Consortium, 2020a). In JavaScript, the pixel multiplier can be queried with the DOM property `window.devicePixelRatio` (World Wide Web Consortium, 2016), and the dimension of an element on the page can be accessed using the `devicePixelContentBoxSize` property (World Wide Web Consortium, 2020b), allowing the precise control of physical pixels with an HTML `canvas` element.

3.2.3. Rendering for digital displays

In computer graphics, the term *rendering* refers to the computations performed for transforming a model of graphical content to actual coloured pixels on the screen. In past decades, much effort has been put into improving efficiency and quality of rendering of three-dimensional scenes (Hearn & Baker, 1996), but also 2D graphics is rarely created by writing directly to the pixel buffer in modern operating systems and may involve complex rendering operations (Mileff & Dudra, 2012). Modern high-level formats for 2-dimensional content such as HTML+CSS or SVG may require significant processing in multiple stages to result in a screen image matching the specification (The Chromium Project, 2009)[20].

While early computer systems allowed applications to directly write to the pixel buffer to define the visual output, today rendering almost always involves a *rendering pipeline* in which

---

[20] The term "rendering" is also sometimes used to refer to the transformation of a proprietary content model to a well-known other content model, such as the "rendering" of HTML code in a web application framework. In this document I will use the term rendering to refer exclusively to the generation of pixel-based output.

a content model is transformed in multiple steps to the desired output pixels, potentially involving multiple software layers (such as the application, graphics libraries and the operating system) or hardware components (such the CPU and dedicated graphical processing units (GPUs) of a graphics hardware subsystem). The multi-step pipeline model aligns with a conceptual model of cartography as a sequence of transformations (Tobler, 1979; Cauvin et al., 2010, Chapter 2) which has led to detailed conceptualizations of a cartographic rendering pipeline, starting with geographic data stored in a primary model, involving potentially multiple intermediate steps of generalisation and stylization, ultimately resulting in a digital image (Semmo et al., 2015). The very last transformation step in such a pipeline is the *display transformation* in which geometric information is transformed to actual pixel values at a specific location on the screen (Cauvin et al., 2010, p. 56). This final transformation is often not at the focus of attention of cartographers and is usually taken care of by software or hardware beyond the cartographer's immediate control – the operating system, a web browser or special-purpose hardware components. However, the display transformation is what sets apart digital maps from earlier map production techniques, in which the other transformation steps were also required (Cauvin et al., 2010). It therefore warrants closer investigation how lines and shapes are transformed to pixel values, particularly in any project interested in details of how the rendered content is perceived by a human observer.

3.2.3.1. Aliasing and anti-aliasing

In computer graphics, aliasing refers to artefacts caused by a mismatch between the spatial resolution of rendered content and the pixel grid of the display. For bitmap graphics, in which pixels can either be fully on or off (or can be set to one colour value out of a limited palette of colours), for any graphical content that does not exactly align with pixel boundaries, aliasing will lead to inaccurate reproduction on screen (Hearn & Baker, 1996). For example, a simple straight line between two points on the screen, if not exactly parallel to the displays x- or y-axis, is a continuous phenomenon that includes an infinite number of points not located at pixel coordinates between its start and end points. Rendering such graphics to the screen will always incur aliasing, because the geometry with infinite precision needs to be represented by discrete pixels. This leads to graphics with a clearly visible "staircase effect" and omission or enlargement of small gaps in the graphics on displays with conventional pixel densities (Neudeck, 2001). Figure 3.12, left, illustrates the staircase effect of aliasing for a simple black line.

Early graphics systems were limited to bitmap graphics with few colours and integer pixel coordinates, and floating-point operations were not supported by all hardware platforms. Drawing algorithms such as Bresenham's line drawing algorithm focused on efficient drawing of graphical primitives to a bitmap pixel buffer using only integer calculations (Hearn & Baker, 1996).

The availability of increased colour depth facilitates the main idea of *anti-aliasing* approaches: to substitute spatial resolution with (colour) intensity resolution. One common approach to antialiasing is supersampling: The graphics is rendered to a bitmap of higher resolution internally, using multiple cells to represent each screen pixel. Each screen pixel is then coloured according to the number of cells covered by the graphics – black for pixels of which

all cells are black, white for all-white cells, and shades of grey in between selected in proportion to ratio of black-to-white cells. On displays of low pixel density, anti-aliasing can significantly enhance the apparent fidelity and legibility of the overall image (Neudeck, 2001). Figure 3.12, right, shows the effect of antialiasing, resulting in a much smoother appearance of the black line, despite being rendered at identical spatial resolution as the aliased example.

Figure 3.12 Aliasing and antialiasing. Left: rendering a line at an oblique angle as a bitmap image results in a jagged appearance (aliasing). Right: Antialiasing substitutes spatial resolution with intensity resolution, colouring partially-covered pixels in shades of grey, which results in a smoother appearance at equal display resolution.

Antialiasing can produce significant enhancement of subjective quality of the resulting images, with the cost of decreasing the crispness of high-contrast boundaries in the image. Particularly horizontal or vertical lines with a width of one pixel can appear quite differently depending on their alignment with the pixel grid – when the boundaries of the line are exactly aligned with the pixel grid, the line will be rendered one pixel wide at maximum intensity, while a shift of ½ pixel will result in a width of 2 pixels (a doubling of size) at half intensity (Neudeck, 2001, p. 53). While applications which use integer pixel coordinates do not suffer from this problem for lines in alignment with the pixel grid's axes, in any applications with graphical coordinates with arbitrary fractional components (such as cartography, where pixel coordinates often result from the cartographic transformations applied to the geographic data) and small geometries, this leads to an overall reduction of sharpness (ibid.).

Interestingly, the principle of antialiasing works as a kind of inverse to visual hyperacuity (discussed in Section 3.1.1.2): a limited spatial resolution is compensated by substituting higher intensity resolution. Consequently, hyperacuity of human observers can be observed also for stimuli rendered with anti-aliasing, for apparent acuity far beyond the physical resolution of the pixel grid (Ware, 2020, p. 65). The use of antialiased rendering is accepted and well established for medical and psychophysical examinations of acuity (Bach, 1997; Ruda, 2013). Bach claims that antialiasing can be used for stimuli in which the critical dimension (e.g. the opening of a Landolt-C or the stroke/gap with of Snellen letters) is equal to or larger than one physical pixel. However, due to the enlargement effect discussed above, this seems a dubious claim, as a gap slightly larger than a single pixel – or even a single-pixel gap not aligned with the pixel grid – would appear as a two-pixel gap (admittedly of lower contrast). Figure 3.13 demonstrates that a Snellen E can exhibit features that are significantly larger than the critical dimension when rendered at small pixel sizes with antialiasing.

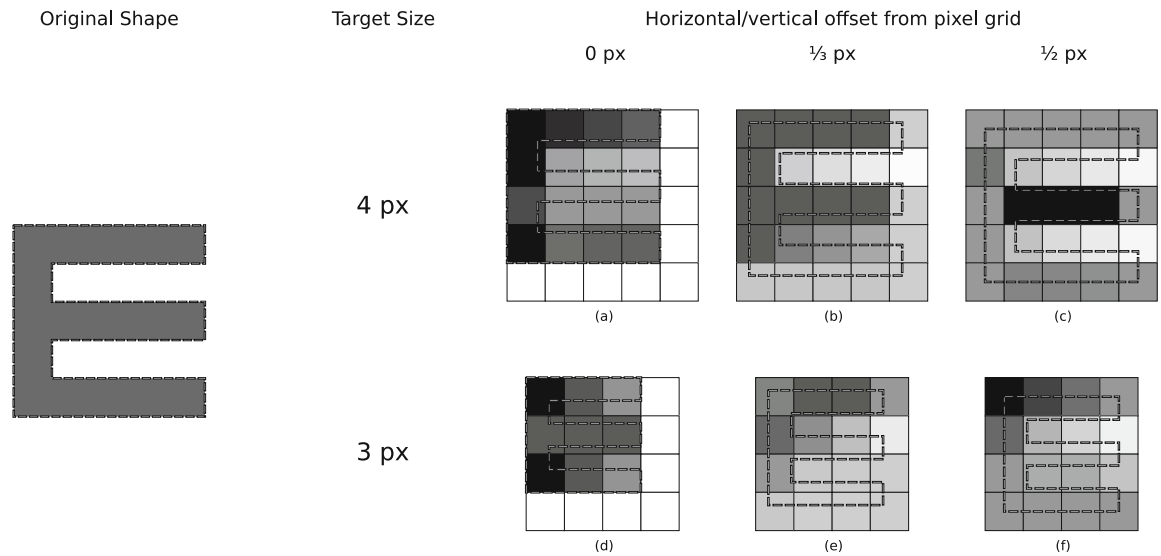| Original Shape	Target Size	Horizontal/vertical offset from pixel grid		
		0 px	⅓ px	½ px

Figure 3.13 Antialiased rendering of Snellen-E at small pixel sizes (Chrome 88.0.4324 on Windows 10). Both the three-pixel and the four-pixel variants exhibit features in the pixel image that are larger than the strokes of the "E" in the stimulus geometry, which could result in inflated performance results for similar stimuli at such small pixel sizes.

Due to its increased burden on the CPU, antialiasing was not generally applied to all graphical content until after the new millennium. For web-based cartographic applications, Jenny et al. (2008) recommended that map authors evaluate the antialiasing routines of graphics programs and browser plugins before making a decision on whether to activate it. For rendering text, the use of antialiasing was generally recommended "because it is graphically more pleasing" (Jenny et al., 2008, p. 12). With the implementation of antialiasing routines in hardware and the increase of processing power of CPUs, and due to the higher perceived attractiveness of antialiased graphics, antialiasing was activated by default on desktop and mobile operating systems. Today, on most systems there is no option to disable antialiasing, and workarounds have to be applied in order to achieve a non-antialiased appearance for graphical output[21].

As discussed in Section 3.2.2, there is today a landscape of heterogeneous display sizes and resolutions, and fractional ratios of logical pixels to physical pixels are common. Thus, the notion of "exact pixel boundaries" is becoming increasingly unreliable or even misleading – even an integer pixel coordinate may not fall precisely on the boundaries of a group of hardware pixels. The prevalence of antialiasing for rendering graphical output corresponds with this development – using antialiasing, graphical elements can be placed on the display at arbitrary non-integer locations, and can routinely have non-integer line widths and dimensions. This represents a paradigm shift for 2D computer graphics: full alignment of graphical elements with the physical pixel grid becomes a rare special case instead of the norm on which the organization of screen content is based on.

---

[21] See the question "Can I turn off antialiasing on an HTML <canvas> element?" on Stack Overflow: https://stackoverflow.com/q/195262 (accessed 2022-06-05)

3.2.3.2. Font hinting and pixel fitting

The limited fidelity of conventional screens presented a particular challenge for rendering text: the small details of individual letters can appear seriously distorted when rendered as a bitmap image, particularly at small sizes. Antialiasing may improve the overall appearance, but will result in a blurred image on low-resolution displays, which reduces legibility of small text. Therefore, for rendering small text on digital displays, alignment of the components of letters with the pixel grid is highly desirable. In addition to geometry information for each letter, digital fonts can therefore contain information on how the elements of a letter should be modified for better alignment with the pixel grid. Adding the information on which parts of a letter should align with pixel boundaries and which parts may be displaced to accommodate such alignment is called *font hinting* (Stamm, 1998, 2009).

According to Spolsky (2008), Apple and Microsoft use different strategies for font rendering in their operating systems, with Microsoft using more aggressive hinting at the cost of omitting fine design details at smaller font sizes. Recently, Apple has stopped fitting fonts to the pixel grid altogether, since with the availability of high-resolution "retina" displays the problem of irregular line widths and spacing has been greatly reduced. This has been perceived to result in a "blurry" appearance on low-resolution screens, while for high resolution screens the geometry of the font is more faithfully preserved (Rozario, 2022).

Similarly to text characters, graphical icons will also be affected by the reduction of sharpness when rendered with antialiasing. Traditionally, small icons intended for screen presentation have been created as pixel-based bitmap images, potentially in multiple discrete sizes. With the prevalence of more powerful CPUs and high-resolution displays, designers have increasingly used vector graphics to define icons, as these will be presented with crisp edges on a higher resolution display without the need to create separate versions. However, on low resolution displays small icons could suffer from blurred appearance when rendered with antialiasing. Icons are therefore often designed according to a (hypothetical) pixel raster of a low-resolution device, and should be rendered at a size that corresponds with that raster. This will increase the probability of edges coinciding with the boundaries of physical pixels on systems with nominal pixel sizes or integer pixel multipliers, if the icon is placed at integer coordinates. This process of designing vector graphics with the underlying pixel grid in mind is called *pixel fitting* or *pixel snapping* (Zhang, 2020a). Grid sizes of 16, 18, 24, 36 or 48 pixels are commonly used, and an icon designed with a specific pixel grid should be used only at a corresponding nominal size to avoid blur (Zhang, 2020b). Figure 3.14 shows an icon from the "Maki" icon set, which, like all icons from the set, was designed using a 15×15 pixel grid.
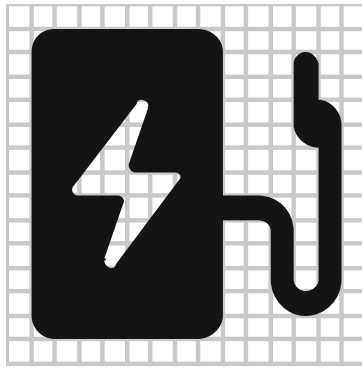
Figure 3.14 The icon for "charging station" from the Maki icon set, defined as of vector graphics designed against a 15×15 units grid (grey lines in the background). Icon designers using a grid-fitting approach are free to use arbitrary geometry not aligned with the grid, but are encouraged to fit the main elements of the icon to grid boundaries. Icon source: https://labs.mapbox.com/maki-icons/ CC-0, published by Mapbox.

3.2.3.3. Subpixel rendering

So far, the physical pixel has been treated as the "atom" of computer graphics, the minimum unit for which visual output can be generated. However, as has been discussed in Section 3.2.2, pixels are composed of subpixels which create the impression of arbitrary colours by additive colour mixing. If the subpixel arrangement of the display is known (such as in a horizontally partitioned RGB LCD pixel), its subpixels can be used to render geometry at a higher resolution than one pixel, at the cost of accurate colour reproduction at the boundaries.
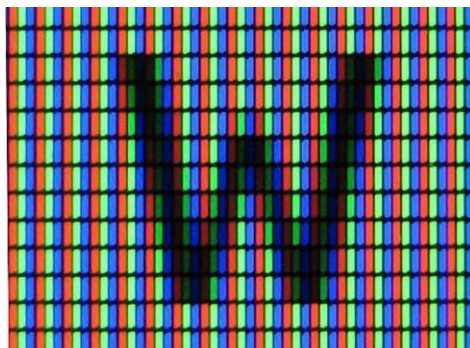

Figure 3.15 Microscopic image of the letter "w", rendered with subpixel rendering. Note that at the boundaries of the letter, the RGB subpixels of the display are used to establish the horizontal resolution. Image source: CC BY-SA Wikimedia user ALexL33.

Subpixel rendering on conventional LCDs with horizontal partitioning has been primarily used for improving font rendering (Sheedy et al., 2008) and has also been shown to be applicable to image resampling (Fang et al., 2012). However, the subpixel arrangement of the display has to be known by the system or correctly configured by the user in order to achieve the desired effect and avoid detrimental artifacts such as colour fringes caused by incorrect assumptions about the subpixel arrangement (Stamm, 2009). With high-density screens and unconventional subpixel layouts now common (see Section 3.2.2.1), subpixel rendering is becoming less relevant as the resolution of full pixels of these devices now surpasses the fidelity that could be accomplished by exploiting subpixels on conventional screens of lower resolution. Furthermore, mobile devices can be used in portrait or landscape orientation,

which the rendering software would need to take into account for correctly applying subpixel rendering. Apple has disabled subpixel rendering in the current versions of its operating systems, and subpixel rendering is generally not available on Android.[22]

3.2.4. Conclusions from the review of technological foundations

The following conclusions can be drawn from the review of the technological foundations of display technology presented above:

- CRT display technology is practically irrelevant today, having been replaced for most use cases by LCD and OLED displays.
- OLED- and LCD-based display technology today allows for practically arbitrary resolutions, not limited by technological factors any more.
- Contrast ratios produced by modern displays in indoor environments is generally well above the levels required for ophthalmological examinations, therefore ensuring highest visual performance of observers for stimuli displayed at or near maximum contrast on these displays. While OLED- and LCD-based displays have different contrast profiles due to their different working principles, in illuminated environments this difference should be negligible.
- Mobile devices are manufactured today with a wide range of pixel densities (250–600ppi). All devices sold today have a pixel density higher than the nominal resolution assumed by operating systems of 160ppi. One reference pixel (referred to as density-independent pixel or CSS pixel, depending on the platform) specified as the dimension of on-screen graphical elements will therefore always be represented by more than one physical pixel on current devices. For desktop monitors, normal-resolution devices in the range of 80–100 ppi are still commonly used, but the landscape of pixel densities deployed in these environments is also heterogeneous. All modern operating systems support the configuration of the mapping from logical to physical pixels in one way or another.
- Logical and physical pixels are increasingly decoupled, with flexible mappings including non-integer and fully user-configurable pixel ratios. However, even for web content, access to physical pixels is still possible for end-user software for purposes of pixel-optimized rendering, by APIs like `devicePixelContentBox`.
- Subpixel arrangements of modern displays can deviate from a RGB arrangement with one subpixel per primary colour. However, the operating system always exposes a model of a homogeneous pixel grid adhering to the RGB colour space to software run by the end user. Without explicit detailed knowledge of the particular subpixel structure of a particular device, addressing individual subpixels to optimize the display cannot be achieved by software outside the operating system. Due to the high resolutions of currently manufactured devices, the relevance of subpixel rendering techniques to increase the smoothness or crispness of graphical elements seems to be waning.

---

[22] Sources:
https://forums.macrumors.com/threads/the-subpixel-aa-debacle-and-font-rendering.2184484/
https://stackoverflow.com/questions/5613706/sub-pixel-rendering-of-fonts-on-the-ipad
https://daringfireball.net/2007/12/anti_aliasing_on_the_iphone
https://www.quora.com/How-are-Androids-fonts-optimized-if-at-all-for-PenTile-sub-pixel-patterns

- Antialiasing is ubiquitously availably today, and cannot be easily deactivated in many 2D graphics APIs. Antialiasing potentially reduces the crispness of graphical output, but this is thought to be compensated by the higher physical pixel densities available for modern devices.

3.2.4.1. Overview of display resolutions, pixel size and corresponding visual acuity scores

Table 3.2 compiles estimates of the resolution capability at various visual acuity scores (see Section 3.1), corresponding metric sizes at typical viewing distances of 30 and 60 cm, and pixel densities of exemplary devices in tabular form, which is hoped to be of use for reference purposes. The mobile devices that will be used in the empirical studies presented in this thesis are highlighted in the table, allowing interested readers to look up the corresponding pixel sizes, logMAR scores and desktop resolutions.

Pixel size @ 30 cm (mobile, print)	Pixel size @ 60 cm (desktop)	Apparent size (minutes of arc)	logMAR	ppi @ 30 cm Mobile resolution	ppi @ 60 cm Desktop resolution	Comments / Device
1.00 mm	2.00 mm	11.46'	1.06	25	13	Metric dimensions
0.10 mm	0.20 mm	1.15'	0.06	254	127	"
0.50 mm	**1.00 mm**	5.73'	0.76	51	25	"
0.05 mm	**0.10 mm**	0.57'	-0.24	508	254	"
0.09 mm	0.17 mm	**1.00'**	**0.00**	292	146	20/20 Vision
0.04 mm	0.09 mm	**0.50'**	**-0.30**	577	289	Upper limit of visual acuity
0.01 mm	0.02 mm	**0.10'**	**-1.00**	2920	1460	Vernier acuity
0.16 mm	0.31 mm	1.79'	0.25	**163**	81	Apple iPhone 1st Gen.
0.11 mm	0.22 mm	1.28'	0.11	**228**	114	**LG P-970 (D$_1$)**
0.10 mm	0.19 mm	1.10'	0.04	**265**	132	**LG K50S (D$_5$)**
0.08 mm	0.16 mm	0.89'	-0.05	**326**	163	iPhone 4
0.07 mm	0.15 mm	0.85'	-0.07	**342**	171	**Sony Xperia V (D$_2$)**
0.06 mm	0.11 mm	0.63'	-0.20	**460**	230	iPhone 13 Pro
0.05 mm	0.10 mm	0.58'	-0.23	**500**	250	Samsung Galaxy S22 Ultra
0.05 mm	0.10 mm	0.56'	-0.25	**522**	261	**Samsung Galaxy Note 4 (D$_3$)**
0.03 mm	0.06 mm	0.36'	-0.44	**801**	401	**Sony Xperia Z5 Pro (D$_4$)**
0.15 mm	**0.30 mm**	1.72'	0.24	169	**85**	Conventional pixel
0.11 mm	0.22 mm	1.26'	0.10	231	**116**	CSS pixel (1/96th inch @ 28 inch)
0.05 mm	0.09 mm	0.52'	-0.28	561	**280**	Dell UltraSharp UP3218K

Table 3.2 Various real-world pixel sizes at 30 cm and 60 cm viewing distances, and their corresponding apparent sizes, logMAR and ppi values. Cells with a bold border indicate anchoring values of special relevance. Rows with light orange background represent devices D$_1$-D$_5$ used in the studies presented in this thesis.

3.3. Minimum dimensions for cartographic symbology

"These scientific studies […] have made only limited
contributions to the understanding of practically
effective design methods." (Wood, 1993, p. 149)

This section will review and compile guidelines on minimum dimensions of cartographic symbology for printed and screen-based maps. Additionally, an attempt is made to contextualize the work undertaken for this thesis, and related work in the past, within the overall development of related ideas and paradigms in academic cartography.

3.3.1. Cartographic communication and representation

"Measurements of user response to map symbol design, it was
hoped, could lead to formulation of scientific principles."
(Wood, 1993, p. 151)

One major paradigm of cartography conceptualizes maps as a form of *communication* between map producers and map users. Communication processes, in turn, can be modelled according to the Shannon-Weaver model as having three fundamental components – a sender, a receiver, and a channel (or medium) carrying information from the former to the latter. For successful communication, any information that the sender wants to transmit needs to be encoded into some representation that can be (a) carried across the channel and (b) decoded by the receiver to retrieve its information content. The channel is also subject to noise which may distort the information carried across (Hake et al., 2002, p. 8f).

Koláčný (1969) proposed an extension to the generic communication model, specifically representing the cartographic communication process. At the centre of the process is the map as the main medium for cartographic communication. As sender, there are one or more cartographers, and as recipient, the user(s) of the map. Both senders and recipient are influenced by psychological and environmental factors, such as their knowledge of the world, prior experience, and tasks or aims they may have with respect to the communication act. A crucial role in Koláčný's model is assigned to cartographic language, "a system of map symbols and rules for their use" (Koláčný, 1969, p. 48), which facilitates the encoding and decoding of the communicated information. While the medium of cartographic communication – the map – has a central part in Koláčný's model, the main focus of the model is placed on the psychological and social aspects of map creation and use.

MacEachren (1995) acknowledges that the communication model has expanded the view of cartography to encompass "more than mapmaking" and to include the formation and negotiation of spatial knowledge in social processes, but criticises the narrow focus of the communication model on the efficiency of maps to communicate predetermined messages (p. 6), and proposes a shift of focus towards the *representation* of our world on maps. Such a paradigm is better suited to include the use of maps for exploratory purposes, in which the message, or even the question, may not be known in advance, as made possible by the use of interactive computer systems in geovisualization scenarios. Maps have the "potential to

convey many meaning at multiple levels of analysis" (p. 9) instead of only a set of predetermined messages.

Both approaches – whether focused on communication or representation – call for information to be encoded in a map in a way that is legible by the map user and allows them to draw valid conclusions about reality. Much attention has therefore been paid in the history of cartographic research to the way spatial information is represented on maps, and how the reading of maps can be made more efficient and reliable. While cartography is a creative discipline that allows virtually infinite variants of how the world is represented on a map, a fundamental framework of the commonly used building blocks used to represent geographic reality is laid out by Imhof as follows:

> "Die graphischen Grundelemente der Karten sind Punkte, Linien, Kleinfiguren, Punkt- und Linienanhäufungen oder –scharungen (Schraffuren, Raster usw.), Farb- und Schummertöne, sowie die Beschriftung" (Imhof, 1972, p. 27)

Points, lines, small figures, groups of points and lines (including hachures and halftone screens), colour, and textual labels make up the basic catalogue of graphical elements frequently used to represent geographic phenomena on maps. Each of these elements can be varied in multiple ways, which Jacques Bertin has systematized as *retinal variables* (later commonly referred to as visual variables): size, value (brightness), texture, colour (hue), orientation, and shape (Bertin, 1983, p. 60f). While other authors have proposed variants of or extensions to this catalogue of visual variables, the basic framework has found to be valid for structuring our thinking about cartographic symbology to this day (MacEachren, 2019).

While the graphical elements and visual variables give map designers a lot of creative freedom, there are some constraints that limit the range of choices when the goal is to produce a sensible and useable map. For the size of map elements, an upper limit is usually established by the media context in which the map will be reproduced, the extent and density of the geographical features depicted, the amount of information that needs to be represented and the desired graphical density of the map. The lower limit for the size of elements will be dependent on the fidelity of the map reproduction medium, the viewing situation, and the map user's intent and their capability to resolve small details. For most of these limiting factors, the circumstances and capabilities of the map user have to be taken into account. To help map designers make decisions that result in a map that is practical and reliable to use without frustration, and avoid a situation in which the designer makes choices based solely on their own preferences and needs, cartography has a history of condensing needs and preferences of map users into guidelines for the design of maps and the symbology on them. Heupel (1977) even posits that the creation of such guidelines and style sheets is one of the two main scientific activities of academic cartography.

The following sections will give an overview of guidelines and models that have been put forward by academic cartographers on the question of minimum dimensions of cartographic symbols. Because the medium through which a map is reproduced has, besides the capabilities of the map user, a major impact on map design and, specifically, on minimum dimensions of symbology that can be reliably reproduced, the discussion will be split in two

parts: first the guidelines established for printed maps will be discussed, following by coverage of guidelines available for maps reproduced on digital displays.

3.3.2. Guidelines on symbology dimensions for printed maps

> *"our efforts to explain map reading in relation to constraints of vision and the resulting influence on discrimination and other low-level processes have resulted in an idiosyncratic set of case-specific conclusions that we cannot confidently extend to other applications."*
> *(MacEachren, 1995, p. 22)*

As argued above, in a view of cartography as communication or representation, there cannot be a map without a medium. Today, two main media types for reproducing maps come to mind: printed paper and digital displays. However, before computers gained the graphical capabilities needed for an acceptable reproduction of maps, printed paper was the sole main medium used for map reproduction in the western world. In contrast to the story of the fish that do not know what water is[23], cartographers were well aware of the properties of the medium, and many discussions of printing technology can be found in the cartographic journals of the 1960s and 1970s. Nevertheless, because print was the only medium, discussions predating the advent of digital displays often imply printed reproduction, instead of explicitly stating it.

In fact, the properties of the printed medium were taken into account in early theoretical treaties on map generalisation and symbology. Töpfer (1974) mentions that "the limits for a scaled-down representation of objects is not only dependent on map scale and the size of an object's footprint, but also the means of visualization dedicated to the task" [24] and acknowledges that there can be no absolute limits for scaled representations but only those related to a specific purpose of the map (p. 21) (today, this would probably be referred to as a specific map use scenario). In an example given on the introductory pages, Töpfer introduces a minimum distance of 0.2 mm for contour lines without further justification or reference to a specific purpose, demonstrating perhaps that some assumptions on map medium and use must be made for any practical map design endeavour.

Arnberger and Kretschmer (1975) provide such assumptions, together with justifications anchored both in human perception and the capabilities of the medium. They claim that a limit of 0.1 mm line width corresponds with the limits of what can be manually drawn and what can be reproduced using modern printing technologies, as well as with the limits of human vision, for which they specify a minimum line width of 0.05 – 0.1 mm for black lines against white background (Arnberger & Kretschmer, 1975, p. 45). For "correct and rapid recognition" and a the counting of parallel lines, a minimum line width of 0.1 mm and a line

---

[23] David Foster Wallace *This Is Water: Some Thoughts, Delivered on a Significant Occasion, about Living a Compassionate Life*. 2009, Little, Brown and Company, New York, NY.

[24] Own translation; German original: "daß die Grenzen der maßstäblichen Darstellung nicht nur vom Kartenmaßstab und der Größe des Objektgrundrisses, sondern auch von den vorgesehenen Darstellungsmitteln abhängig sind." (Töpfer, 1974, p. 20)

separation of 0.25 mm is recommended. From this, the authors derive elaborate tables for graduated line widths for different topographical elements at various scales. Later in the book, a table for minimum dimensions for isolated graphical elements is given, specifying as minimum dimension at high contrast a line width of 0.05 mm, a diameter of 0.24 mm for circles and a side length of 0.4 mm for filled squares (p. 227).

Other authors in the German-language literature at the time come to similar conclusions, and treat the topic with various levels of elaboration. Imhof (1972) mentions that the diameter of map point symbols typically lies between 0.6 and 10 mm (p. 71), and specifies minimum dimensions for the diameter of various point symbols, starting from 0.3 mm for isolated graphical points (smaller values for points as part of a group or pattern), 0.6 mm for simple geometric figures and 1.5 mm for simple iconographic figures (p. 67). The Swiss Cartographic Society refers to perceptual limits as motivation for minimum dimensions in their publication on map generalisation[25], citing a resolution capability of human vision of 0.02 mm at a distance of 30 cm, mandating a minimum line width of 0.04 mm, which, according to the authors, coincides with the capabilities of printing technology (Schweizerische Gesellschaft für Kartografie, 1980, p. 13). However, it is remarked that pushing cartographic symbology to the limits of perception and reproduction technology may not be advisable, for four reasons: a) important objects must be recognized rapidly, not only be barely noticeable; b) differences in form must be clearly recognizable; c) low lighting conditions and brighter colours will reduce contrast; and d) the best means for reproduction will not always be available (ibid.). Thus a table of minimum dimensions at full contrast is given, including values of 0.05 mm for line width, 0.25 mm line separation and 0.3 mm for discriminating a solid black square from a circle (see Section 3.3.2.1 for more data extracted from this publication and the previously mentioned authors). The publication also lists the following artifacts of the printing process that may negatively affect the quality of the reproduced map: widening of fine lines, fusion of crossing lines, merging of lines and areas, rounding of shapes, loss of contrast, mismatch of colour layers, rough paper surface (ibid., p. 15).

None of the authors cited above cites any empirical studies or explains the process of how the specified numbers were derived. This seems to be reflected in increased calls for empirical verification of postulated guidelines in the German-language literature. Knorr (1967) states that "a systematic investigation on the perception of cartographic means of expression has not yet been done"[26] and calls for extensive investigation of map symbols and the possibilities of size, form, colour and combinations thereof (p. 157). He concludes that the necessary work would be "without doubt very extensive and time-consuming", but due to the high interest considered indispensable[27]. Werner (1970) decries the lack of psychological foundations for map design guidelines (p. 103) and delivers the verdict that judgement of map designs is usually done by a small group of professionals by their own subjective criteria (p. 104), and

---

[25] First edition published in 1975.

[26] Own translation. Original: „Es ist bis heute eine systematische Untersuchung über das Auffassen von kartographischen Ausdrucksformen noch nicht erfolgt; […]" (Knorr, 1967, p. 150)

[27] Own translation. Original: „Es besteht kein Zweifel, daß die hierfür durchzuführenden Arbeiten infolge des vielseitigen Materials sehr umfangreich sein und viel Zeit in Anspruch nehmen werden. Sie werden aber im Interesse der Aufgaben für unabdingbar notwendig gehalten." (Knorr, 1967, p. 157)

that the "self-image of map designers has so far prevented a detailed analysis of map users" (p. 106)[28]. He goes on to state that scientific cartography should compare the "information goals" of a cartographic product with their realisation in a map, and that any theoretical conceptualisation of cartographic symbology without taking into account the laws of seeing and psychology are "worthless" (ibid.). Werner concludes that "cartography needs an extraordinary expansion of its empirical foundation" to gain insight into genuinely *cartographic* questions of information transmission (p. 108)[29]. In years following these calls for more empirical investigation, multiple authors in the German journal *Kartographische Nachrichten* reported on studies undertaken in the US and UK, where they sensed more activity in the desired direction.

According to Alan MacEachren (1995), indeed a greater interest of US post-war cartography in the empirical investigation of perceptual processes can be traced back in origin to Arthur Robinsons book *The Look of Maps* (1952), which is also identified as the origin of modern scientific cartography by other scholars (Petchenik, 1983; Dobson, 1985; Wood, 1993). Robinsons cites in detail an empirical study on the legibility of various typefaces (Luckiesh & Moss, 1937) and posits that variation of map symbology by size "is of fundamental significance" (p. 66). However, "no [studies] of relative visibility or degree of visual distinction between line weights" would be known (p. 67), and several cartographic procedures such as minimum width difference for lines "could be evaluated by testing" (p. 72), while others "seem likely forever to remain essentially subjective insofar as their evaluation is concerned" (p. 73).

Subsequent work towards the called-for empirical investigation includes the influential work by Flannery on graduated map symbols (undertaken before 1956 and reported in a paper in 1971) and by Jenks & Knos (1961) on the perceived value of grey patterns focused on the retrieval of quantitative information from thematic maps. Such emphasis on symbology representing quantitative information seems plausible, as in the age of printed maps, thematic maps served as storage of information which would not be commonly available to map users in other form, and therefore the accuracy of extraction of quantitative information from maps was of crucial importance.

In the UK, several studies on map perception were undertaken by the Experimental Cartography Unit in collaboration with the Department of Psychology of the University College London, including tests on the legibility of relief maps (Phillips et al., 1975) and contour lines (Phillips, 1979). However, in both studies, line width and symbol size were not among the tested variables. Michael Wood from the University of Glasgow published a paper on visual perception and map design, in which visual acuity or minimum dimensions are not discussed; instead the focus is put again on the extraction of quantitative information from graduated symbols, and other forms of processing at higher cognitive levels (Wood, 1968).

---

[28] Own translation. Original: "Das Selbstverständnis der Kartenentwerfenden hat bisher von einer detaillierten Benutzeranalyse abgehalten." (Werner, 1970, p. 106)

[29] Own translation. Original: „[…] daß die Kartographie eines ganz außerordentlichen Ausbaues ihrer empirischen […] Basis bedarf, um die bisher überwiegenden sachgebundenen und daher nicht kartographischen Ergebnisse durch solche des kartographischen Informationsvorganges zu ersetzen."

In the literature review undertaken for this thesis, with limited access to older, uncited and unpublished work from the US and UK, no study on minimum dimensions from the era was found or cited in any other of the reviewed works. US textbooks on cartography seem to dedicate much less space on the issue than their counterparts from Europe. The fifth edition of Arthur Robinson's textbook *Elements of Cartography* (A. H. Robinson et al., 1984) mentions that there is "some disagreement as to the exact measure" of minimum dimensions, and propose as a "practical" measure of 1 minute of arc representing "perfect vision and perfect conditions of viewing" (p. 145). Because such ideal conditions can rarely be assumed, it would be "wise" for cartographers to establish a somewhat larger minimum size, for which 2 minutes of arc is proposed and a table of sizes derived from this measure for different viewing distances is presented (p. 146). MacEachren (1995) attributes little relevance to the issue of minimum dimensions, "because lines or symbols too small to be seen cannot be consistently printed" (p. 125) and states that "no guidelines exist because little empirical testing has been done" (ibid.). Terry Slocum's book on thematic cartography (Slocum, 1999) does not specify any minimum dimensions and only advises to "avoid type smaller than 6 points" in the first edition (p. 37). Gardiner (1981) bases his theoretical considerations for cartographic symbology on the properties of the human eye, and cites an acuity of one minute of arc as "normal vision", but mentions that lower than normal vision would be "common". Among more informal suggestions, he derives a limit of 0.05 mm for black lines at a "normal" viewing distance of 25 cm.

According to Montello (2002), the first wave of cognitive map design research reached its peak in the late 1970s, with increasing scepticism towards the relevance of its findings for real-world map production being voiced in the 1980s. Wood (1993) claims that the psychologically oriented studies "have made only limited contributions to the understanding of practically effective design methods." (p. 149) and that the study of simple map symbols in isolation would be of limited relevance in the context of more complex real-world maps.

> Unfortunately, the more complex the map-like stimuli the less predictable and more variable were observer-responses. […] The result, yet again, was a lowering of confidence in response-based research of this type as a source of map design knowledge. […] With a few notable exceptions […], psychophysical studies have largely failed to produce the hoped-for answers to design questions" (Wood, 1993, p. 151)

Wood, however, acknowledges that a "full understanding of maps as representations will require a multi-directional approach at all cognitive levels from first sensation of the image" (Wood, 1993, p. 152). Barbara Petchenik (1983) sees "limited utility of academic research in map design", not least because "the analytic process of scientific research and the synthetic process of design are distinctly different, both cognitively and epistemologically" (p. 40). Dobson (1985) defends the need for controlled experiments against the "ritual slaughter of psychophysical research in cartography" (p. 27) and remarks:

> In a discipline whose product is a visual display, one would assume that it would be necessary to know or determine the capabilities of the visual system in respect to the processing of visual symbols. (Dobson, 1985, p. 31)

> Guidelines for effective design can have little significance unless they are based on a knowledge of the abilities of subjects in respect to visually acquiring, differentiating, and processing graphic signs or symbols. […] More importantly, the specific display characteristics that influence the speed and accuracy of task solution may be isolated."
> (Dobson, 1985, pp. 37–38)

Worth (1989) formulates what could be read as an attempt to reconcile the necessary strict simplification of experimental research with the creativity and complexity of applied map design in the conclusions of his article on the challenges of experimental map design research:

> Although we can talk about the science of cartography with some confidence, it is not a pure science. […] Subjective decision making and aesthetic considerations are extremely important in map design. […] By their very nature, experimental tests in cartography must involve many subjective decisions, and although we must do our best to apply the scientific method, subjectivity will always be involved. (Worth, 1989, pp. 152–153)

These debates in the scientific community coincided with the advent of computer-based map production and the urge to explore the possibilities of these new technologies. Among engineers and a new group of GIS professionals, there was little time to engage in theoretical academic debates and critical discourse because, as one interviewee of Nadine Schuurman put it, "everyone was just too busy" at the time (Schuurman, 2000, p. 582). Even the design and implementation of controlled experiments on basic aspects of map perception may have lost its appeal in the face of exciting new technological possibilities.

In the German-speaking countries, experimental cartography continued mainly at research groups in Dresden, Berlin and Vienna (Montello, 2002), and later in Trier (Bollmann, 1981). At the Academy of Sciences in Vienna, Erich Vanecek (1980) conducted a series of experiments with printed map symbols, including recognition of complex, combined point symbols and map search tasks. It is in this context that the only work on empirical verification of minimum symbology dimensions for printed maps that was found in the literature review for this thesis was conducted: Chlupac's (1982) PhD dissertation *Die Erkennbarkeits- und Unterschiedsschwelle verschiedener geometrischer Signaturformen* (own translation: "Thresholds for recognizing and discriminating diverse geometrical map symbols"), supervised by Vanecek.

Chlupac used a stack of 139 white PVC cards, on each of which a single geometric shape, filled solid black, had been printed in the centre. The shapes included: circle, square, rectangle in horizontal / vertical orientation, triangle pointing upwards / downwards, and pentagon (see Figure 3.16). Symbols were scaled to various sizes to cover an area in the range of 0.04 – 0.6 mm². The cards were presented to participants in random order, illuminated by a 40 Watt lightbulb, at a "typical reading distance" of 30 cm. For each card, participants were asked to identify the shape printed on it, and rate how confident they were in their judgement.

In Chlupac's experiment, participants could identify with a correctness rate above 95% shapes at the following sizes or larger[30]: circles of a diameter of 0.83 mm, squares with a side

---

[30] Chlupac specifies all sizes as the area covered by the shape in mm². The values have been converted to length measurements for the purpose of comparison with other guidelines. Also, the proportions of the rectangular shapes are not specified by Chlupac – a side length ratio of 2:1 is assumed, based on a figure of the shapes.

length of 0.48 mm, triangles with a base of 0.50 mm and rectangles with a short side of 0.21 mm. Subjective confidence of participants generally corresponded with their success rate, showing the ability of participants to reflect on the confidence of their perception.

Figure 3.16 The shapes used by Chlupac (1982) in his study on the legibility of map symbols.

Despite the call for empirical research to back up map design guidelines (Werner, 1970; Grünreich, 2008), such guidelines were continued to be published without citing any of the studies that may have informed them. The Swiss Cartographic Society published the latest version of their detailed guidelines on map symbology in 2002 (see Figure 3.17 for an excerpt), mentioning a "resolution capability" of the human eye of 0.2 mm (ibid., p. 26) as a foundation, but without discussing any procedures for, or results from, empirical verification. Other works published by the Swiss Cartographic Society contain more detailed guidelines for specific aspects of cartography, including for representation of the built environment (Spiess, 1990b) and thematic cartography (Spiess, 1990a). Hake, Grünreich and Meng (2002, p. 110f) present the most detailed specification of minimum dimensions of any of the textbooks reviewed for this thesis, proposing dimensions for various types of symbols, both for maximum contrast as well as reduced contrast reproduction. Minimum dimensions specified by the authors mentioned above are compiled in the tables in Section 3.3.2.1.

Most studies of printed maps assume a viewing distance of 30 cm for a typical map use scenario (Morgenstern, 1974; Chlupac, 1982; Schweizerische Gesellschaft für Kartografie, 2002; Jenny et al., 2008) – only a few works explicitly state different assumptions, for example 50 cm are specified by A. H. Robinson et al. (1984, p. 146) as the smallest viewing distance considered.

Enlarged 4:1	1:1	Minimum Size	Minimum Stroke	Minimum Distance	Remarks
Point Symbols					
+ ✕	+ ✕	0,80 mm	0,12 mm		Cross
△	△	1,20 mm	0,08 mm		Triangle, hollow
□	□	0,70 mm	0,08 mm		Square, hollow
•	·	0,30 mm			Round Dot
○	○	0,60 mm	0,10 mm		Colored, hollow
Line Symbols					
———	———		0,08 mm		Black / White
═══	———		0,08 mm	0,25 mm	Double Line
≣≣≣	≣≣		0,08 mm	0,25 mm	Line Bundle 3 Lines / mm
〜〜〜	〜〜	0,30 mm	0,08 mm		Minimum Wiggle
• • • • • •	··········	0,15 mm	0,15 mm	0,40 mm	Dotted Line

Figure 3.17 Guidelines on minimum dimensions of point and line symbology at maximum contrast, adapted from Schweizerische Gesellschaft für Kartographie (2002, p. 27).

3.3.2.1. Overview of proposed guidelines for printed maps

Tables 3.2–3.5 compile the recommendations for minimum dimensions given by various authors for the depiction of point, line and area features, and textual labels on printed maps.

Point Symbols

Value	Symbol type / dimension	Source
0.1 mm	Point diameter (as part of fill pattern)	Imhof (1972)
0.2 mm	Point diameter (in groups)	Imhof (1972)
0.3 mm	Point diameter (circular)	Imhof (1972)
0.6 mm	Point diameter (simple geometric figures, filled)	Imhof (1972)
1.0 mm	Point diameter (simple geometric figures, hollow)	Imhof (1972)
1.5 mm	Point diameter (simple graphical icons)	Imhof (1972)
0.24 mm	Point diameter (dot)	Arnberger and Kretschmer (1975)
0.44 mm	Point diameter (circular)	Arnberger and Kretschmer (1975)
0.40 mm	Side length of square	Arnberger and Kretschmer (1975)
0.3 mm	Point diameter (circle/square)	Schweizerische Gesellschaft für Kartographie (SGK) (1980)
0.3 mm	Point diameter (hollow circle)	SGK (1980)
0.15 mm	Point diameter (circular)	SGK (1980)
1.0 mm	Side length (hollow triangle)	SGK (1980)
0.48 mm	Side length of square symbol (>95% correct identification)	Chlupac (1982)
0.50 mm	Side length of triangle symbol (>95% correct identification)	Chlupac (1982)
0.21 mm	Short side of 2:1 rectangle symbol (>95% correct identification)	Chlupac (1982)
2 arcmin (≈ 0.3 mm @ 50 cm)	Point diameter	A. H. Robinson et al. (1984)
0.4 mm	Point diameter (circle)	Spiess (1990b)
0.5 mm	Point diameter (square)	Spiess (1990b)
0.25 mm	Point diameter [31]	Hake, Grünreich and Meng (2002)
0.5 mm	Point diameter (circle/square) [30]	Hake, Grünreich and Meng (2002)
0.6 mm	Point diameter (circle/square, hollow) [30]	Hake, Grünreich and Meng (2002)
0.30 mm	Point diameter (circular)	SGK (2002)
0.70 mm	Side length (hollow square)	SGK (2002)
0.80 mm	Cross symbol	SGK (2002)

Table 3.3 Minimum dimensions for point symbols on printed maps, as specified by various authors. Values set in italics denote values derived from empirical studies.

---

[31] Hake et al. specify values for optimal contrast and reduced contrast – only the value for optimal contrast is listed here.

Line Symbols

Value	Symbol type / dimension	Source
0.2 mm	Line separation	Töpfer (1974)
0.05 mm	Line width	Arnberger and Kretschmer (1975)
0.1 – 0.15 mm	Line width (contour lines)	Arnberger and Kretschmer (1975)
0.15 mm (better: 0.25 mm)	Line separation (for counting contour lines)	Arnberger and Kretschmer (1975)
0.05 mm	Line width	Schweizerische Gesellschaft für Kartographie (SGK) (1980)
0.1 mm	Line width (dotted line)	SGK (1980)
0.25 mm	Line separation	SGK (1980)
0.08 mm	Line width	Spiess (1990b)
0.16 mm	Line separation	Spiess (1990b)
0.05 mm	Line width [30]	Hake, Grünreich and Meng (2002)
0.25 mm	Line separation (fine lines) [30]	Hake, Grünreich and Meng (2002)
0.15 mm	Line separation (thick lines) [30]	Hake, Grünreich and Meng (2002)
0.08 mm	Line width	SGK (2002)
0.08 mm	Line width (double line)	SGK (2002)
0.25 mm	Line separation (double line)	SGK (2002)
0.15 mm	Line width (dotted line)	SGK (2002)

Table 3.4 Minimum dimensions for line symbols on printed maps, as specified by various authors.

Area Symbols

Value	Symbol type / dimension	Source
0.25 mm²	Area size (footprint of lake)	Töpfer (1974)
0.4 – 0.6 mm	Area diameter (for recognizing form)	Arnberger and Kretschmer (1975)
0.25 mm	Area separation	SGK (1980)
0.4 mm	Minimum extent of area, black on white	Spiess (1990a)
0.6 mm	Minimum extent of area, with outline	Spiess (1990a)
1 mm²	Area size (for recognizing color)	Hake, Grünreich and Meng (2002)
0.3 mm	Area size (side length of rectangle) [30]	Hake, Grünreich and Meng (2002)
0.20 mm	Area separation (rectangle, small areas) [30]	Hake, Grünreich and Meng (2002)
0.15 mm	Area separation (rectangle, large areas) [30]	Hake, Grünreich and Meng (2002)
0.35 mm	Side length of square	Schweizerische Gesellschaft für Kartographie (SGK) (2002)
0.20 mm	Separation of rectangular shapes	SGK (2002)
0.80 mm	Minimum diameter of colored area	SGK (2002)

Table 3.5 Minimum dimensions for the depiction of areas on printed maps, as specified by various authors.

Text labels

Value	Symbol type / dimension	Source
1.2 mm	Text height	SGK (1980)
0.6 mm	Text height (optimal contrast)	Hake, Grünreich and Meng (2002)
1.0 mm	Text height (reduced contrast)	Hake, Grünreich and Meng (2002)
6 points (≈ 2.1 mm)	Text height	Slocum (1999)
3.2 points (≈ 1.13 mm)	Text height (uppercase, sans serif)	Slocum et al. (2009)

Table 3.6 Minimum dimensions for text labels on printed maps, as specified by various authors.

3.3.3. Guidelines on symbology dimensions for screen-based maps

"The graphical design of a web map must be coarser and simpler than the design of a paper map so that it conveys the desired information under the less than ideal conditions of low screen resolution, increased viewing distance and shorter reading time."
(Jenny et al., 2008)

Digital systems were used to store, manipulate and output representations of geographic information beginning in the 1960s (Marble, 2015). However, the production of output that resembled in any way a high-quality cartographic product was technologically not feasible, and therefore not at the focus of most researchers and practitioners, until the late 1980s, when computer graphics methods, digital printers and screen displays had improved enough to allow for the necessary fidelity of graphical output (Fish & Brewer, 2015). Tobler (1988) addresses some issues related to the resolution of digital data, including output resolution. He remarked that "the smallest physical mark which the cartographer can make is about one half millimeter in size" (p. 131), and derives from sampling-theoretical considerations a table of the smallest features detectable on maps of various scales. Tobler defines resolution as "the capability of making distinguishable the individual parts of an object" (p. 132). However, the discussion focuses mainly on the resolution of geographical data and processes of transformation and resampling of such data, with photographic film as the only graphical output medium discussed.

An early empirical comparison of map reading performance on paper and screen-based maps, with the CRT screen at a "typical resolution" of 640×480 pixels showing the printed map enlarged by a factor of about 3.7, concluded that performance was consistently lower on the CRT screen for all tested map reading tasks such as locating, counting and identifying symbols (Gooding & Forrest, 1990). The authors mention that "few guidelines" are available for map designers targeting computer screens (p. 15), and conclude that display resolution is one of the main limiting factors for reliable map use. Brown (1993) gives an overview of the state of display technology at the time, with resolution stated as an aspect of "ultimate importance" for representing fine details, with current state of the technology not allowing for reproducing "at actual size very small text and symbols typical of maps" (p. 130). The article makes a clear distinction between the three usage scenarios of editing GIS data,

preparing designs for paper maps, and the final presentation of cartographic output to the end user on the digital display. For the latter scenario, cartographers are advised to make their design "sufficiently robust" to be read under the adverse conditions of on-screen presentation, and may have to come up with "very simple" designs, "containing only the truly relevant information", particularly in the context of safety-related information that needs to be reliably read (p. 133). However, already at that time Brown points out that screen-based presentation not only limits the cartographic presentation, but also offers novel possibilities such as animation, three-dimensional views, and user interaction (p. 135). The article ends with a vision for future developments:

> So, what do users (and designers) of screen map displays really want? From the visual point of view they would like very high resolution and very good quality colour. From the practical convenience point of view they would like large format yet easy portability […], without necessarily having to be attached to a bulky computer. […] If (when?) all this becomes possible, screen displays will be fully emancipated and accepted as a normal, perhaps the normal display mode. Then any discussions on designing for special types of display will consider, for example, the particular characteristics of printed paper maps. (Brown, 1993, p. 135)

At the time, some recommendations for dimensions of cartographic symbology for screen-based presentations can be found. Eaton (1993) mentions that a size of "about 12 pixels is enough for most symbols" (p. 185) in the context of design guidelines for digital nautical charts, without elaborating on the rationale for this statement. Morrison and Forrest (1995) examine point symbols for tourist maps and mention that a point symbol size of 5.5 mm² performs "significantly worse" than larger sizes, and use a symbol size of 7 mm² (corresponding with a side length of approximately 2.65 mm for a square symbol) as a "reasonable compromise in terms of performance" (p. 136).

By the end of the 1990s digital display technology had matured, with good quality CRT screens widely available. While these displays did not yet achieve the vision of a portable, high-resolution device expressed by Brown, technology at the time was considered mature, stable and relevant enough to justify the creation of more systematic and detailed guidelines for the design of cartographic output targeting such devices. Malić (1998) derived fundamental guidelines for minimum symbology dimensions that can be accurately reproduced on CRT screens from theoretical considerations and the analysis of example graphics at various sizes, rendered uniformly without antialiasing. The derived minimum dimensions for many common graphical elements, at the highest screen resolutions considered (108 ppi), are larger than the corresponding recommendations for printed maps, by a factor of 4 to 6 times. Matching intuition, the minimum dimensions specified by Malić for line width and line separation, as well as the minimum size of dents or protrusions of complex shapes, correspond with the size of a single pixel. In order to avoid aliasing effects to cause lines to be fused into one, a distance for line separation larger than one pixel is recommended. As the minimum size for distinguishing geometric shapes at the highest resolutions, a diameter of 1.0 mm for triangles and squares, and 1.8 mm for circles is specified (p. 109). For text labels, a minimum size of 7 pixels (1.4 mm) is recommended for horizontal text direction and sans-serif fonts, while for curved and rotated labels a size of 10 pixels (2.0 mm) is recommended (p. 104). Brunner (2000, 2001) assesses the size of a pixel to be in the

range of 0.19 – 0.30 mm for common screens available at the time, including laptop computers, and derives similar conclusions than Malić for minimum dimensions of graphical primitives (see Section 3.3.3.3).

Based on these investigations of basic geometric elements and available screen resolutions, Neudeck (2001) proposes more differentiated guidelines for screen-based cartographic symbology. The possibility of antialiasing is discussed in detail, highlighting its potential to largely improve the appearance of text and complex details, but also lamenting the introduction of a blurred appearance of sharp boundaries, fine lines and small text (p. 52f). Due to the differences in pixel size across displays, Neudeck advocates for the specification of symbology sizes in pixels instead of metric units (p. 62), and specifies a minimum size of 10 pixels for filled circles (which may be reduced to 8 pixels if antialiasing is used), 6 pixels for squares, and 15 × 15 pixels for iconic map symbols (10 × 10 pixels with antialiasing). For lines, a minimum width of 2 pixels is recommended, which can be reduced to 1 pixel if antialiasing is available (p. 63). A line separation of 2 pixels is mandated under any circumstances, as a smaller gap can lead to fusion of lines into a single shape (p. 66). Having established the physical pixel as the minimum unit of differentiation, Neudeck hypothesizes that once display resolutions have reached 300 ppi, the guidelines established for printed maps could be applied for screen-based presentation (p. 62).

The Swiss Cartographic Society also adopted a pixel-based approach for their specifications of minimum dimensions of screen-based maps. In the same publication containing the latest iteration of the society's recommendations for printed maps (see Section 3.3.2; Figure 3.17), recommendations for screen-based maps are given in a similar manner. A minimum size of 5 pixels is recommended for various simple geometric shapes, and a minimum width of 1 pixel and minimum separation of 2 pixels is mandated for lines (see Figure 3.18).

The guidelines established around the turn of the millennium, solidifying the intuition that significantly larger symbology needs to be used on screen-based maps, also had consequences for recommendations concerning the overall appearance of the map. Birsak (2000) notes that expectations commonly associated with maps at particular scales need to change, with the parameters of generalisation adjusted for screen-based maps. Users' expectations for digital maps to potentially provide global coverage of many related layers of information may require a larger pool of map symbols, which, however, need to be reliably discriminated at the lower resolution. At the same time, the permanent display of a legend may not be advisable due to the reduced overall space available for the map. All in all, Birsak concludes that the limited resolution of the digital medium may call for more "catchy"[32] map design. Lobben and Patton (2003) state that the interface of a digital map should be "structured and consistent" in order to not divert attention from the map itself, and that map symbols may be explained upon interaction instead of providing such explanation by way of a static legend. They conclude that design guidelines for on-screen maps "may be distinctly different than those adopted for the printed map" (p. 55) and that the former may be more effective when designed with less graphical complexity. However, Ellsiepen and Morgenstern (2007) conclude that, overall, screens expand the possibilities of cartographic

---

[32] in German original: "plakativ"

design, and have the potential to increase the efficiency of cartographic communication by making use of interaction and animation.

Bildaufbau am Bildschirm				Bildschirm 1:1	Mindestgrösse	Mindestabstand	Bemerkungen
Punktsignaturen							
✚	✖	✚✚✚		+ ✕ +++	5 Pixel	2 Pixel	Kreuz
△	△			△ △	7–9 Pixel		Hohlform
☐		☐☐☐		▫ ▫▫▫	5 Pixel	2 Pixel	Hohlform
■		■■■		• •••	5 Pixel	2 Pixel	Runder Punkt
☐		☐☐☐		◦ ◦◦◦	5 Pixel	2 Pixel	Hohlform
Liniensignaturen							
▬▬▬▬				────	1 Pixel		ca. 0,3 mm
▬▬▬▬				════	1 Pixel	2 Pixel	Doppellinie
▬▬▬				═══	1 Pixel	2 Pixel	Doppellinie
～～～				～～～	4 Pixel		Mindestausschlag
■ ■ ■ ■ ■ ■ ■				········	2 Pixel	3 Pixel	Punktierte Linie

Figure 3.18 Guidelines on minimum dimensions of point and line symbology for presentation on screens, at maximum contrast. Excerpt from Schweizerische Gesellschaft für Kartographie (2002, p. 30).

The possibility of multimodal interaction is just one aspect in which the digital medium has advantages over printed paper. Printing technology does not, in fact, guarantee perfect reproduction, and potentially introduces artifacts on its own due to the behaviour of ink flowing on the paper, as has been mentioned in Section 3.3.2. While some authors have expressed the possibility that the guidelines for printed maps may eventually become applicable for digital displays with a high-enough pixel density (Neudeck, 2001; Lechthaler & Stadler, 2006) – a consideration of only theoretical value around the turn of the millennium, as digital display technology with sufficient resolution was not available yet –, none of the works reviewed for this thesis considered even in theory that digital displays could actually allow for smaller symbology dimensions and higher density of information than the printed medium.

3.3.3.1. The advent of web-based and mobile maps and the resulting "big mess"

At around the same time when cartographers started to develop detailed guidelines for screen-based map design, two developments resulted in even greater challenges for creators of digital cartographic products: the advent of the web as an important dissemination medium had the consequence that map creators would not know the exact capabilities and properties of the device that ultimately displayed the map to the user; and the increased availability of mobile devices posed the challenge of designing cartographic output for even smaller displays and coarser resolutions, compared to conventional computer screens.

Mobile devices became of relevance for cartographic output several years before the release of the first iPhone, which established the new class of devices known today as the smartphone. Pre-smartphone mobile phones connected to affordable GPS receivers, in-car navigation systems, and GPS receivers with built-in displays were among the novel devices allowing for location-based services and applications. Such application scenarios required graphical depiction of the user's spatial surroundings, despite their severe limitations of display size and graphical fidelity. The pixel count of these displays could be as low as 180 × 180 pixels, at a size of 45 × 45 mm (Nissen et al., 2003). Similar to maps designed for conventional computer screens, some authors saw adaptivity and interaction as key elements that could compensate for the shortcomings of early mobile map displays (Reichenbacher, 2001; Burghardt et al., 2005). However, many early devices did not support touch-based interaction, so maps had to be augmented with additional cues for interaction purposes, such as adding numbers or letters to "hot spots" that could be interacted with (Brunner-Friedrich & Nothegger, 2002), further cluttering the limited display space.

Some authors applied the minimum dimensions established by Neudeck in the context of wayfinding maps on early mobile devices (Brunner-Friedrich & Nothegger, 2002). Others called for much more simplified depictions due to mobile displays "not allowing for the map graphics we are used to from printed media"[33], drawing on the existing body of work on cartograms and other schematic representations (Brunner, 2002). Due to the small display area, multiple levels of detail and interactive zooming were seen as important features of mobile maps, which required suitable generalisation strategies which are rooted, among other parameters, in minimum dimensions of graphical elements (Burghardt et al., 2005; Cheung et al., 2009).

Several authors proposed items for a possible research agenda of cartography in the context of pre-smartphone mobile devices. Reichenbacher and Meng (2003) mostly focus on the user's context, and include some high-level questions related to the user interface, such as the interaction modalities offered or the use of highly abstracted cartographic visualizations, but do not include questions specifically targeting the quality of the mobile display. Zipf (2003) discusses the adaption of maps to the user context in mobile systems, and lists display size and resolution among the primary factors influencing the map graphics. In his PhD thesis on mobile cartography, Reichenbacher (2004) asserts that "for screen display in general hardly any studies have been conducted" (p. 43). For the question of minimum dimensions, the work

---

[33] Own translation. Original: „Die kleinen Bildformate ermöglichen jedoch [...] keine aus dem Printmedium gewohnte Kartengraphik." (Brunner, 2002, p. 103)

of Neudeck is referred to. However, Reichenbacher comes to the conclusion that such "minimal dimensions have to be much larger to ensure legibility on small screens" (ibid.), without further elaboration on whether this statement is motivated by the coarser resolution of this class of devices or by other factors. To improve mobile cartographic applications, adaption is established as the central paradigm, and display size and resolution are listed among the main technological factors which make adaption of the visualization necessary (ibid., p. 100).

As has been shown in Section 3.2, the years after the introduction of the original iPhone in 2007 have seen and explosion of display resolutions and form factors, creating a heterogeneous landscape of display devices. One field in which cartography always had to deal with such heterogeneous end-user devices with potentially unknown properties was in the dissemination of maps online, on the World Wide Web. To ensure reliable communication under these circumstances, many authors have advised map designers to use simpler symbology, assume lower resolution, use only few "web-safe" colours and minimize the file size of any assets used (van den Worm, 2001; van Elzakker, 2001; Cartwright, 2003; van Elzakker et al., 2003). However, most authors agree that some of the disadvantages of the medium can be made up by offering suitable interaction possibilities to let the user retrieve additional information on demand.

More detailed guidelines for designing maps for the web are proposed by Jenny et al. (2008), who argue that maps on the web "should be legible at a glance", the information depicted on them must be unambiguous, easy to remember and should instil trust in the map (p. 31f), and generally the density of information on maps disseminated through the web should be reduced (p. 40). The authors mention that of the compiled guidelines, "some were confirmed by user surveys" while others may be "self-evident" or based on established conventions and yet others "have never been scientifically verified" (p. 32) – unfortunately, however, the paper does not cite any empirical studies investigating in detail the legibility of cartographic symbology. A pixel density of 86–100 ppi is assumed for LCD monitors, which, by the time, were commonly used as desktop monitors. Due to the fact that display resolution "may vary by 20% or even more" (p. 36), map creators are advised to not use numeric indicators of map scale, but instead use a graphical scale bar. The recommendations for symbology dimensions are based on a model of 1 minute of arc for smallest graphical detail, which translates to a size of 0.17 mm at a viewing distance of 60 cm, which the authors note is "clearly smaller than the size of a pixel". The authors acknowledge the widespread availability of antialiasing (facilitated at the time by the Adobe Flash Player 9 plug-in), and generally recommend to use antialiasing where possible. This would also allow for the specification of dimensions as fractional pixels, which the authors routinely make use of for specifying symbology dimensions. For example, a distance of 1.5 pixels is recommended for the separation of areas or parallel lines. For point features represented by simple geometric shapes at maximum contrast, a minimum size of 6 pixels is recommended. For more complex map icons, the authors advise to use even larger sizes. The advice to use antialiasing is also consistent with advice given by Ware (2020, p. 464) to "antialias visualizations wherever possible", which was first included in the 2013 edition of his textbook on information visualisation.

The guidelines proposed by Jenny et al. did not yet address the availability of high-resolution desktop monitors or smartphones as relevant target devices for information disseminated through the web, and thus assumed relatively homogeneous pixel densities of map users' displays. The relevance of such new classes of devices is acknowledged a few years later by Muehlenhaus (2014) in his book on web cartography, in which several pages are dedicated to the discussion of screen resolution and pixel density. While the potential of higher resolutions to positively impact "visual clarity" is acknowledged, the need to adapt designs for a wide range of heterogeneous devices is seen as a "nightmare" and as creating a "big mess" for map designers (p. 70). As for practical advice how to adapt designs for various display characteristics, Muehlenhaus refers to using the possibilities of scripting and CSS to reformat map layouts "on the fly depending on the type of screen, resolution, or device your map user is viewing" (ibid.). Due to the continually changing specifications of devices, the author "dare[s] not to make a suggestion in print" (p. 69) for specific resolutions and recommended pixel dimensions of elements of the cartographic interface, referring the reader instead to websites tracking the usage statistics of most frequently used screen resolutions.

As has been shown in Section 3.2.2, Muehlenhaus' book was indeed published at a time when display resolutions started to diversify, but the development had not been stabilized yet, and universally agreed concepts for dealing with the resulting "big mess" were not yet developed. Few studies have tackled the issue of heterogeneous screen resolutions, and the challenge to provide useful guidelines to map designers, since then. Mańk (2019) investigated the legibility of point symbols and variants of line symbology on the smartphone of highest pixel density available at the time, the Sony Xperia Z5 Premium, at 806 ppi, as well as on a high-resolution desktop monitor at 217 ppi. The performance of participants on the high resolution phone was not compared to different display hardware, but to images rendered at a lower resolution presented on the same display. At the highest resolution, squares and triangles could be distinguished at a success rate above 95% down to a size of 0.4 mm, and circles were identified at a success rate of 93% at the largest presented size of 0.6 mm. Similar success rates were accomplished on the high-resolution desktop monitor for circles of 0.6 mm and squares of 0.7 mm size. At a size of approximately only 2–3 pixels at conventional monitor resolution, these sizes are significantly smaller than what was commonly recommended by established guidelines. Unfortunately Mańk's study failed to accurately control the rendering of stimuli at lower resolutions, and also did not control participants' viewing distance or assess their visual acuity. The results show, however, that modern high-resolution displays have the potential to accurately reproduce symbology much smaller than conventionally assumed by existing guidelines.

3.3.3.2. Contextualizing the question of display resolution and minimum dimensions in contemporary cartographic research agendas

As of 2022, the recent development and increased dissemination of high-resolution display technology has not yet resulted in the publication of updated practical and comprehensive guidelines for map design. Gartner acknowledged already in 2009 that the development of useful cartographic guidelines for the design of screen- and web-based maps may only be at the beginning:

> Kartographen stehen erst am Beginn der Erstellung von neuen Gestaltungsrichtlinien für moderne Webkarten, die sowohl der interaktiven Nutzerumgebung gerecht werden als auch eine effiziente Vermittlung raumbezogener Informationen ermöglichen und ästhetischen Gesichtspunkten entsprechen. (Gartner, 2009, p. 286)

While Gartner's argument points mainly towards cartographers' understanding of the role of the map user as an active agent with influence in the map design process, a clear understanding of the properties of the map display medium would certainly form the basis for such improved guidelines for the design of digital cartographic products.

Arguably, the digital medium has in some aspects now reached, or surpassed, the fidelity of printed reproduction, which makes it timely to revisit the fundamental properties of mediated reproduction of both media. Buchroithner (2016) claims that printed maps still "radiate another aura" than their digital counterparts (p. 4). Although display resolution is not explicitly addressed in his essay, the accompanying figures show examples from printed maps with a high density of information – land cover, rock formations, contour–, and boundary lines – which would not have been possible to reproduce on a low-resolution digital display. Yet one can look at the PDF of the article on a high-resolution screen, and understand what the author refers to as the "aura" of printed maps, part of which may be related to a densely layered image of textures, lines and text, clearly outside of the parameters of design that have commonly been recommended for screen-based maps.

The ICA research agenda published in 2009 (Virrantaus et al., 2009) lists small screen size as one of the main factors limiting the usability of digital maps, and suggests continued research on visual perception of maps in order to derive a theoretical foundation for map design rules (p. 67). The issue of screen resolution is not explicitly addressed, but would fall well into the direction of research called for. In his contribution to a workshop on the future of cartographic research, Roth calls for an "update [of] methodological procedures and analysis techniques to study interactive, online, & mobile maps" (Roth, 2015, p. 2). In a later publication in a special issue on the future of cartographic research (Griffin, Robinson, et al., 2017), Roth et al. (2017) address the need to bridge the gap between studies in psychology, which often use very reduced stimuli or artificially limited viewing situations (e.g. very short timed exposure), and, on the other extreme, the evaluation of complete map designs in usability studies. The specific need of cartography to understand the properties of display devices, including screen size and resolution, and to compare results across devices and usage settings, is mentioned (p. 70). In order to make research better comparable and reproducible, the authors call for better reporting of the details of studies, particularly of the aspects of participants, materials, and procedures used. Finally, the authors call for the development of updated design guidelines that take into account "variable user and use case contexts" (p. 78). In the same issue, Griffin et al. (2017) stress the need to develop knowledge for designing maps that work across map use contexts, with the properties of the display device being an important technological factor constraining or facilitating particular map design strategies (p. 14).

Most recently, Roth (2019) calls for the development of design principles for "mobile first" cartographic design, a term borrowed from the discourse on web design, reflecting the fact that web content may today be primarily consumed on mobile devices (Xia, 2017). The further

development of symbolization techniques and guidelines developed specifically for mobile map design is listed among the most pressing needs identified. In a survey undertaken in the context of a recent proposal to develop an updated ICA research agenda (Meng et al., 2021), "visualization technologies" is listed as the top research topic of interest across ICA commissions, with "map design" ranked at fifth place. This may be read as underlining the need to continuously evolve academic engagement with technologies and principles of visual communication, but also to produce knowledge that is of practical use for creators of maps and cartographic visualizations, particularly for map use scenarios involving mobile devices.

3.3.3.3. Overview of proposed minimal dimensions for screen-based maps

Tables 3.2–3.5 compile the recommendations for minimum dimensions given by various authors for the depiction of point, line and area features, and textual labels on maps presented on screens.

Point Symbols

Pixel dimension	Metric dimension	Dimension	Source
12 pixel		Nautical symbols	Eaton (1993)
	2.6 mm	Side length of tourist map symbols	Morrison and Forrest (1995)
	1.2 mm @ 60 cm	Side length of triangle [34]	Malić (1998)
	1.2 mm @ 60 cm	Side length of square [33]	Malić (1998)
	2.1 mm @ 60 cm	Diameter of circle [33]	Malić (1998)
	3.0 mm @ 60 cm	Diameter of circle	Brunner (2000)
	2.0 mm @ 60 cm	Side length of square	Brunner (2000)
10 pixel		Diameter of circle	Neudeck (2001)
8 pixel		Diameter of circle (with antialiasing)	Neudeck (2001)
6 pixel		Side length of square	Neudeck (2001)
8 pixel		Side length of square (hollow)	Neudeck (2001)
10 pixel		Side length of triangle (with antialiasing)	Neudeck (2001)
15×15 pixel		Graphical icon	Neudeck (2001)
	2 mm @ 60 cm	Point symbol	Hake, Grünreich and Meng (2002)
5 pixel		Diameter of circle	Schweizerische Gesellschaft für Kartografie (SGK) (2002)
5 pixel		Square (hollow)	SGK (2002)
5 pixel		Circle (hollow)	SGK (2002)
6 pixel		Diameter of geometric symbol (square/circle/triangle)	Jenny et al. (2008)
	0.4 mm @ 30 cm	Side length (square / triangle) on 801 ppi display	Mańk (2019)

Table 3.7 Minimum dimensions for point symbols on screen-based maps, as specified by various authors. Values set in italics denote values derived from empirical studies.

---

[34] Malić produced elaborate tables of minimum dimensions for different screen resolutions and sizes. Here, the values for the highest resolution listed (1280x1024) at a screen size of 20 inch (diagonal) are extracted.

Lines

Pixel dimension	Metric dimension	Dimension	Source
	0.3 mm @ 60 cm	Line width [33]	Malić (1998)
	0.3 mm @ 60 cm	Line separation [33]	Malić (1998)
2 pixel	**0.4 mm @ 60 cm**	Line width	Brunner (2000)
	0.5 mm @ 60 cm	Line separation	Brunner (2000)
2 pixel		Line width (without antialiasing)	Neudeck (2001)
1 pixel		Line width (with antialiasing)	Neudeck (2001)
1 pixel		Line width	Hake, Grünreich and Meng (2002)
1 pixel		Line width	Schweizerische Gesellschaft für Kartografie (SGK) (2002)
2 pixel		Line separation	SGK (2002)
2 pixel		Line width (dotted line)	SGK (2002)
1.5 pixel		Line separation	Jenny et al. (2008)

Table 3.8 Minimum dimensions for line symbols on screen-based maps, as specified by various authors.

Areas

Pixel dimension	Metric dimension	Dimension	Source
	0.3 mm @ 60 cm	Area separation [33]	Malić (1998)
	10 mm² @ 60 cm	Area to distinguish color	Brunner (2000)
12×12 pixel or 5×20 pixel		Area size	Neudeck (2001)
16×16 pixel or 10×20 pixel		Area size (with outline, no antialiasing)	Neudeck (2001)
14×14 pixel or 5×20 pixel		Area size (with outline, with antialiasing)	Neudeck (2001)
	3×3 mm² @ 60 cm	Area size	Hake, Grünreich and Meng (2002)
2 pixel		Area separation	Schweizerische Gesellschaft für Kartografie (2002)
1.5 pixel		Area separation	Jenny et al. (2008)

Table 3.9 Minimum dimensions for the depiction of areas on screen-based maps, as specified by various authors.

Text labels

Pixel dimension	Metric dimension	Dimension	Source
	7 pt (≈1.4 mm) [35] @ 60 cm	Text height [33]	Malić (1998)
	10 pt (≈2.0 mm) [34] @ 60 cm	Text height (curved) [33]	Malić (1998)
	14 pt (≈ 3.6 mm) [34] @ 60 cm	Text height	Brunner (2000)
14 pixel		Capital letter height (without antialiasing)	Neudeck (2001)
10 pixel		Capital letter height (with antialiasing)	Neudeck (2001)
	10 pt	Font size	van den Worm (2001)
	12 pt (≈ 4.2 mm) [34] @ 60 cm	Text height	Hake, Grünreich and Meng (2002)
	12 pt @ 60 cm	Font size (rotated)	Schweizerische Gesellschaft für Kartografie (2002)
	12 pt @ 60 cm	Font size	Jenny et al. (2008)
	10 pt @ 60 cm	Font size (for fonts optimized for screen)	Jenny et al. (2008)

Table 3.10 Minimum dimensions for text labels on screen-based maps, as specified by various authors.

3.3.4. Conclusions from the review of cartographic literature

The guidelines on symbology dimensions that were extracted from the reviewed literature have already been presented in tabular form in sections 3.3.2.1 and 3.3.3.3. The following further conclusions can be drawn from the review of the related work in cartography presented above:

- Academic cartography has a long history of proposing guidelines for symbology dimensions, both for printed maps as well as for screen-based presentation. However, publications on the subject from the 20th century rarely cite empirical studies, or explain in detail the considerations from which those guidelines originated. Also, no publication was found in the course of the literature review that ever revoked or disputed a proposed guideline. Overall, the relationship of practical guidelines to the body of academic work in cartography seems to be weakly established.
- The question of minimum dimensions could be considered less a *persistent challenge* in cartography (Çöltekin et al., 2017) than rather a temporary blind spot in cartographic research, that may become apparent with each significant technological development in the realm of visual media. For the guidelines on printed maps, the history of academic publications on the subject can be read as indicating that just at the time when empirical investigation seemed relevant and feasible (with the development of automated, computer-based cartographic methods), other problems emerged that were considered more pressing, which resulted in such a temporary blind spot. For screen-based maps, the limited resolution of early computer screens was so apparent that spending much time on empirically assessing the obvious was probably not considered worthwhile. As display resolutions increased, the "big mess" caused by the heterogeneous device

---

[35] Conversion from point to metric sizes is taken from the authors, since this is not consistent between authors.

properties was probably found as obscuring the issue, again creating a blind spot that prevented systematic investigation for a while.

- For the purpose of empirical verification, it is not always well documented or implicitly clear what *exactly* the guidelines established for printed cartography refer to. For example, if the minimum separation between two parallel lines is specified as 0.1mm, what does this limit denote? Confusion of two parallel lines with a single grey line? The ability of map users to count the parallel lines? Preventing ink flowing together into a single fused region in the printing process? The limits of photographic reproduction techniques? Certainly, cartographers from the era of printed maps had a lot of knowledge and experience across the domains of perception, reproduction technology, cartographic techniques, and production tools. Often, this integrated knowledge was reproduced in a corresponding integrated fashion in cartographic publications of the time, which makes it hard to separate and analyse in detail those aspects today.

- Worth's attempt to reconcile "objective" research with the creativity and aesthetic considerations that are involved in all cartographic endeavours (cited in Section 3.3.2) is seen as programmatic also for the research proposed in this thesis. Even strictly empirical investigations, working with reduced stimuli and asking participants to solve artificial tasks, will involve creativity and many subjective decisions. Cartographic research of any school or paradigm should embrace the creativity facilitated and necessitated, instead of attempting to get rid of such "subjective" influences.

4. An apparatus and software framework for empirical studies of cartographic stimuli in controlled lab settings

Details on the practical aspects of research – laboratory equipment, device setup and software implementations – are often relegated to the appendices of scientific publications, if discussed at all. While such an approach may deliver a concise view of the results of a study to third parties, researchers trying to replicate findings of others often struggle to find detailed information on how and under which circumstances experiments have been conducted, or may not be able to conduct an identical experiment at all if important information on the practical aspects of the work is not revealed. Consequently, Roth et al. (2017) argue that particularly an interdisciplinary field like cartography calls for detailed reporting of method designs, materials and procedures. As Koenker and Zeileis (2009) put it: "the real challenge of reproducible […] research lies in restructuring incentives to encourage better archiving and distribution of the gory details of […] research" (p. 845). This is not the place to comment on whether the incentive structure of academic research has already changed to the desired effect. Nevertheless, this chapter shall shine a light on the "gory details" and discuss implementation aspects of the research that will be presented in chapters 5, 6 and 7.

The importance of the availability of source code for reproducibility has been outlined by Giraud and Lambert (2017). Together with the source code of all experiments and the underlying software framework, which will be released under an open-source license, it is hoped that the presented methodology may serve as a starting point for other researchers and students of cartography to conduct novel studies or replications of existing findings in a modern setting.

4.1. A software framework for controlled lab studies for cartography

4.1.1. Requirements and available software solutions

Based on the conclusions of the Chapter 3, the practical needs of the presented research project, and the author's wider experience with cartographic research and teaching, the following requirements for a contemporary state-of-the-art environment for conducting cartographic research have been formulated:

(1) Studies must be able to include a wide range of devices, including desktop computers, laptops, tablets and mobile phones, which are the most common display media for digital cartography. Less commonly used devices, such as outdated mobile phones, e-ink readers or VR displays should also be supported by any cartographic research framework. Studies with heterogeneous setups, incorporating multiple types of devices, should also be supported.

(2) A single experiment run may incorporate multiple devices even for a single user (e.g. to compare different display devices), or multiple users with potentially different roles (such as participant and supervisor).

(3) Visual output must be controllable down to individual hardware pixels, if the system software of the device allows for it. The system should support the specification of on-screen dimensions in physical pixels, real-world metric units, or angular units (if supplied with a pre-configured or interactively measured viewing distance), which should be accurately converted to corresponding screen pixel values.

(4) The framework must not assume a particular cartographic paradigm in order to allow the free development of new visualization and interaction techniques. Anything from "raw" rendering of stimuli using 2D or 3D graphics operations and primitives, up to well-known paradigms for interactive maps such as slippy maps should be supported.

(5) Devices should be able to re-join a running experiment after a crash or network outage. The overall state of an experiment should be explicit and should be modified upon actions of participants, supervisors, or in accordance with timed or scripted events.

(6) For the purpose of the planned studies, the framework should be tailored towards a lab-based setting incorporating multiple devices, including several devices for stimulus display, a device for entering participant responses, and a device for the experiment supervisor to monitor the progress.

There are many existing solutions that can be considered when implementing perceptual cartographic user studies. In the past, cartographers have adapted software from other domains for their purposes, or created software frameworks tailored to their specific requirements (e.g. Tainz & Weber, 1996; Roth et al., 2021). Software packages for psychological experiments, such as psychtoolbox (Kleiner, 2013) or PsychoPy (Peirce, 2007) support the accurate display of stimuli and recording of participants' responses. More recently, Roth et al. (2021) presented their software framework MapStudy, which is based on PHP and JavaScript. MapStudy is tied to the Leaflet library and therefore exclusively supports a "slippy map" map paradigm, and only supports one session per user. The evaluated existing software solutions do not support distributed or multi-device scenarios, and most do not run on mobile phones.

A. Robinson (2011) presented a series of case studies that demonstrate that the web is a powerful, versatile platform for conducting remote, asynchronous usability studies. He concludes that using off-the-shelf software to implement cartographic experiments is feasible, but further software development would be desirable to meet the needs of cartographic user studies. One advantage of web-based solutions is that they can be viewed on any device with a web browser, including older devices or e-book readers.

Many authors have used survey software such as LimeSurvey[36] or Google Forms[37] to conduct asynchronous online studies. These solutions allow the creation of survey-style online studies, the inclusion of graphical stimuli in the form of images or, for some products, interactive web content implemented in JavaScript, and provide powerful data collection facilities. However, it is difficult to control the precise dimensions of stimuli (Mańk, 2019), and experimental sessions spanning multiple devices are usually not supported.

---

[36] https://www.limesurvey.org/ - accessed 2022-04-01
[37] https://docs.google.com/forms/ - accessed 2022-04-01

4.1.2. Software implementation

Given that none of the evaluated software solutions fulfils the requirements identified in the previous section without significant modifications or extensions, a custom software framework to support the planned lab studies, *stimsrv*, was implemented. The stimsrv framework is informed by the requirements analysis presented above, and exhibits the following core features:

- The software is optimized for conducting local, lab-based studies, running one experiment session (one participant) at a time.
- Multiple devices can be part of the experiment, and devices can be heterogeneous, including desktop computers and mobile phones, running various operating systems.
- All components of the user interface – stimulus presentation, response collection, supervisor interface – are provided by default by a web browser on each of the participating devices. Custom clients (e.g. using native app frameworks) can be implemented if needed.
- Multiple devices can contribute to the overall user interface of the experiment – e.g. stimulus presentation can be done on one device, response collection on another device, and a third device can provide a control monitor for the experiment supervisor.
- A server component keeps track of the overall state of the experiment (e.g. current task and condition parameters, responses received etc.). Devices can (re-)join the experiment at any time, and will be supplied with the current state. Upon each response received from the participant, a new state is derived on the server and sent out to all participating devices, which update the user interface according to the received state. Such unidirectional data flow design based on centrally managed state is known to lead to simpler and more robust program structures (Madsen et al., 2020).
- An experiment is conceptualized as a sequence of *tasks*, defined in a central experiment specification file. For each task, an initial *condition* is stored in the experiment state, and rendered by the user interfaces of participating clients. Any participant *response* collected by a user interface component is sent to the server and processed to generate the next condition, which is again sent out to clients as part of the new experiment state and rendered. If, upon a response, no new condition is generated, the experiment advances to the next task, or ends if the list of tasks has been completed.
- Part of the overall state is information about the *context*, which can carry information from one task to the next. Tasks may modify the overall context, or read information from it. For example, a task testing the visual acuity of the participant could store the result in the context, with later tasks adopting their stimuli to match the assessed acuity of the participant.
- For each device, the user interface components to display (e.g. stimulus display, response user interface, etc.) as well as device-specific information (e.g. pixel density, display size etc.) is stored in the experiment definition file. This *device context* is available to the rendering component of a task, and can be used to adapt stimuli to the device context. For example, stimulus dimensions can be specified in metric units, and converted on each device to appropriate pixel values, using the information on pixel density defined in the device context. Therefore the rendering of stimuli does not have to rely on information on

device metrics (pixel density, display size, etc.) provided by the web browser or operating system, which may be inaccurate, but instead can be based on accurate and empirically verified measurements.

- Context, device context, conditions and responses are stored as plain JavaScript objects which can be serialized to JSON for transmitting between clients and server.
- To implement a stimulus or a user interface component, only two functions have to implemented in JavaScript: `initialize()`, which receives the context object (including context information specific to the device the component is initialized on) and returns a HTML DOM element, and `render()`, which is called with the current condition object as a parameter and is expected to update the component's DOM element, created earlier, accordingly. Standard JavaScript event handling can be used to capture participant input and generate responses which are sent to the server. The server in turn generates a new condition, which is sent to all clients' `render` function, or advances to the next task conditions for the current task are exhausted.
- All conditions, responses and context values are stored in a JSON file for each experiment session, including timestamps, for later analysis.
- While the framework is implemented using modern JavaScript features and APIs, older and very simple devices can be incorporated in an experiment by rendering the stimulus on the server using the Puppeteer library, which provides a "headless" web browser (Balabash & Oskin, 2020) which can be used to capture screenshots of browser content. The captured image can then be sent to a device with an outdated or primitive web browser for display, without requiring any JavaScript capabilities. If JavaScript is not available at all on the target device, the HTTP refresh mechanism is used to periodically refresh the web page and load a new rendered image from the server. This allows the incorporation of older devices for which the web browser cannot be updated any more, or devices for which only a simple web browser is available, such as e-book readers.

Figure 4.1 shows the flow of information in the stimsrv framework, with the server managing and distributing the central experiment state to multiple clients, updating the state upon receiving response data, and updating the clients with new state.

The stimsrv framework has been published under the GNU Affero GPL open-source license at https://github.com/floledermann/stimsrv, where more detailed documentation and links to example code can be found.

A series of case studies that demonstrates the versatility of stimsrv beyond the research presented in this thesis has been published elsewhere (Ledermann & Gartner, 2021).
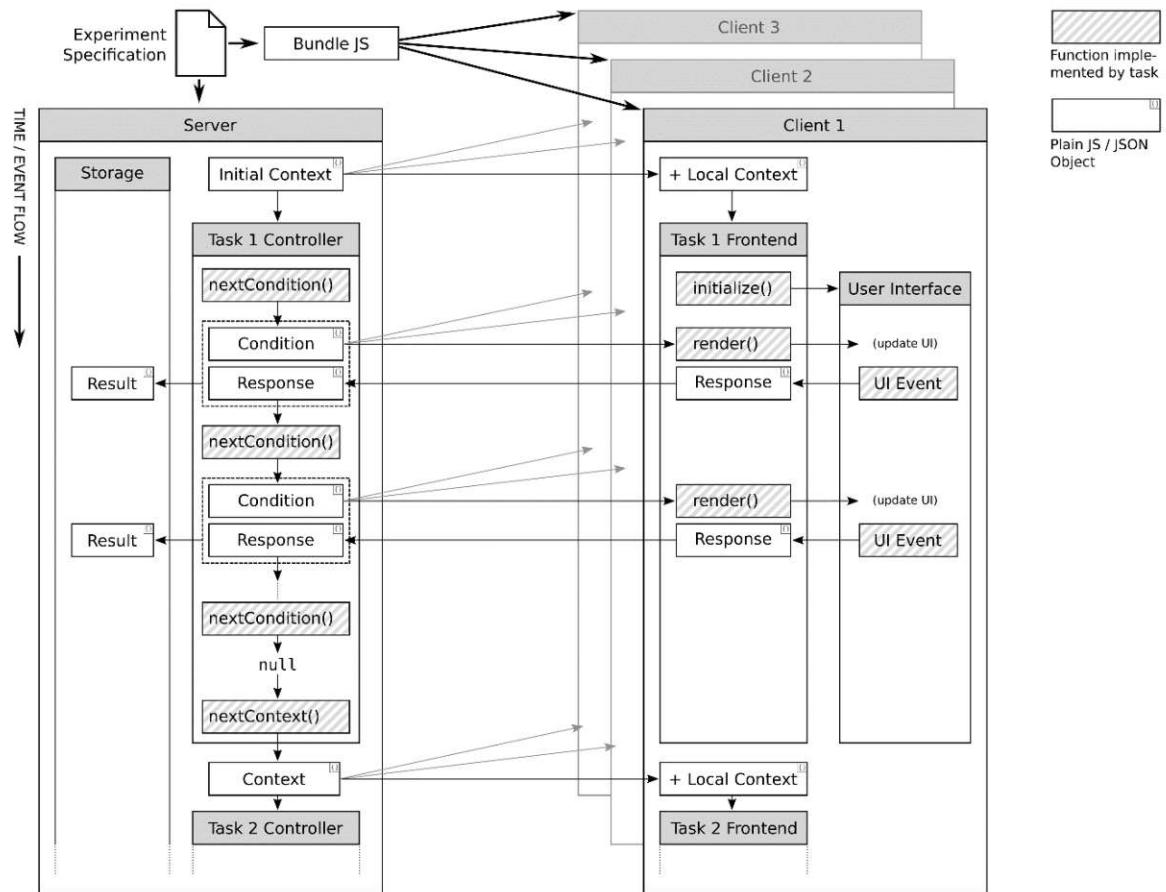
Figure 4.1 Flow of information and control between server and multiple clients in the stimsrv framework. Detailed documentation of the framework can be found in the stimsrv repository at https://github.com/floledermann/stimsrv

4.2. An environment for controlled lab studies for cartography

Researchers in cartography and GIS may not always have access to dedicated facilities for conducting controlled lab studies, and may not have the resources to establish such facilities for a single study. Therefore, a pragmatic solution has to be found that allows for the generation of reliable results within the available resources. This section describes the lab environment that has been adapted under such constraints for the studies presented in this thesis.

4.2.1. Environment

Psychological experiments are often conducted in an "experimental chamber" with black walls, controllable lighting and devoid of anything not part of the experimental apparatus (Cunningham & Wallraven, 2011). Setting aside a room for this single purpose for an extended period of time and making the necessary adaptions (which may not be compatible with alternative uses of the room, e.g. painting the walls black) will not be feasible for many

cartographic research groups. Therefore, minimum requirements for the experiment room have been formulated for the purpose of this study as follows:

(1) The room should be small to medium size, comfortably accommodating the furniture and apparatus needed for the experiment. Ideally, no other furniture is present in the room, but if this cannot be avoided it should be of neutral appearance (closed cupboards, empty desks etc.)

(2) The walls should be painted in a uniform, neutral, non-coloured way (white or grey or black) and devoid of any decorative elements (posters, paintings etc.).

(3) The ambient lighting must be uniform and controllable. As a minimum requirement this means that lighting conditions must be identical across all participants, ideally the light can be dimmed in a reproducible way to accommodate the particular requirements of the experiment. This also requires covering of any windows or glass doors with light-blocking material, in order to assure lighting independent of time of day and weather conditions. If the experiment involves multiple devices or screens, identical ambient lighting conditions need to be assured at each location.

(4) The room needs to be reserved for exclusive use during experiment sessions. Any doors should be closed and marked with "do not enter" signs, and any co-workers need to be informed in advance that the room is not available.

(5) Ideally, another room near the experiment room can be used for welcoming, briefing and debriefing of participants, including signing the participant consent form. This room needs to be set up with a desk and the necessary items for the paperwork (consent forms, information sheets, pens etc.).

(6) Ideally, the experimenter can sit in a room adjacent to the experiment room during the experiment, to be quickly available in the case of questions or problems, and to supervise the participant's progress if a monitoring facility if provided by the experiment apparatus.

(7) There should be a way to quickly ventilate the room between participants (e.g. unblocking and opening any windows) – this was of particular relevance during the COVID pandemic when the main part of this study was conducted.

The three studies described in chapters 5, 6 and 7 have all been conducted in the same room at the cartography research group at TU Wien, which was adapted in accordance with above requirements. The room measures 3.8 × 5.6 m, and has two doors (one to the foyer of the research group, one to another office room with a single desk in it), and one window facing a courtyard. All wall decorations (maps and posters) were removed, with exception of a permanently mounted whiteboard, which was outside the field of view of participants and was cleared before experiment sessions. Besides the furniture used as part of the experiment apparatus (see below), the room contained two cupboards, which remained closed during experiments. The window was blocked off using a sheet of 8 mm plywood covered with black "molleton" stage fabric for sealing any gaps, which could be quickly removed for ventilation purposes. The small office room next door was occupied by the experimenter during the experiments, and both rooms were reserved for exclusive use during experiment sessions.

4.2.2. Apparatus

All three studies intended to compare the legibility of cartographic symbols on mobile phone screens of different pixel densities in a within-subject repeated measures design. Consequently, multiple such displays have to be presented to each participant in sequence to display the stimuli in a controlled manner, and responses need to be collected by the participant. Mobile phones vary not only by pixel density, but also by display and device size, thickness, weight and other factors, which may suggest different technological "advancement" to the participant, a potential source for bias. In order to prevent such confounders influencing the results, and to ensure similar viewing conditions across devices, the following consequences were drawn: (a) participants should not hold the device in their hands, but only look at the display in a manner controlling for equal viewing distance and angle for each device; (b) an opaque bezel should cover the device, revealing only a square portion of the screen, identical in size for all displays; (c) responses to a the presented stimuli should not be entered on the devices themselves (which is made difficult or impossible by measures a) and b) and also could result in fingerprints diminishing the visual clarity of the stimulus), but use a separate device for this purpose; (d) the display devices should be mounted in fixed, static locations, with participants moving from one display to the next as required, instead of participants or experimenter needing to handle the devices during the experiment.

To realize the experiment in accordance with these requirements, two custom furniture elements ("shelves") have been constructed that can be placed on top of standard office desks and provide physical support for mounting the displays at equal height. Each shelf is 940 mm wide and provides a vertical surface with a horizontal support rail at a height of 283 mm ± 1 mm above the table surface. For each phone to be included in the study as a stimulus display, an adapter module measuring 288 × 200 mm is constructed from multiple layers of corrugated cardboard. The material of the adapter module was carved out to accommodate the mobile phone at its centre, with the power supply cable directed to the back side of the adapter module (see Figure 4.2, left). Each shelf provides enough horizontal space to accommodate two such adapter modules, which are supported and accurately positioned vertically by the horizontal rail. The power cables are routed through holes in the shelf's vertical plane to the back side of the shelf, where a multi-port USB power supply provides power to the phones. Neodymium magnets were embedded in the adapter modules at fixed locations to hold the cardboard bezels in place. The bezels were cut out to the appropriate size (varying for each experiment) from thin cardboard (300 g/m^2), and thin metal strips were glued to their back to be held by the magnets in the adapter modules. This allows for quick placement and removal of the bezels (in case access to the full screen is necessary for device configuration or troubleshooting), and allows for adjustment of the bezel position with sub-millimetre accuracy.

Mobile phones used as stimulus displays need to be fitted only once to their adapter module, and can remain in place there permanently (if set aside for use in the experiment exclusively). These preparations allow for placement of the mobile displays on the shelves with an accuracy of about ±1 mm in all 3 cardinal directions (using cardboard or plastic shims as necessary between the rail and the adapter to adjust vertical position and levelling), allow for

quick swapping or reordering of displays (because of the identical adapter modules), and allow for the placement of bezels in front of each display with sub-millimetre accuracy. The four possible positions for mobile phone displays (two positions on each of the two shelves) were labelled "Station A" to "Station D" with labels printed on a laser printer and glued to the top part of the shelves.

To ensure identical, undisturbed viewing conditions across all four display stations, the shelves were placed on two regular office desks in parallel to the room's strip light source (see Figure 4.2, right). Measurement of incident light with an uncalibrated luxmeter (since absolute values were not highly relevant) showed 201 – 208 lux across the locations of the four displays (a ~4% variability). A standard office chair with rollers was provided for participants to sit in during the experiment, and participants were instructed to move from station to station as prompted on screen during the experiment.

Figure 4.2 Left: adapter module to accommodate a mobile phone for the study, made from corrugated cardboard. Right: The two shelves constructed for the experiment, each holding two adapter modules, with grey bezels revealing only a portion of the screen of each device. A chin rail with a curtain to prevent reflections is mounted in front of the shelves.

To assure consistent viewing distance and minimize disturbing reflections on the stimulus display, a horizontal wooden rail was mounted in front of the display stations, with a short "curtain" made from black molleton fabric attached to it (see Figure 4.2, right). The vertical position of the rail was set in such a way that when participants placed their chin on the rail, they would look slightly down onto the display, and the entire area of the display would reflect the black curtain. This means that viewing of the display was not exactly orthogonal, but participants would view the display at an angle of approximately 15°–20° degrees down from horizontal orientation (see Figure 4.3, left, for a schematic cross-section of the apparatus). This corresponds with research finding that in natural viewing situations, mobile phone users will adjust the angle of their device to minimize disturbing reflections (Kelley et al., 2006).

For entering responses throughout the experiment, a separate, untethered mobile phone was provided. Participants were instructed to take the response device with them as they moved from station to station, and to slip their hands underneath the rail with the curtain to hold the response device comfortably in their field of vision (see Figure 4.3, right).

On a separate desk orthogonal to the two desks supporting the stimulus displays, a normal desktop monitor (Samsung SyncMaster SA450, with a display size of 474 × 298 mm at a pixel density of 90 ppi) labelled "main monitor" was placed that provided initial instructions to participants and presented an initial questionnaire, which had to be answered using the response device. A Laptop (Lenovo ThinkPad T495 running Windows 10) was placed on a separate table outside the view of the participant, running the *stimsrv* server that all mobile devices connected to, and a web browser that ran in full screen mode on the main monitor, which was connected to the laptop with an HDMI cable. A local wireless LAN router provided connectivity for all devices independent from the university wireless LAN, to ensure low latencies.
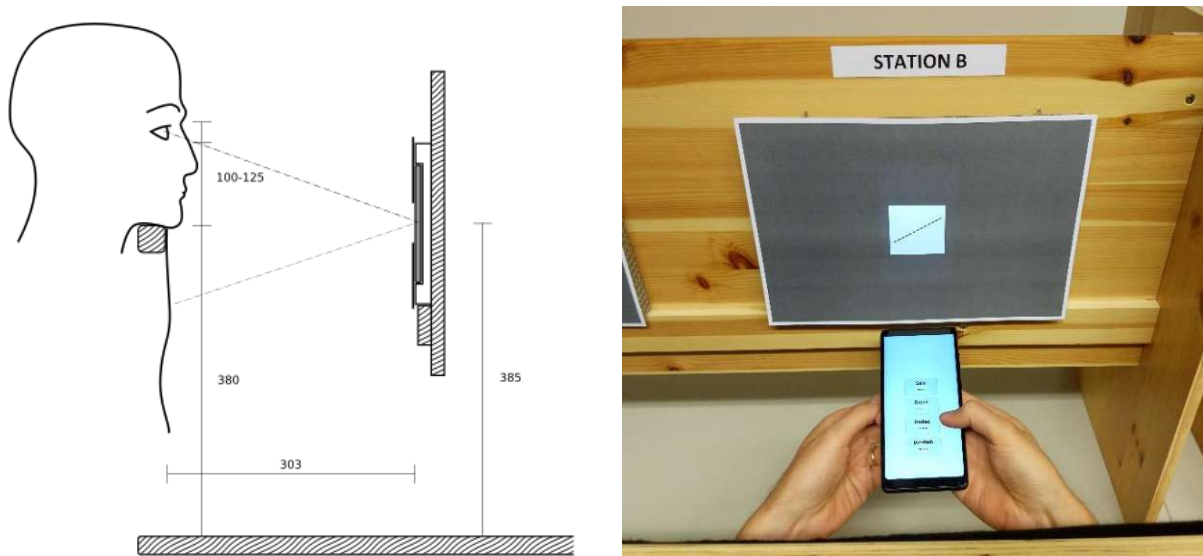
Figure 4.3 Left: Schematic cross-section of the experiment apparatus, showing the chin rail with attached curtain, supporting the participant's chin on the left, and the mobile device mounted on the vertical support plane of the shelf on the right (dimensions in millimetres). Right: Simulation of the participant's view onto the stimulus display at one of the display stations, holding the response device in their hands.

Due to the COVID pandemic, preparations had to be made to disinfect all surfaces between participants (see Section 4.3.5 below). A separate round table was used to hold the items necessary for this procedure (disinfectant, single-use gloves, paper wipes).

4.2.3. Device selection and preparation

Using commercially available mobile phones for an empirical study allows for easy and affordable access to a wide range of devices from the new and used device market. Using the web browser as a platform for rendering stimulus and response user interfaces, with the facilitation for supporting old and simple web browsers described in Section 4.1.2, allows for incorporation of a wide range of devices, including e-book readers and very old smartphones into an experiment. Nevertheless, devices with suitable properties as needed for a given study must be identified, acquired and prepared for incorporating into the overall experiment setup.

Two collections of data on mobile devices and their display properties (pixel density, size, pixel count) were compiled into a single spreadsheet for the purpose of getting an overview of available devices: pixensity.com, a crowdsourcing-based collection of 166 device specifications[38] and Wikipedia, where a crowdsourced list of 562 device specifications[39] was found. The consolidated list was sorted by pixel density, allowing for quick identification of phone models of the desired properties. Information of identified candidate models was verified by searching for the selected model on gsmarena.com and double-checking the information on resolution and pixel density found there.

At the outset of the study, some smartphones had already been acquired at our group for research purposes, including the Sony Xperia Z5 Premium, which at 801 pixels per inch[40] was the device with the highest pixel density in the study. Some older phones were donated by institute staff and their friends and families to create an initial pool of devices. However, devices for the study were not selected primarily by immediate availability, but based on the goals of the study, as will be discussed for each study in chapters 5, 6 and 7. Once a desirable pixel density had been identified for inclusion in the study, the spreadsheet was consulted to identify candidate models that most closely matched the desired target value. These phones were then searched for on willhaben.at, an Austrian used-goods online market. After purchase of a candidate phone, it was inspected for the presence of any visible scratches on the display, in which case it would not have been used in the study. Any protective foil was removed from the display, if present. Three of the five different devices used for stimulus presentation in the studies were acquired in this way from the second-hand market.

The study design requires the screen of all phones to be switched on and at constant brightness during the whole experiment. Not all versions of the Android operating system provide settings for keeping the screen on indefinitely. On phones where such a setting was not available, the "No Screen Off" app[41], which allows disabling the display timeout, was installed.

As the device for entering responses, the Xiaomi Mi Mix 2, taken from the pool of donated phones, has been used for all three studies. This phone has a pixel density of 403 ppi, at a display size of 68 × 136 mm. The response user interface was rendered in Google Chrome, which was put in full-screen mode on the response device. Because the experiment was run in a private LAN not connected to the internet for security reasons, Google Chrome would display a warning message reading "No internet connection" on top of the display even in full-screen mode. Since no software solution was found to suppress this message, the top 3 mm of the display were covered with black electrical tape to hide this message from participants to prevent distraction.

---

[38] https://pixensity.com/list/phone/ - accessed of 2021-07-07

[39] https://en.wikipedia.org/wiki/Comparison_of_high-definition_smartphone_displays - accessed of 2021-07-07

[40] As measured; the manufacturer's marketing materials state a resolution of 806 ppi, which was found to be inaccurate.

[41] This app was available at the time of the study preparation, but has since then been removed from the Android app store. A similar app seems to be available at
https://play.google.com/store/apps/details?id=com.sapphire_project.screenwidgetdemo (as of 2022-05-21)

For calibrating the brightness of the devices, an X-Rite i1 Display Pro display calibration sensor has been used. The i1 Profiler software that ships with the device can be used for acquiring absolute luminance measures by selecting "Monitor" > "Profiling" > "Density" > "Measure" > "Second Monitor", and then measuring a full white screen on the mobile device. This will report an absolute luminance value which can be used to calibrate all displays used for stimulus presentation, as will be described in Section 4.3.3.

A dummy experiment file for *stimsrv* has been created for each study for calibration purposes. When started, this script alternates between a full-white image (for calibrating brightness) and a square shape (for positioning the bezel) upon pressing a button on the response device.

4.3. A protocol for conducting controlled lab studies for cartography

For conducting the studies reported in chapters 5, 6 and 7, a number of checklists were prepared to ensure identical conditions and operations across all participants.

4.3.1. Preparing mobile phones for inclusion in the experiment

This checklist has to be completed once for every device that is prepared for being used as stimulus display in a study.

Required material:

- Mobile phone to be prepared for the study
- Charging cable for the phone with USB Type-A connector to plug into a regular USB power supply
- USB power adapter
- Corrugated cardboard for constructing the adapter blocks, approximately … for a 3-layered adapter block.
- Neodymium magnets, 2 × 2 × 2 mm cubes.
- Contact adhesive

Procedure:

- Remove any protective covers from the device.
- Remove any protective foil or additional layer of glass from the display.
- Clean the display using window cleaner. Inspect the display for scratches, dead pixels or other inconsistencies, and discard the device if any such display damage can be detected with the naked eye.
- Uninstall any non-essential apps. Update Google Chrome to the latest version available for the device.
- Remove all app shortcuts from the start screen. Add app shortcuts for Google Chrome and device settings to the start screen.
- Enable developer options for the device.
- Disable screen lock ("Security" > "Screen Lock" > "None")

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.

The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

TU Bibliothek
WIEN
Your knowledge hub

- Disable sleep ("Developer options" > "Stay awake") if available.
- Disable adaptive brightness ("Display" > "Adaptive Brightness" > Off)
- Disable "night light" ("Display" > "Night light" > Off)
- Disable screen timeout ("Display" > "Screen timeout" > "Never") if available.
- If disabling screen timeout is not possible, install the "No screen off" app and add to home screen.
- Check if the device runs stable and the display stays on over several hours when connected to a USB power supply.
- Prepare an adapter block for the device made from corrugated cardboard, as described in Section 4.2.2.
- At suitable locations, pierce holes into the adapter block and embed neodymium magnets, securing the magnets with contact adhesive.

4.3.2. Preparing the device assembly

This checklist has to be completed once when preparing the study, once all devices have been selected and prepared according to the previous checklist.

Required material:

- Mobile phones mounted in adapter blocks with USB charging cables (see above)
- Phone mounting shelves (as described in Section 4.2.2)
- Chin rail assembly (as described in Section 4.2.2)
- Grey cardboard for display bezels (as described in Section 4.2.2)
- Thin strips of ferromagnetic metal (e.g. cut from the lid of a tin can), approximately 20×20mm, 4 pieces per phone
- Shimming material of 0,1 mm – 0,5 mm thickness (plastic or cardboard)
- Contact adhesive
- Spirit level
- Multi-port USB power supply

Procedure:

- Place the shelves on empty office desks at the desired distance from the front edge.
- Place the adapter blocks containing the phones on the shelves, routing the USB power cables through the back of the adapter block and a hole in the shelf to the multi-port USB power supply mounted on the back of the shelf.
- Using a spirit level and shimming material placed under the adapter blocks as needed, make sure the display edge of each phone is level.
- From the cardboard, cut rectangular pieces to the desired bezel size.
- In the centre of each bezel, cut a rectangular opening of desired size.
- Glue the metal strips to the back of the bezels, at suitable locations so that they are centred on the magnets in the adapter block when the opening is exactly centred on the display.

- Check that each bezel is well seated, the central portion of the display is revealed and all four magnets engage with the corresponding metal strips when attached to the adapter block.
- Mount the chin rail assembly in front of the shelves. Measure the horizontal distance between the front edge of the chin rail to each display surface, adjust as needed for the desired viewing distance. Verify that each display is located at the same distance from the chin rail.
- Mark the positions of the shelves and the chin rail in a suitable way (e.g. tape) in order to be able to reconstruct the arrangement if inadvertently moved.
- Prepare a *stimsrv* experiment file for the purpose of calibration. When run, this should display in alternating fashion a full white screen for calibrating the brightness of the display, and a rectangular shape centred on the screen for calibrating the bezel position.

After this procedure, the physical arrangement of the experiment apparatus is complete.

4.3.3. Display calibration

Specialized devices for measuring luminance at arbitrary angles are expensive and will not be readily available at many cartography labs. As an alternative, for ensuring consistent *relative* luminance across devices without establishing an accurate absolute measurement, a workflow based on a digital single-lens reflex camera (SLR) and a common display calibration device is proposed.

Measuring display brightness with a calibration device such as the X-Rite i1 Display Pro that has been used in this study can produce an absolute measured value of display luminance. However, this measurement does not take into account the considerable variations in display brightness when viewed at an angle (Kelley et al., 2006). Since the proposed apparatus has participants view the device at an angle to prevent reflections, and the devices may vary widely in their angle-dependent brightness, absolute measurements of display brightness would not produce an accurate representation of the brightness seen by the participant.

Therefore, a calibration procedure using an SLR camera and a display calibration sensor is proposed here. In a fully dark room, the camera is put to manual exposure mode and pointed to the display with attached bezel, at the expected viewing angle of an average participant, and the camera's exposure meter is used to measure each display's apparent brightness at that angle. The brightness adjustment of the operating system is used for each display to result in an equal brightness value when viewed from the camera's position. Once all displays have been equalized in this way, the light in the room is switched on and the absolute luminance values of each display are measured with the display calibration sensor and noted down. These values serve as target values for subsequent calibration of the displays.

Required material:

- SLR camera which can be put to manual exposure mode, equipped with a varifocal lens (in this study: Canon EOS 60D)

- Display calibration sensor which can produce an absolute measurement of display luminance (in this study: X-Rite i1 Display Pro)
- Tripod
- Tape measure

Procedure:

- Set all phones to display a full-white image, and mount the bezels in front of them.
- Mount the camera on the tripod and configure the tripod in such a way that the focal point of the lens is approximately … cm vertically above the front edge of the chin rail, and the centre of the image in the viewfinder points to the centre of the first display.
- Adjust the focal length of the lens such that the display area revealed by the bezel covers approximately one third of the horizontal space in the viewfinder. If the camera supports multiple exposure metering positions, select the central position.
- Measure the horizontal distance between display surface and a distinct point on the tripod or camera (mark a location with tape if needed).
- Make sure the tripod can be placed in front of each display at exactly the same distance, adjusting its legs if needed.
- During measurement and adjustment, make sure that any windows are blocked, the doors to the lab are shut and the light is switched off, so that the displays are the only remaining source of light.
- Place the camera in front of each display, measuring the horizontal distance noted earlier, and use the exposure meter to measure the brightness.
- Use the brightness settings of the operating system to adjust each display's brightness, and measure iteratively until all displays show the same result in the camera's exposure meter.
- When finished, switch on the light to resemble the lighting conditions during the experiment.
- Measure the brightness of each display, and note down the absolute luminance values. These are the target values to which each display should be set before each experiment session.

An estimation of the accuracy of this method is beyond the scope of this research, as it would require a significant investment in the required special-purpose measuring equipment. However, given the available equipment it is the best solution that has been found for calibrating the brightness of the displays. Since only high-contrast stimuli will be used in the studies, and research has shown that contrast values of digital displays generally produce contrast far beyond the levels for which perception would be negatively affected (see Section 3.1.1), and subjective inspection of displays calibrated in the proposed way has confirmed a similar appearance of stimuli, it can be stated with some confidence that this method assures that contrast will not affect the outcome of the experiment. However, for studies investigating stimuli of very low contrast or in which colour plays an important role, more sophisticated calibration procedures may be required.

After completing the brightness calibration procedure, the experiment is fully set up to be run.

4.3.4. Running an experiment

Start of a lab session

The procedures described in sections 4.3.1 to 4.3.3 need to be performed once to prepare the incorporated devices for a study. During the course of the study, the experiment can be launched quickly and prepared for the next participant by following the procedures described in the next sections.

At the start of each "session" (a run of the experiment with potentially multiple participants, i.e. at the start of a day in the lab) the following procedure was followed:

Procedure:

- Remove all bezels from the stimulus display devices
- Switch on all mobile devices (stimulus displays and response device) and the laptop
- Connect the response device to the USB charger
- Connect the monitor HDMI cable to the laptop
- Run *stimsrv* with the calibration settings for the experiment
- Start Google Chrome on the laptop and enter the URL for the *stimsrv* server. Configure the role for the main monitor via the browser interface and put the *stimsrv* window in fullscreen mode on the external monitor (by pressing [F11]).
- Start the "No screen off" app on devices where it is needed.
- Clean the screens of all devices with window cleaning fluid.
- Launch Google Chrome on all devices, and enter the URL for the stimsrv server. Configure the role of the device and start the stimulus display via the browser interface.
- Block the window and switch the lights in the room to the setting that will be used during the experiment.
- Using the display calibration sensor, verify the brightness level of a full white display (as configured as part of the calibration experiment) to match the reference brightness values ±5%.
- Bring up the *stimsrv* calibration mode for centring the bezels. Attach the bezel to each display to centre on the display.

Before each participant

Procedure:

- Unplug the feedback device and put on the desk in front of the main monitor.
- Verify that all screens are clean and free of dust, scratches or streaks.
- Bring the chair back to the highest position.

Participant introduction and guidance

The consent forms used for the experiments can be found in the supplementary materials.

Procedure:

- Participants were sent a confirmation e-mail with their time slot, the address of the research lab, and a PDF copy of the consent form and the COVID-19 protection measures one day before their participation.
- The participant was greeted in the foyer of the research group, and asked to take a seat at a desk placed there on which the consent form was prepared for the participant's signature.
- The introduction by the experimenter was spoken freely, covering the following items in order:
  - Participation in the experiment is fully voluntary and the experiment can be abandoned at any time (course credits, if applicable, will be given for showing up on the date and are not conditional on participation or completion of the experiment)
  - Participants were asked about any viewing corrections they would use for reading or using their phone, and asked to use them during the experiment.
  - Participants were given a brief explanation of what they have to do during the experiment, and it was explained that the experiment involves graphics that are too small to read for anyone even with perfect vision, and that we ask from them to best guess what kind of graphics is displayed. When a figure is too small to be discriminated at all, a random button should be pressed.
  - Participants were asked if they had any questions at this time, and they were notified that they could contact the researcher during the experiment in case they had any questions or problems, or they wanted to abort the experiment.
- Participants were then guided to the experiment room and asked to take a seat at the chair.
- The experimented would close the door to the experiment room, and go to the next room to monitor the experiment.
- Further instructions on adjusting the chair, positioning themselves in front of the display stations, and the overall structure of the experiment were given to the participants on the main monitor (see supplementary materials for details).

After each participant

Procedure:

- Connect the feedback device to the charger.
- Disinfect the response device and the chin rail by wiping with disinfectant solution.

At the end of a lab session

Procedure:

- Switch off all devices
- Unblock the window

4.3.5. COVID-19 protective measures

The experiments were run in the period from March 2021 to July 2022, during which time mandatory protective measures due to the COVID-19 pandemic where instantiated. The experiment protocols were adapted to adhere to regulations mandated by local and university administrations and to provide maximum safety for participants and staff. The following additional measures were taken to adapt the experiment during the pandemic.

Participant invitation

- Participants were asked to not come to participate in the experiment if they felt sick or showed any COVID-19 related symptoms.
- Participants were asked to wear an FFP2-mask when arriving for the experiment, and to register with TU Wien's access management system.
- Participation was scheduled so that only the experimenter and one single participant were present at any time.

Before each participant

- The experiment room was ventilated by unblocking and opening the window and opening the window in an adjacent room.
- All surfaces potentially touched by the participant (desk surfaces, chair, response device, chin rail, pens) were disinfected.

Participant introduction and guidance

- A distance of 2 meters minimum was kept between participant and experimenter.
- Participants were provided with a disinfected pen to sign the consent form.
- Participants were advised on the protective measures in place.
- During the experiment, participants were on their own in the experiment room, with the experimenter sitting in a room next door to help with questions or problems. The participants were allowed to take off their mask during the experiment, as long as they were alone in the room. In the case of any difficulties requiring the intervention of the experimenter, both the participant and the experimenter would have to put on their mask before the experimenter entered the room.
- In the case of technical difficulties that could not be resolved while keeping a distance of 2 meters, the experiment would be aborted and the problems would be assessed after ventilating the room.

After each participant

- The window was un-blocked and opened for ventilation.
- All surfaces were disinfected in preparation for the next participant.

5. Study 1: Fundamental perceptual thresholds of cartographic stimuli on smartphone displays

The goal of the first study was to "cast a wide net" and generate initial empirical data on the legibility of various stimuli related to cartographic symbology at small dimensions. This study was also the main use case in the context of which the *stimsrv* software framework (as described in Section 4.1.2) had been developed.

5.1. Study design

The study adopts a basic psychophysical design, presenting stimuli of decreasing size using a staircase method of stimulus adjustment (Cornsweet, 1962), which has the advantage of converging relatively quickly to an individual's perceptual limit.

Stimuli were designed to be related to the fundamental cartographic symbology types of (categorical) point icon, dashed and parallel lines, and text labels. For each stimulus class, a set of multiple variants is created, from which the stimulus for a particular trial is chosen at random, and the participant is presented with choices representing all stimulus types of the class as alternatives, from which they are asked to select the one best matching the presented stimulus in an n-alternative forced choice (n-AFC) design. Participants will complete all trials at each of four different display stations, varying in screen pixel density, in a within-subject design. To control for practice and fatigue effects, the order in which participants are assigned to display stations will be randomized for each participant.

The independent variable is the display pixel density, modelled as an ordinal categorical variable defined by the phone display at a particular station. The dependent variable is the limit size at which reliable correct identification of the stimulus is still possible for any given participant.

5.1.1. Devices and experiment configuration

The study employed the software, setup and protocol that have been described in Chapter 4. Table 5.1 lists the phones that have been selected as display devices for this study, which will be referred to as D_1 – D_4 in order of increasing pixel density. The devices have been selected in a way that they approximately cover the spectrum of pixel densities that had historically been available for smartphones, and that each increase of pixel density approximates a factor of ~1.5. The Xiaomi Mi Mix 2 has been used as response device, with a pixel density of 403 ppi at a display size of 68 × 136 mm.

ID	Model	Year	Display Technology	Subpixels	Display Dimensions	Pixel Density[42]
D_1	LG P-970	2011	LCD	RGB	52 × 87 mm	228 ppi
D_2	Sony Xperia V	2012	LCD	RGB	54 × 95 mm	342 ppi
D_3	Samsung Galaxy Note 4	2014	OLED	PenTile	71 × 125 mm	522 ppi
D_4	Sony Xperia Z5 Premium	2015	LCD	RGB (ZigZag)	68 × 121 mm	801 ppi

Table 5.1 Devices used for stimulus display in study 1.

The bezels for the display devices (see Section 4.2.2) were produced from 300 g/m² white cardboard in format A3, onto which a uniform 50% grey colour was printed using a colour laser printer[43]. The bezels were cut to reveal a 48 × 48 mm square portion of the screen of each phone, as mandated by the smallest display in the study.

To prevent participants from hypothesizing about the display qualities, the displays were arranged in pseudo-random order, with D_2 mounted at station A, D_4 at station B, D_1 at station C and D_3 at station D (assignment of participants to stations during the experiment was also randomized, so participants would be instructed to move from station to station, after completing all tasks on a station, in random order).

The experiment apparatus has been set up with a horizontal distance of 303 mm between the inner edge of the chin rail and the display surfaces, resulting in a viewing distance of 31.5-33 cm to the centre point of each screen, at an angle of 17-21.6° down from horizontal.

In contrast to the procedure proposed in Section 4.3.3, for this initial experiment all stimulus displays were calibrated for equal luminance of a full white screen of 328 cd/m² ± 2% (as mandated by the maximum brightness of the display of least brightness D_3), without correcting for angular variation in brightness.

5.1.2. Stimulus and task design

Stimuli for the first study were designed to include a wide range of cartographically relevant graphical elements, presented to participants in isolation with the task of identifying a given stimulus among other stimuli of the same class. All tasks used an *n*-AFC forced choice design, with the alternatives to choose from being represented as on-screen buttons on the response device. The buttons contained a text label representing the stimulus type as well as an enlarged, prototypical rendering of a matching example (except for the text recognition task, as will be discussed below). Tasks 1-6 will be denoted as T_1-T_6 in subsequent discussion.

---

[42] Pixel density values have been derived from measurements of the actual devices, and may deviate slightly from information found elsewhere.
[43] The study was prepared during a lockdown due to the COVID-19 pandemic, during which only shops for essential necessities were allowed to open. Therefore, coloured cardboard could not be acquired for producing the bezels.

All stimuli were implemented using the *stimsrv* framework with canvas-based two-dimensional rendering. The source code for all stimulus classes is available in code of study 1 in the supplementary materials.

5.1.2.1. Task 1: Tumbling E's

The first stimulus class was not included for its cartographic relevance, but to assess the effective near visual acuity of participants. Since participants likely would be from different cultural backgrounds (see below), a culturally independent test of visual acuity was preferred, which ruled out using roman letters for the test. The "tumbling E" stimulus was chosen as a standardized, widely used test of visual acuity (Chang, 2017). Performance of the participant on the tumbling E task can be converted to their logMAR score to control for visual acuity for their performance on other tasks.

5.1.2.2. Task 2: Laterally structured line

The stimuli for this task were designed based on the hypothesis that multiple parallel lines cannot be reliably distinguished from a single wider grey line when display resolution or visual acuity is insufficient. For each trial, the stimulus geometry was chosen at random from three variants of laterally structured lines: three solid black parallel lines, four solid black parallel lines, or a solid grey line (see Figure 5.1). The black parallel lines were dimensioned at a width of $1/9^{th}$ of the width of the overall line, with the remaining space distributed evenly for the gaps between the lines. The grey line was rendered at an intensity value randomly chosen between 0.5 and 0.6, which the author perceived as resulting in similar brightness values as the parallel lines when viewed from a larger distance. For each trial, a single straight line segment of 50 mm length was presented at a randomized angle between 0 and 360 degrees.
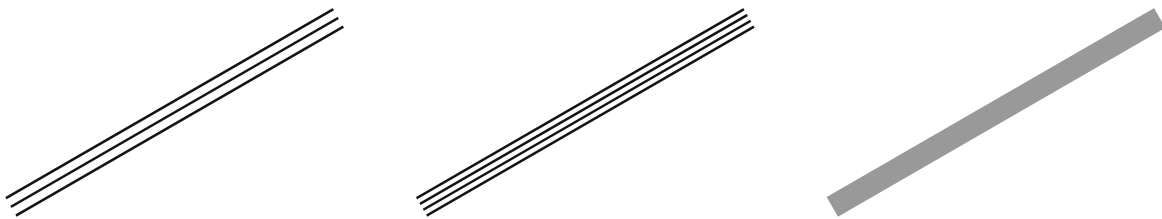
Figure 5.1 Stimulus geometries used for Task 2. Lines are shown at the original length used for the task (50 mm), at increased line width of 3 mm. Starting width for the task was 1 mm.

The response device presented the participant with three buttons, each showing a graphical representation of the line geometry, at horizontal orientation and a width of 2.5 mm, accompanied by a label reading "3 Lines", "4 Lines" or "Grey Line", respectively.

5.1.2.3. Task 3: Longitudinal line pattern

Dashed and dotted lines are commonly used on topographic and thematic maps, and many digital graphics environments allow the specification of dash patterns to create a variety of line types. For the stimuli of Task 3, dash and gap patterns were specified as follows, in correlation to line width *w*: a dotted line (*1w – 2w*), a dashed line (*3w – 1w*), a dash-dot line

(*3w – 2w – 1w – 2w*), and a solid black line (See Figure 5.2). For each trial, a single straight line segment of 50 mm length was presented at a randomized angle between 0 and 360 degrees.

Figure 5.2 Stimulus geometries used for Task 3. Lines are shown at the original length used for the task (50 mm), at increased line width of 1 mm. Starting width for the task was 0.28 mm.

The response device presented the participant with four buttons, each showing a graphical representation of the line geometry, at horizontal orientation and a width of 0.7 mm, accompanied by a label reading "Dotted", "Dashed", "Dot–Dash" or "Solid", respectively.

5.1.2.4. Task 4: Point symbols

For printed maps, guidelines for various point symbols based on geometric primitives, such as circle, square and triangle, have been given in the literature, as has been discussed in Section 3.3. However, geometric shapes are less frequently used in modern digital maps, for which the availability of vector graphics and icon libraries has allowed map designers to use iconographic map symbols for many purposes. Thus, it was considered of little practical use to test the legibility of geometric shapes which are arguably more rarely used as map symbols on digitally produced maps today. However, with a number of icon libraries available, each containing potentially hundreds of icons to choose from, the question arises which icons to select for a controlled user study? This question will be revisited in the second study, where a procedure to select icons from icon libraries based on similarity analysis is proposed.

For this first study, the "Auckland Optotypes" (TAO) icon set, developed in the context of ophthalmological research, was used. TAO is a set of 10 icons developed for testing the visual acuity of children, and has some properties that are advantageous for psychophysical research: The icons can be easily named and remembered by participants, have a similar "ink ratio" i.e. overall darkness, and have prominent features that are similar within subgroups of the overall set (Hamm et al., 2018). For example, both the "rabbit" and the "butterfly" symbols contain elongated features at the top, the "duck" and "car" shapes have a prominent feature at the top right, and several shapes feature a dent in the bottom centre.

Butterfly　Car　Duck　Flower　Heart　House　Moon　Rabbit　Rocket　Tree
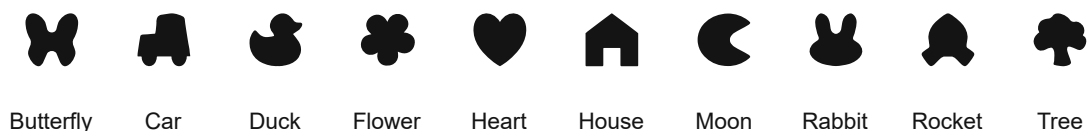
Figure 5.3 Symbols of the "Auckland Optotypes" set, used as stimuli for Task 4. Names of the icons were only shown as labels on the buttons on the response device. Icon shapes CC-BY-NC-SA Dakin Lab.

For the purpose of Task 4, the icons of the TAO set were rendered with a solid black interior against a white background (as shown in Figure 5.3). For full control of the rendering in a canvas-based environment, the SVG geometry of the TAO shapes has been converted to canvas path-drawing commands using an online converter[44].

For each trial, a single symbol chosen at random was shown at the centre of the stimulus display, scaled to target size. The response device presented the participant with 10 buttons, each showing one symbol at size of 3 mm and a label with the corresponding human-readable name of the symbol (see Figure 5.3).

5.1.2.5. Task 5: Point symbols, vanishing

For Task 5, a variant of the TAO symbols used in Task 4 was used, for which the outline of the symbol is drawn in two passes against a 50% grey background, first in white with a width of $2w$, then in black with a width of w (see Figure 5.4). Viewed from afar (or with insufficient display resolution), the symbol is expected to "vanish" into the background due to the overall intensity of the outline of 50% grey (Hamm et al., 2018). This stimulus requires the detection of high-frequency information, which may be relevant for detecting small details on a map or hollow symbols at non-optimal contrast.

Figure 5.4 The "Auckland Optotypes" symbols, rendered as a "vanishing" variant against a grey background, as used as stimuli for Task 5. Icon shapes CC-BY-NC-SA Dakin Lab.

For each trial, a single symbol chosen at random was rendered in the described way at the centre of the stimulus display, scaled to target size, against a background of 50% grey (RGB code #808080). The response device presented the participant with 10 buttons, each showing one candidate symbol, filled solid black against a white background (identical to Task 4, shown in Figure 5.3), at size of 3 mm and a label with the corresponding human-readable name of the symbol.

5.1.2.6. Task 6: Text labels

In contrast to normal reading, where information can be inferred form context and familiarity, it cannot be assumed that map readers are familiar with the toponyms present on a map, or that toponyms resemble dictionary words from any language. Therefore, words have been constructed artificially to serve as stimuli for a task related to map label reading, according to the following procedure:

- Letters and letter combinations that are easily visually confused are identified and grouped. For example: a/e, m/nn/rn, ff/fl/lf etc.

---

[44] http://www.professorcloud.com/svg-to-canvas/, accessed 2021-03-05

- Groups of artificial words are created, for each of which the words differ only by a few letters taken from the letter groups identified in step 1. Each word should sound believable as a toponym, but, ideally, not be a frequently used dictionary word.

Beier (2009, p. 63) gives an overview on existing research on frequently confused letters. As she mentions, most of these studies have been undertaken with typewriter or serif fonts, and results will differ widely for different fonts. Since a higher relevance of sans-serif fonts such as Helvetica was assumed for cartographic labelling (Jenny et al., 2008), the letters listed by Beier served only as a starting point to define letter groups based on the subjective judgement of confusability by the author.

Figure 5.5 Examples of stimuli used for Task 6, reproduced here at an enlarged letter size of 3 mm. Starting size for Task 6 was a letter height of 1 mm.

The following letter groups and word groups were used as stimuli for the text recognition task:

Group: a / e
- Kamao / Kameo / Kemao / Kemeo
- andarn / andern / endarn / endern
- Rasta / Raste / Resta / Reste

Group: m / nn / rn
- Lemos / Lennos / Lenos / Lernos
- Semato / Senato / Sennato / Sernato
- Kame / Kane / Kanne / Karne

Group: ff / fl / lf / ll
- Stoffen / Stoflen / Stolfen / Stollen
- Saffe / Safle / Salfe / Salle

Group: f / l
- Kofifa / Kofila / Kolifa / Kolila
- fokef / fokel / lokef / lokel

Group: il / li / ll
- Deila / Delia / Della
- Monail / Monali / Monall

Group: i / l
- Aiganei / Aiganel / Alganei / Alganel

The text labels were rendered in the "Roboto" font (a font available on Android devices which is similar to the sans-serif font families Arial or Helvetica), at a randomized angle between −60 and +60 degrees (with 0 degrees corresponding with horizontal text), in black with a

white outline against a 50% grey background (see Figure 5.5). For the chosen font, values specified as font size were found to result in capital letters having a height of 72.5% of that value. The starting font size of 1.4 mm thus resulted in capital letters of a height of ~1 mm for the initial trial.

The response device presented the user with one button for each of the alternatives from the same word group as the stimulus, shown in all uppercase letters and with increased letter spacing, in order to ensure participants had to correctly read the text in the stimulus and couldn't derive the correct answer from graphical similarities.

5.1.3. Presentation of stimuli

At each station, a participant would complete all trials of all six types of tasks before being instructed to move to the next station. While the physical arrangement of stations was identical for all participants, the order in which a participant was instructed to visit the stations was randomized. Participants started the experiment at the main monitor, where instructions were displayed and participants completed a short questionnaire with questions on age, gender, native language and self-assessment of viewing capability. The questionnaire and instructions also served to familiarize participants with the response device, which was used to answer the questionnaire and advance the instructions. Participants were then instructed by an on-screen message to move to the first (randomly selected) station to begin the main part of the experiment.

At each station, participants were greeted by a message confirming that the experiment should be continued at that station. The screens of stations not currently the target station were switched to black. Participants were instructed that they may take a short break and/or adjust the chair as needed before continuing with the first task. Upon confirmation, a very brief introduction was displayed on the screen of the station: "Press the button on the response device that best matches the shown graphics", before the first task was started.

At each station, an identical sequence of tasks was to be completed. Each task started with a stimulus that proved sufficiently large for most viewers in pilot experiments. Upon each response, stimulus size was adjusted according to a staircase method. For the initial sequence of correct responses, stimulus size was reduced upon each correct response in order to rapidly advance to a critical level for the participant. After the first incorrect response, stimulus size was adjusted upwards after each incorrect response, and downwards only after a sequence of three correct responses. After the first three reversals of adjustment direction, the adjustment of stimulus size changed to a smaller factor. After a total of five reversals, the task was completed, and the experiment continued with the next task. Table 5.2 shows the initial sizes and the coarse and fine adjustment factors for the staircase adjustment for each task.
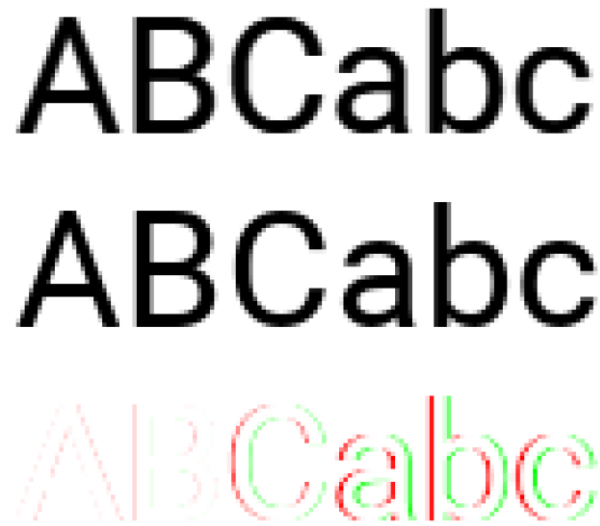
Task	Initial size	Coarse step	Fine step	Random angle
Tumbling E	1.163 mm	1.2	√1.2	-
Lateral Line Pattern	1 mm	1.2	√1.2	0 – 360°
Longitudinal Line Pattern	0.28 mm	1.2	√1.2	0 – 360°
Point Symbols	1.447 mm	1.2	√1.2	-
Point Symbols, Vanishing	2.315 mm	1.2	√1.2	-
Text Labels	1.4 mm[45]	1.1	-	-60° – +60°

Table 5.2 Initial size and staircase adjustment factors for each task.

For each trial of each task, the stimulus type displayed was chosen at random from the set of stimulus types. For tasks 1, 2 and 3, which utilize only a small number of stimulus types, randomization was implemented by drawing from a shuffled array containing two entries for each stimulus type, in order to prevent sequences of more than two repetitions of identical stimulus type, and to ensure similar frequency of each stimulus type at similar size levels for each participant. For the text recognition task, for each trial the list of words was chosen in sequence from the shuffled list of word groups, to ensure similar frequency of stimuli from each group.

For devices D_2, D_3 and D_4, stimuli were rendered by the web browser on the device using an HTML canvas. The browser on the oldest device in the study, D_1, could not be upgraded to a version that could run the *stimsrv* framework. Therefore, the mechanism described in Section 4.1.2 was used: the stimulus was rendered on the server using the Puppeteer framework, a "headless" version of the Chrome browser that can be controlled via software, and sent to the client as an image. The client displayed the updated image using the web browser available on the phone. It has to be noted that even though the same browser engine was used to render the stimuli for all devices (the Chromium engine is used by both the Puppeteer framework and the mobile version of Google Chrome running on the phones), the rendering of text for Task 6 for display D_1, which was performed on the *stimsrv* server and sent to the device as an image, differed slightly from the rendering on the Android devices. Despite the Roboto font being made available on the server by installing it on the computer, the server-side rendering resulted in different spacing of characters by about ⅓ pixel per character (see Figure 5.6).

---

[45] For the text label task, the size specified was the font size, which resulted in capital letters of 72.5% of that height.

Figure 5.6 Results of text rendered by two different rendering engines. Top: Canvas in Chrome version 90.0.4430 on Windows 10, middle: Rendering on Android, Google Pixel 2 phone, Chrome version 90.0.4430; Bottom: differences between the two renderings, with darker pixel values in Cairo coloured green, and lighter pixel value in Cairo coloured red.

5.1.4. Participants

Participants were recruited among students of the course "Web Mapping" of the summer term 2021. Bonus points amounting to 5% of the courses overall points were awarded as a compensation for the time spent travelling to and from the experiment. It was made clear that participation in the experiment itself was entirely voluntary (bonus points were awarded as soon as the participants arrived at the experiment location), and an alternative task to get the bonus points was offered to all students. As mentioned in Section 4.3.4, participants were reminded verbally as part of the experiment protocol that their participation in the experiment was entirely voluntary, that they could abort the experiment at any time and should do so if the felt unwell, and that the bonus points for the course had been assigned as soon as they showed up for the appointment and would not be rescinded.

30 participants made an appointment for taking part in the experiment, of which 27 showed up for their appointment. 14 identified as male and 13 as female on the questionnaire. 16 participants were aged 16-25 years, 10 were in the 26-35 years age group, and one participant was in the 35-45 years age group.

The course from which participants were recruited was taught as part of the "Cartography M.Sc.", an international study programme attracting a student audience from all over the world, therefore participants had a diverse cultural background. To control for familiarity with the Latin alphabet for the label reading task, a question on participants' native language was included in the questionnaire. 14 participants declared to have a native language using the Latin alphabet, while 13 participants had a native language using a different writing system, including Cyrillic, Chinese and Japanese.

Participants were asked to arrive wearing any vision correction (glasses or contact lenses) as they would normally use for reading a book or looking at their phone for an extended period. 13 participants wore some form of vision correction during the experiment.

5.1.5. Pilot

A pilot study was run with the author and three colleagues participating. During the pilot, the stimuli for laterally structured lines were adjusted to a narrower width of the internal lines, to achieve more equal difference between all three stimulus types. Unfortunately, this change was not activated in the experiment code until after the 3rd participant – therefore, data for participants 1-3 has to be discarded for this task.

5.2. Hypotheses

As mentioned in the introduction of this chapter, the goal of this first study was to "cast a wide net" and gather initial data on user's performance with respect to the proposed stimuli. Since no previous studies on the effects of display pixel density on map reading performance were found in the literature review, this study served to establish initial data on which subsequent studies could be based. Therefore, the hypotheses going into the experiment were formulated in a general manner to cover some initial assumptions.

A participant's individual limit size for correctly discriminating among the task's stimuli will be established as $l(D_i, T_j)$ for task j on display i. The overall mean of all participants' limits for the same task and station can be established as $L(D_i, T_j)$. Based on these metrics, the following hypotheses can be formulated:

(H1) An increase in pixel density is never detrimental to task performance. This means $L(D_1, T_i) \cong L(D_2, T_i) \cong L(D_3, T_i) \cong L(D_4, T_i)$ for all tasks T_1-T_6.

(H2) The pixel density of mobile phones has already surpassed the limit at which a further increase provides a measurable benefit for cartographic applications. Therefore $L(D_4, T_i)$ was expected to not be significantly lower than $L(D_3, T_i)$ for any task i.

(H3) As visual acuity is not normally distributed across the population (see Section 3.1.2.1), we do not expect to find participant's thresholds to be normally distributed. The statistical methods for analysing the data must be chosen accordingly, if this is indeed the case.

(H4) There are no significant learning effects (better performance at stations assigned later in the experiment) affecting discrimination performance within the course of the experiment.

(H5) There are no significant fatigue effects (worse performance at stations assigned later in the experiment) affecting discrimination performance within the course of the experiment.

5.3. Results

All 27 participants completed all six tasks at the four stations, with an average total of 606 trials (stimulus-response pairs) per participant. For each trial, the stimulus parameters (stimulus type, size, angle etc.) and the participant's response were recorded by the *stimsrv* server.

For each participant, their individual threshold for a task was derived by taking the smallest stimulus size for which at least 3 consecutive correct responses were given. For a task with n choices this results in a probability of $\frac{1}{n^3}$ of guessing 3 consecutive correct responses by chance. For the only 3-choice task in the study, T_2, this means there is a 3.7% chance that a participant exceeded their true threshold by one level by guessing; for tasks involving four choices or more (all other tasks), this probability is below 1.6%. The probability of successfully advancing to a lower threshold two times by random guessing is about 0.1% or lower for all tasks. While the thresholds derived in this way are not an exact assessment of the level at which a participant can reliably perform the task (deriving this level would require many more trials at "safe" levels, which is not facilitated by the staircase method), it should be at most 1 level off the level of reliable discrimination for all tasks and participants.

The resulting threshold levels for all tasks and participants are shown in Figure 5.7. Each dot in the beeswarm plots represents the threshold of one participant for a particular task and display station. Box plots show the median and 25% and 75% quartile values of the distribution of participants' thresholds.

To assess the visual acuity of participants, each participant's two best thresholds for task T_1 (tumbling E's) across the four display stations was used to calculate the participant's logMAR score. A participant's visual acuity was classified as "good" with a logMAR score of lower or equal to 0 (25 participants), and as "reduced" with a score lower than 0 (2 participants). Furthermore, visual acuity of participants with a logMAR score of <-0.25 was classified as "exceptional" (14 participants), while a score between -0.25 and 0 was classified as "regular" (11 participants). As the main interest of this study lies on participants with good visual acuity, the results of the two participants with reduced visual acuity were excluded from subsequent statistical analysis.
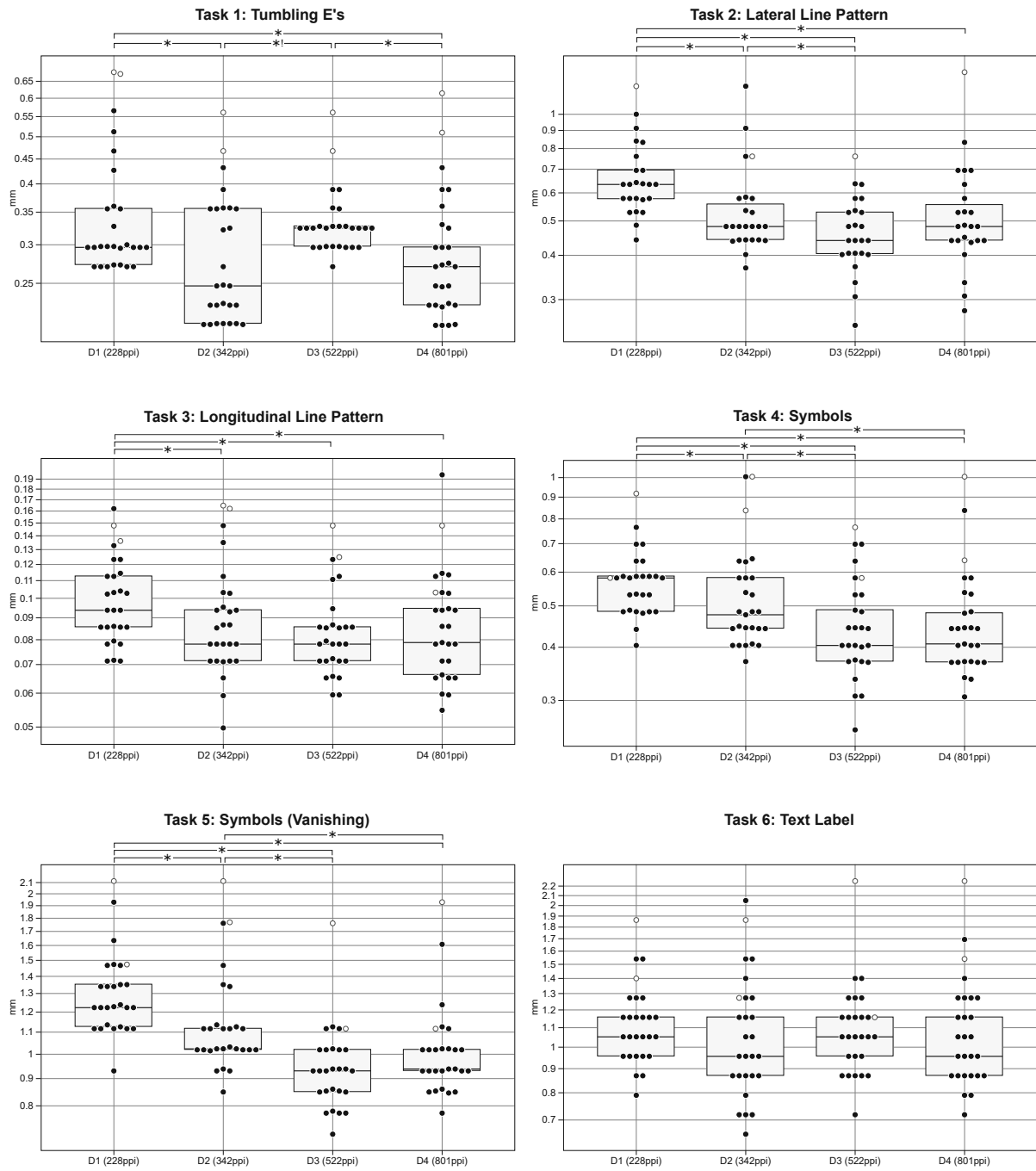
Figure 5.7 Resulting threshold values for all six tasks of study 1. Smaller values represent better performance. Results by participants with poor visual acuity scores are represented by hollow circles ∘ and were excluded from statistical analysis. ∗ denotes significant differences ($p < 0.05$) supporting H1. ∗! denotes a significant difference contrary to H1 (which only occurred for Task 1).

5.3.1. Statistical analysis

Psychophysical thresholds are assumed to not be normally distributed, which is reflected in hypothesis H3. To verify the distribution of the threshold results, a Lilliefors test for normal distribution was applied to the threshold data for each task and display. Assumption of normal distribution was rejected for 15 of the 24 result sets. After performing a log transformation of threshold values, the assumption of normal distribution was rejected for 8

of the 24 threshold distributions. Subsequent statistical analysis thus proceeded with methods suitable for results which do not require assumption of normalcy (A. Field, 2017).

The core hypothesis for this study refers to comparing the performance of the four different displays in the experiment. Friedman's ANOVA was used to compare the threshold distributions of all display pairs for all tasks. Pairs of displays for which a significant 1-tailed (due to the directionality of the hypothesis) difference of threshold distributions has been detected are indicated in Figure 5.7. Display D_1, the display with the lowest pixel density in the study, caused significantly worse performance of participants than D_2 for five out of six tasks (all tasks except T_6 text labels). Display D_2 was outperformed by participants on display D_3 in three tasks (T_2 lateral line pattern, T_4 point symbols and T_5 vanishing symbols), with no significant improvement for tasks T_3: longitudinal line patterns and T_6 text labels for this display pair, and significantly *worse* performance of D_3 over D_2 for task T_1 (tumbling E's). Display D_4 showed significantly better performance of participants compared to D_3 only for task T_6 (tumbling E's), with no significant difference in this task when compared to D_2. When compared to D_2, D_4 allowed for better performance in two tasks (T_4 point symbols and T_5 vanishing symbols). Notably, the text label reading task T_6 shows no significant improvement of performance on any display compared to any other.

Hypothesis H1 (performance on higher resolution is equal or better to performance on lower resolution) holds for five out of six tasks, but needs to be clearly rejected for the tumbling E's task T_1, for which $L(D_2, T_1) \ll L(D_3, T_1)$ (p < 0.001). It is also worth pointing out that a slightly lower mean performance of D_4 over D_3 could be observed for four of the six tasks, but this difference is not statistically significant to a point that warrants rejection of H1 for these tasks.

Hypothesis H2 (no significant improvement of the device with highest resolution D_4 over D_3) is retained for five out of six tasks, but needs to be rejected for T_1 due to the diminished performance of D_3 in this task. However, since participants' performance on D_4 is not significantly better than on D_2 for this task, and performance on D_3 was significantly worse than both D_2 and D_4, the cause for this can be attributed to reduced performance of D_3 instead of an increase in performance on D_4 for this task. H2 can be retained for all tasks except T_1.

Hypotheses H4 and H5 refer to the effect of the order in which participants were assigned to display stations. To control for such effects, participants' relative thresholds were calculated for each task and display by dividing their individual threshold by the geometric mean of all participants' thresholds. This relative threshold allows the different performance levels at different stations to be compared to each other. A Kruskal-Wallis test on the effect of station order on relative performance shows no significant differences (*H(3)=0.84, p=0.840*), thus H4 and H5 can be retained.

5.3.2. Discussion and further analysis of results

While the general trend of results mostly supports the hypotheses postulated in Section 5.2, results for particular tasks and display stations warrants discussion and closer analysis.

5.3.2.1. Limitations of the staircase method of stimulus adjustment

The staircase method for adjusting stimulus size allowed rapid convergence to participants' limits, but exposes some drawbacks when detailed analysis of results is attempted: not all stimulus sizes are shown to all participant an equal number of times on all display stations. Participants who perform well on larger stimulus sizes are only exposed a single time to stimuli at initial sizes, until the first incorrect response, with the staircase rapidly progressing to smaller sizes on each correct response, while participants who entered an incorrect response at a larger size (because of a lapse or because they may have already struggled to correctly identify the stimulus) were presented with three stimuli at each size, potentially more if repeated incorrect responses were selected. We can therefore expect error rates in raw trial data to appear inflated at larger sizes, because of the disproportionate number of trials presented to poorer performing participants. On the other hand, participants who fail to correctly identify multiple stimuli at larger sizes would never proceed to smaller sizes at all, because the task would be aborted after five reversals (i.e. three incorrect responses causing a reversal upwards). This means that the raw trial data for smaller sizes only includes the results of participants who reached those sizes at all – i.e. the better performing participants, while not including any information on participants that never reached those smaller sizes.

This means that any detailed analysis of results at particular size levels that have been generated using a staircase procedure will not be representative of the overall population, but disproportionally represent participants failing at those size levels. Furthermore, this inconsistent sampling applies to each participant at each station separately – a participant may have been presented stimuli at a certain size on one particular display station, but not on another, where previous incorrect responses may have prevented them from reaching the smaller sizes. Comparisons of trial data for different stations, even for the same individual, can therefore not reliably be made.

Despite these limitations of the chosen staircase method of stimulus adjustment, a provisional analysis of trial data may still provide valuable information for further studies, which is why it is included here. For these analyses, trials across a range of stimulus sizes will be grouped in size bands, and data from all four stations will be collated, in order to try to compensate for some of the limitations discussed above and reach larger sample sizes for each group. The limitations of the staircase method will be addressed in the second and third studies, where a different strategy for stimulus adjustment has been implemented that allows for the collection of representative data which can be compared across stations and participants.

5.3.2.2. Task 1: Tumbling E's

The most striking result in contradiction to the general hypotheses are those for task T_1, the tumbling E's, which has been included in the experiment not for its cartographic relevance, but as a well-established and standardized test to establish visual acuity of participants. In particular the comparably poor performance achieved at display D_3 when compared both to D_2 and D_4 warrants further investigation. So what could explain the diminished performance of participants identifying this simple stimulus?

A microscopic image of the stimulus at small sizes reveals the possible explanation (see Figure 5.8). On display D_2, displaying the stimulus at a size smaller than 0.37 mm will result in the graphics covering less than 5 pixels in horizontal and vertical extent. Due to pixel aliasing, this causes the middle bar of the E to spread across two rows of pixels at reduced contrast, resulting in the middle portion of the "E" to appear less dark than the 3 bars along its perimeter. While the stimulus is reproduced less truthfully, it allows participants to infer the orientation of the E by only looking at the outer three strokes, thus effectively making use of a much larger gap (approximately 2.5 times the size of the gaps between the bars of the "E") as an orientation cue. Therefore, performance of participants is amplified on D_2 by this aliasing artefact. D_3 also shows noticeable aliasing at this size level, but in this case it leads to the inner portion of the E to be filled in with pixels at medium intensity which, combined with the finer lines of the perimeter reproduced by the higher resolution display results in a more difficult to discriminate stimulus. Only on D_4 the stimulus is recreated somewhat truthfully.
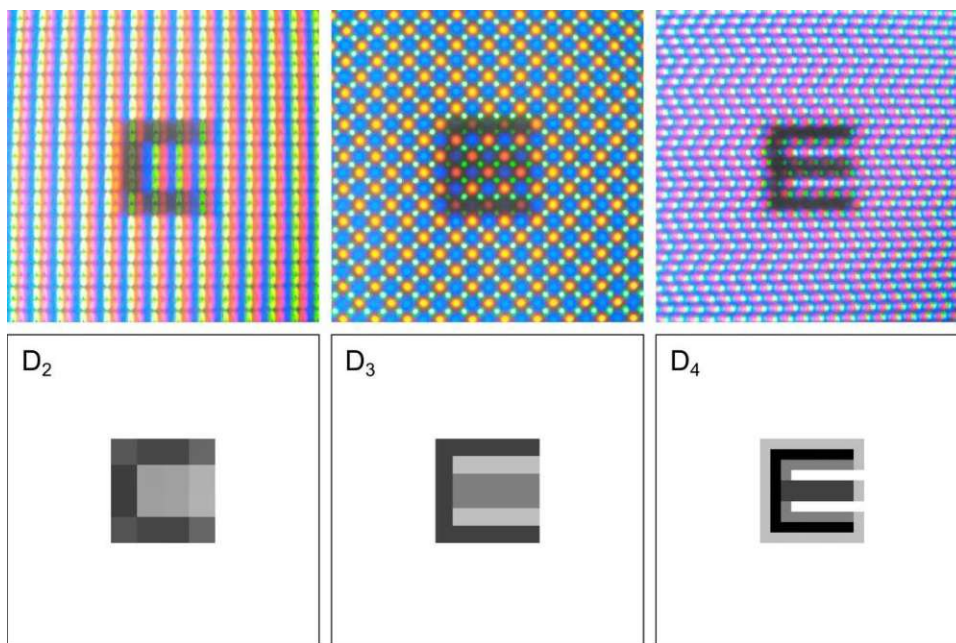
Figure 5.8 Microscopic images and bitmap patterns of stimuli for the "Tumbling E's" task at 0.27 mm stimulus size.

While the tumbling E stimulus is not of direct relevance for cartographic symbology, it shows that care has to be taken to avoid aliasing artefacts for graphics that involves lines parallel to the display's main axes. Also, an amplification effect through aliasing, such as observed on D_2, can be the reason why displays of comparably low resolution perform better than expected for some types of graphics.

5.3.2.3. Task 2: Lateral line patterns

Task 2 is the first task of direct cartographic relevance; therefore, a closer inspection of participant's performance is justified. It has to be stressed that because of the limitations of the staircase method discussed above, any detailed analysis of trial data has to be treated

with caution, as it will present a somewhat distorted view due to the irregular sampling of stimulus sizes, and comparison across display stations cannot reliably be made.

The analysis presented here are confusion matrices (matrices of stimulus/response pairs), which give for each type of stimulus presented (triple line / quadruple line / solid grey line) the relative frequencies of responses received. Figure 5.9 shows such confusion matrices for three size bands (results are collated from across all four displays).

response → condition ↓	Solid Line	3 Parallel	4 Parallel
Solid Line	99%	0%	1%
3 Parallel	1%	84%	15%
4 Parallel	3%	2%	95%

0.8mm – 1mm (408 of 2707 trials)

response → condition ↓	Solid Line	3 Parallel	4 Parallel
Solid Line	98%	1%	1%
3 Parallel	1%	79%	21%
4 Parallel	14%	3%	83%

0.6mm – 0.8mm (548 of 2707 trials)

response → condition ↓	Solid Line	3 Parallel	4 Parallel
Solid Line	99%	0%	0%
3 Parallel	2%	78%	20%
4 Parallel	37%	9%	54%

0.4mm – 0.6mm (1343 of 2707 trials)

Figure 5.9 Confusion matrix for stimuli of Task 2 in three different size bands. The caveats discussed in Section 5.3.2.1 apply.

For stimuli in the largest size band (0.8 – 1.0 mm), the data shows reliable identification of the solid grey line (99% correct responses), and a somewhat reliable identification of the quadruple line (95% correct). Closer inspection of the data shows that the misclassification of quadruple lines as solid lines is mainly an issue on D_1 (7%) and D_2 (4%). The overall data also shows a bias for misclassifying the triple line as a quadruple line (15% such misclassifications, with a 84% rate of correct identification), which can be observed in the data of all four stations. Readers are once again reminded that incorrect responses are expected to appear inflated at the larger sizes, due to the fact that well-performing participants will see fewer trials at those sizes.

In the confusion matrix for the next smaller band (0.6 – 0.8 mm), we see a similar pattern, with the solid line being identified at 98% accuracy, and the tendency of misclassifying the triple line as a quadruple line increased to 21%. This effect can be observed for all four stations, but is more pronounced at higher resolutions (32% misclassification at D_4, about 20% at D_2 and D_3, and 15% at D_1). Additionally, at this size frequent misclassification of quadruple lines as solid lines (at 14% such incorrect responses to the quadruple line stimulus) becomes

apparent, mainly driven by poor performance on D_1 (25% responding "solid line" to the quadruple line stimulus on that station). Again, the trial data will overrepresent those participants who submitted incorrect responses in this size band.

The smallest band contains the majority of trials (1343 of 2707 trials or roughly 50% of all trials), reflecting the fact that most participants made mistakes and therefore saw multiple trials with in the size range of this band. Again, 99% identified the solid line correctly at that size, but correct identification of the quadruple line fell to 54%, with 37% wrongly classifying it as a solid line (again most on D_1 with 50% misclassifications of that type, followed by D_2 and D_3 with about 36%, and D_4 with 29%), and 9% misclassifying it as a triple line. The triple line was identified correctly in 78% of trials, with 20% misclassifications as quadruple line and few (2%) misclassifications as solid line and no strong deviations from the overall trend on particular stations.

5.3.2.4. Task 3: Longitudinal line patterns

Similar to above analysis, confusion matrices have been calculated for Task 3, where stimuli involved various dash patterns (see Figure 5.10). At sizes of 0.15 – 0.20 mm (13% of all trials), all stimuli were correctly identified with ≥93% accuracy. However, already in this size band, a tendency to misclassify dashed lines as dotted lines (5% of responses to dashed line stimuli) and dot-dash lines as either dashed or dotted lines (6% of responses to dot-dash stimuli) is apparent. Upon closer inspection of the data, misclassification of dashed lines as dotted lines seems to be mainly a problem on D_2, whereas misclassification of dot-dash lines as dotted lines mainly happened on D_4. The caveats discussed above about over-representation of participants with poor performance applies.

At the next smaller size band (0.10 – 0.15 mm, 29% of all trials), performance for dotted and dot-dash lines remained largely unchanged, and accuracy for solid lines dropped slightly to 94%. Recognition of dashed lines however fell to just 56%, with a significant portion of misclassifications as dotted lines (33% of dashed line stimuli). This effect can be observed in the data for all four stations, with the strongest contribution of D_4 and D_2 with approximately 40% misclassifications, and somewhat less pronounced at D_1 and D_3 (with 29% and 24% misclassifications, respectively).

0.15mm – 0.2mm
(381 of 2864 trials)

response → condition ↓	Dotted	Dashed	Dot-Dash	Solid
Dotted	98%	1%	1%	0%
Dashed	5%	93%	1%	1%
Dot-Dash	3%	3%	93%	0%
Solid	1%	1%	1%	97%

0.1mm – 0.15mm
(819 of 2864 trials)

response → condition ↓	Dotted	Dashed	Dot-Dash	Solid
Dotted	98%	0%	0%	2%
Dashed	33%	56%	5%	6%
Dot-Dash	4%	3%	93%	0%
Solid	2%	4%	2%	94%

0.05mm – 0.1mm
(1328 of 2864 trials)

response → condition ↓	Dotted	Dashed	Dot-Dash	Solid
Dotted	81%	4%	3%	12%
Dashed	46%	26%	4%	23%
Dot-Dash	6%	9%	84%	0%
Solid	1%	2%	0%	98%

Figure 5.10 Confusion matrix for stimuli of Task 3 for three different size bands. The caveats discussed in Section 5.3.2.1 apply.

At the smallest size band (0.05 – 0.10 mm, 46% of all trials), recognition of solid lines remained high at 98%, but with a strong bias to misclassify both dotted and dashed lines as solid lines (at 12% and 23% of trials involving those stimuli, respectively, and all four stations contributing to the effect). Recognition rate for dashed lines collapses to only 26% correct responses at these sizes, with a strong bias to misclassify as dotted lines (46%) or solid lines (23%). This effect can be observed in the data for all stations, with correct recognition rates for dashed lines in the range of 25-30% at stations D_2, D_3 and D_4, and at only 17% at D_1.

5.3.2.5. Task 4: Point symbols

Due to the large number of stimulus types for this task (the 10 different shapes of the Auckland optotypes set), splitting the results data into size bands would result in only very few trials for each type and run the risk of individual outliers and lapses having a disproportional effect. Therefore, the results have been collated into a single dataset, and a single larger size band of 0.3 – 0.8 mm has been extracted for further analysis (with 73% of all trials falling within this size band). Figure 5.11 shows the confusion matrix for that data across all four stations.

Figure 5.11 Confusion matrix for Task 4 for trials at sizes 0.3 – 0.8 mm (73% of all trials).

For further analysis, the most often confused shapes have been ranked, and the five highest ranked confusions compared across displays. The four highest ranked erroneous responses (stimulus *rocket* → response *house* (30% of responses for that stimulus), *flower* → *heart* (21%), *duck* → *car* (16%) and *tree* → *heart* (14%)) were also found among the five highest ranked confusions for each individual display. The confusion ranked at fifth position overall (*house* → *rocket* at 13%) was present among the top five only at second rank for display D_3, with at a 25% confusion rate more than twice the rate of any other display (D_1: 8%, D_2: 11%, D_4: 9%). Among the top 5 ranked confusions on individual displays, only *flower* → *tree* on display D_2 was standing out to a similar extent, with a confusion rate of 16% twice higher than on other displays with each at a consistent value of ≈8% on D_1, D_3 and D_4.

5.3.2.6. Task 6: Text labels

Task 6 stood out in the statistical analysis of participants' thresholds, as no display was found to give participants a statistically significant advantage over any other display for this task. This contrasts findings for all other tasks, for each of with some pairings of displays with significant differences in performance were identified. A closer inspection of the results is thus warranted.

Stimuli for Task 6 were structured in a hierarchical way, for which groups of letters or letter pairs hypothesized to be visually similar were identified, and for each group, one or multiple sets of similar words were constructed. The choices offered as potential responses consisted only of the three or four words from the selected set of similar words, and not of the overall collection of words (50 words overall). There is therefore not a single confusion matrix for this task, but confusion matrices can only be constructed for each subset of similar words.

For a high-level analysis of the overall performance of participants, trials have been collated in two size bands, 1 – 1.6 mm (1549 or 60% of trials) and 0.5 – 1mm (846 or 33% of trials). For each group of letters, overall success rates at these sizes are shown in Table 5.3.

The overall analysis shows a consistent picture that the difficulty of the task varies with the letter group the stimuli are chosen from. Ranking of success rates of letter groups is consistent between larger and smaller sizes. It can also be seen that the success rate for the "il / li / ll" group and the "i / l" group are below 80% already for the larger size band. Further analysis of the data shows that this is also true for even larger sizes, with an overall success rate of only 66% at sizes above 1.5 mm for both groups (it has to be repeated, however, that poor performing participants are overrepresented in the trial data because of the irregular sampling of the staircase procedure).

Letter Group	1 – 1.6 mm	0.5 – 1 mm
a / e	92%	84%
f / l	92%	77%
ff / fl / lf / ll	87%	66%
m / nn / rn	82%	66%
il / li / ll	75%	56%
i / l	65%	56%

Table 5.3 Overall success rates for groups of similar letters and letter pairs for Task 6.

For further analysis, the success rates have been calculated and ranked for each station for each of the two size bands, which allows identification of deviations from the overall trend in rank (relative difficulty of correct identification of letters and letter pairs compared to other groups on a given display) and success rates (performance of participants compared to other displays). On display D_1, the group "m / nn / rn" stands out as being ranked a position 1 (highest success rate) on that display, while being consistently ranked at position 4 or 5 for the other three displays. In terms of success rate, the difference of that group is largest compared to D_2, which has a 17% lower success rate for that stimulus group, and D_4 with a 12% lower success rates. Other stimulus groups on D_1 are consistent with the overall trend. On all displays, all other letter groups are consistent with both the overall results and compared to each other, both in terms of ranking (at most a rank difference of one) and success rates (at most a 5% difference from the overall success rate).

In the smaller size band (0.5 – 1 mm), the highest variability in rank and success rate can be found for the "ff / fl / lf / ll" group, which is ranked at position 1 for D_2 at a 81% success rate, and at position 4 with a success rate of only 53% for D_3. Other stimulus groups also show larger variability at the smaller size. It has to be noted that participants' performance on D_2 and D_4 is consistently better for all rank positions than on D_1 in the recorded trial data.

The high variability of success rates across stimulus groups, and low success rates for some groups even at larger sizes may serve as an explanation why the statistical analysis failed to find significant differences between displays. The staircase method requires three correct responses to proceed to the next smaller level. With success rates of below 70% for the lowest-ranking stimulus group, and below 90% all but the two "easiest" letter groups for most stations, the probability for long streaks of correct responses required to reach smaller stimulus sizes could have been too low to allow for sufficient differentiation between stations. Any potential signal – the difference in performance of participants across stations – would have been drowned out by the noise introduced by the variation in difficulty of the tasks.

Success rates for stimulus groups of Task 6 at sizes 1 – 1.6 mm

	D1			D2			D3			D4	
1	m / nn / rn	90%	1	f / l	95%	1	a / e	94%	1	a / e	94%
2	a / e	88%	2	a / e	95%	2	f / l	92%	2	f / l	92%
3	f / l	88%	3	ff / fl / lf / ll	85%	3	ff / fl / lf / ll	90%	3	ff / fl / lf / ll	89%
4	ff / fl / lf / ll	86%	4	il / li / ll	77%	4	m / nn / rn	86%	4	il / li / ll	80%
5	il / li / ll	74%	5	m / nn / rn	73%	5	il / li / ll	72%	5	m / nn / rn	78%
6	i / l	64%	6	i / l	64%	6	i / l	63%	6	i / l	68%

Success rates for stimulus groups of Task 6 at sizes 0.5 – 1 mm

	D1			D2			D3			D4	
1	a / e	71%	1	ff / fl / lf / ll	81%	1	a / e	88%	1	a / e	94%
2	ff / fl / lf / ll	70%	2	f / l	80%	2	f / l	78%	2	f / l	80%
3	f / l	67%	3	a / e	78%	3	m / nn / rn	75%	3	m / nn / rn	69%
4	il / li / ll	65%	4	i / l	72%	4	ff / fl / lf / ll	53%	4	ff / fl / lf / ll	69%
5	i / l	57%	5	m / nn / rn	65%	5	il / li / ll	49%	5	i / l	54%
6	m / nn / rn	48%	6	il / li / ll	61%	6	i / l	43%	6	il / li / ll	53%

Table 5.4 Ranked success rates for stimulus groups of Task 6 at for the two size bands.

The inconsistencies in success rates and ranks of stimulus groups between stations could be explained by sampling-theoretical considerations. The high success rate of the "m / nn / rn" group on the display of lowest pixel density suggests that the gaps between words may have appeared enlarged (to at least a single pixel) on that display, making the task easier in a similar way as has been discussed and demonstrated above for the "Tumbling E's" of Task 1.

Despite the failure to detect statistically significant differences in participants' thresholds, the slight difference in mean performance apparent in the plot of thresholds (Figure 5.7), showing an advantage of both D_2 and D_4 over both D_1 and D_3, can also be identified in the detailed analysis conducted above. For the smaller size band, D_4 shows higher success rates than D_3 for all but one stimulus groups, and higher success rates than D_1 for all stimulus groups (with a dramatic difference of 23% for the highest rank). D_2 shows higher success rates than D_1 for all but one stimulus groups (and for all ranks), and higher success rates than D_3 for all but two stimulus groups (with dramatic difference of over 25% for the "ff / fl / lf / ll" and "i / l" groups) and all but one rank.

Three groups of letters have consistently appeared in the lowest ranks of performance across stations and sizes: "i / l", "il / li / ll" and "m / nn / rn". Cartographers may therefore be well advised to test legibility of their map labels for a particular design using these letter combinations.

5.4. Conclusions

The study presented in this chapter delivered initial insights into map users' ability to discriminate graphical stimuli related to cartographic symbology at various sizes, and how

such thresholds change with the pixel density of the display on which the stimulus is presented.

As expected, the display of lowest pixel density in the test, the LG P-970 at 228 ppi (D_1), corresponded with significantly poorer performance in five out of the six tasks. While this resolution is slightly below the lowest resolutions of devices manufactured today (see Figure 1.1 in Chapter 1), this finding is also relevant as it corresponds with a pixel density of ~110 ppi at a twice higher viewing distance, which approximates the viewing parameters of conventional desktop monitors. It should therefore be expected that the improvement of performance found in the study for increased pixel density of the mobile phone screen should also translate to a scenario of upgrading a conventional desktop monitor to a high-resolution screen of significantly higher pixel density.

The 342 ppi of the device with second-lowest pixel density in the study, the Sony Xperia V (D_2), correspond roughly with the pixel density of the iPhone 4, whose display was marketed as "retina display", suggesting that the capabilities of the human visual system are matched or exceeded with such resolution. In the study, performance of participants on this device was significantly lower than on a higher-resolution display for three out of six tasks, suggesting that the claim is not supported by empirical results. This is also reflected by the trend of phone manufacturers to produce devices with continuously higher resolutions after the release of the iPhone 4. However, the results of the "tumbling E's" task and the label reading task show that a lower resolution can, paradoxically, be of advantage for some particular stimuli, since gaps or small details in the graphics may be enlarged to a detectable size by the coarser pixel structure of the display. However, this effect has not been shown to appear for any stimuli not containing fine lines exactly parallel to the pixel grid, and therefore can be considered of little practical relevance for cartographic content.

Display D_3, the Samsung Galaxy Note 4, features a resolution near the top end of the spectrum of pixel densities of mobile phones marketed today, and its PenTile subpixel arrangement can be found in many modern phones. Participants showed significantly higher performance on this phone compared to the Xperia V (D_2) for three out of six tasks. But there appears to be a large caveat: performance for the "tumbling E's" task was significantly worse on this phone than on D_2, with the largest median threshold of all four displays in the study for this task, and the median threshold for the text label reading task was also larger than on D_2, although this difference has not been shown to be of statistical significance. While the "tumbling E's" task is not by itself directly relevant for cartographic applications, it shows that a set of stimuli exists for which performance on this display deteriorates in comparison to lower resolution displays.

The fourth device in the study, the Sony Xperia Z5 Premium (D_4), features a pixel density that is unmatched by any phone manufactured today (as of 2022). This suggests that phone manufacturers either did not see a marketable "selling point" in such ultra-high resolutions, or that the drawbacks of a device equipped in such a way (increased demand on memory and computation) outweigh its benefits. This is backed up by the results of the presented study, in which performance on D_4 was not significantly better than on the best-performing of the other three phones for all six tasks, and median threshold was slightly larger (although

not statistically significant) for three tasks. Paradoxically, this may be the result of a more truthful rendering of the stimuli, as has been discussed in Section 5.3.2.2: some graphical elements may appear enlarged, and therefore easier to detect by humans, on phones with lower resolutions. But D_4 is also the only display that did not perform significantly worse than any other phone for all six tasks, due to the irregularities in performance observed for D_3 for the tumbling E's task. This shows that when accurate stimulus reproduction is of crucial importance, the display with the highest resolution seems to be the recommended choice.

What do the results of the study mean from the perspective of the cartographer? Firstly, if information on the particular device or display resolution is not available, the results of D_1 can be used for deriving guidelines, as the resolution of this display marks the lower end of the spectrum of display resolutions available today. Where differentiation is possible, D_2 and D_3 are typical representatives of medium- and high-resolution devices, respectively. The poor results for D_3 for Task 1 may be discounted, since the effect was not replicated for any stimulus of direct cartographic relevance, and seems to be relevant mainly for narrow bundles of parallel lines in alignment with the display's primary axes. Care has to be taken, however, for designs that include such arrangements of lines as part of the design of map frame or legend.

The study presented in this chapter has some important limitations. One primary concern must be the choice of demographics for the study, mostly students at bachelor or master level in the age group of 20 to 35 years old. While this is not a representative sample of overall demographics, it is hoped that within this age group there are a sufficient proportion of participants with good visual acuity to produce valid results for such a particular demographic segment. As the medical research discussed in Section 3.1.2.1 has shown, optimally corrected visual acuity among adults can be expected to lie below 0 logMAR for healthy individuals until the retirement age of 65. This means that the range of participants' visual acuity among the study participants can be seen as representative for settings in which a controlled viewing situation, assessment of visual acuity of map users and optimal correction of visual acuity by means of glasses or contact lenses are generally available, as would be the case in many professional settings like military or emergency services. However, it has to be mentioned that the ophthalmological literature on habitual visual acuity (the visual acuity of persons in their everyday life) does not provide an unambiguous model, and habitual visual acuity of adults over 50 years old may need to be expected to be below logMAR 0. In any case, the author believes that settling the question of habitual or best corrected acuity in a comprehensive manner for all age groups in a way representing the total demographics presently lies outside the possibilities and expertise of cartographic research, at least without special funding for the purpose, as such study would have to involve serious recruiting efforts and a large participant count, without being of primary and direct relationship to (only) cartography. As a consequence, the presented study has established results only for the demographic segment of "good" visual acuity (logMAR < 0, which is also established as a threshold in the medical literature).

Nevertheless, the chosen strategy for recruiting participants can be seen as a weakness of the presented study. Researchers in cartography have recently called for participant recruitment

strategies that ensure a more diverse demographics (Roth et al., 2017), while at the same time acknowledging that for studies on visual perception, selecting students as primary user pool may be appropriate (ibid., p. 69). It is hoped that by making available the materials, procedures and software used for the study, researchers who see the need to verify or expand upon the presented results with a wider participant demographics find it feasible to do so.

The study also has some intrinsic limitations, some of which became apparent only once the results were gathered and a change of study design was not possible any more without the need to discard earlier collected data. The results of the "tumbling E's" task are worrisome, as they show that performance for a particular stimulus can be dependent on the alignment of the stimulus geometry with the pixel grid. To mitigate the issue and prevent the introduction of systematic bias like the ones that can be seen distorting the results for this task, stimulus position on the pixel grid can be randomized in the range of -0.5 to +0.5 physical pixels. This would not fully neutralize the issue (which is theoretically impossible, as sampling to the pixel grid is an inherent part of the presentation of stimuli on digital displays), but prevent the consistent reproduction of specific degenerate cases that make "guessing" the correct response significantly easier or harder for participants, as seems to have been the case for certain stimulus sizes for this task. Such random placement would also more realistically reflect cartographic rendering, in which on-screen position is often determined by transformation of real-world coordinates, and therefore not expected to align consistently with pixel boundaries.

6. Study 2: Minimum dimensions of cartographic point symbols on smartphone displays

In the results of the first study it became apparent that display pixel density had a significant impact on the legibility of point symbols at small sizes. However, some questions contributing to a fuller understanding of point symbol legibility remained unanswered.

Firstly, the "Auckland Optotypes" (TAO) icon set used in the first study was designed for applications in ophthalmology, not cartography. While the TAO icon set has some properties that appeared to make it a good choice for an initial investigation (similar ink ratio of icons, pairwise similarities, set of only ten icons), it is unlikely to be used in any cartographic application. It would be desirable to use icons designed for cartographic applications in an empirical study, in order to derive conclusions that are valid in real-world application contexts.

Secondly, the results of study 1 have shown that the confusion matrix for the TAO icon set is not symmetrical, and that some incorrect responses are more likely than others. Unfortunately, the sampling strategy chosen for the study (staircase adjustment of stimulus size) prevents a solid statistical analysis of overall trials, because each participant was potentially confronted with stimuli at different sizes. A different sampling strategy, ensuring that each stimulus size is presented an equal number of times to each participant, therefore allowing statistical analysis of overall success and confusion rates, would seem desirable.

Thirdly, we have seen that sampling effects of the pixel grid can lead to distortion of results, in some case diminishing success rates below the expected level, and in other cases artificially increasing success rates by enlarging small or high-frequency components of the stimuli. Both effects have been encountered in the results for the "tumbling E's" task in the first study, which by itself is not a stimulus of direct cartographic relevance. However, it cannot be ruled out that similar effects may influence the success rates of other stimuli, too. Therefore, strategies to prevent sampling effects biasing the overall results should be applied, and a better understanding of the impact of alignment fine graphical details with the pixel grid would be desirable.

As a fourth conclusion, a suspicion remains that the potential of the highest resolution displays was not exploited to its fullest by all stimuli and methods of presentation of the first study. The results have shown that, for example, the display with highest pixel density has an advantage over the others for some stimuli, but not for others. Also, the confusion matrices showed that confusion rates are not equally distributed for pairs of stimuli. This raises the question of whether the higher resolutions can be made productive by rendering the stimuli in a way that subtle differences are amplified, in order to improve performance for often-confused stimulus pairs.

Finally, the results of an investigation that presents stimuli in isolation at maximum contrast (a single symbol, rendered in black against a white background) with the sole task of discriminating one symbol from a set of others may have limited validity in real-world

application scenarios. Therefore, a task involving more realistic stimuli and going beyond a simple discrimination question should be included in the next iteration of the study to gain some insight into how findings transfer to scenarios more closely related to actual cartographic applications.

6.1. General considerations for point symbol legibility

The ideas and questions raised above shall inform the design of the second study, which should contribute to a better understanding of the perception of map symbols used to represent points of interest on a map. Before the particular study design is elaborated in detail, some general considerations on icon selection and presentation will be made in this section.

6.1.1. Candidate icon sets

One goal for the second study expressed above is to use icons which were designed for cartographic applications, instead of the generic shapes of the Auckland Optotypes used in study 1. Traditionally, cartographic guidelines have contained advice for the dimension of simple geometric figures (square, circle, triangle), but in the face of the now ubiquitous possibility to reproduce icons of arbitrary graphical complexity at no additional cost with the help of computers, and the availability of icon sets with hundreds of icons for map creators to choose from, it could be argued that such simple geometrical figures are of less relevance today. Hence, the icons to serve as stimuli for the second study shall be chosen from actual sets of icons available to map designers. The following requirements were formulated for icon sets to be considered for inclusion in the study:

- The icons of the set must be available as individual SVG files.
- Each icon file must only contain the icon graphics, no frames, ornaments or background colour.
- The icons must be monochrome, designed for full black-white contrast.
- The icons must be designed as filled shapes (not purely line-based).
- All icons in the set must be designed for being reproduced at roughly equal size (equal maximum extent of bounding box).
- The icons in the set should cover a wide variety of themes.
- The set should contain a large number of icons (about 100 icons or more).
- The set should be available under an open source or creative commons license.

The following icon sets were found by online searches and considered for inclusion in the study[46]:

1. OpenStreetMap map icons, available at
   https://wiki.openstreetmap.org/wiki/SymbolsTab (only the subsets for tag keys

---

[46] The availability of all online sources for icons was verified on 2022-06-23

"aeroway", "amenity", "emergency", "historic", "leisure", "man_made", "railway", "shop" and "tourism" were considered, a total of 229 icons)

2. Maki map icons, available at https://labs.mapbox.com/maki-icons/ (204 icons)
3. The US National Park Service NPMap Symbol Library (22px variant), available at https://www.nps.gov/maps/tools/symbol-library/ (297 icons)
4. AIGA symbols, available at https://www.aiga.org/resources/symbol-signs (68 icons)
5. Military map icons MIL-STD-2525 (US Department of Defense, 2014), available from http://www.mapsymbs.com/ (a system for constructing icons from components, therefore the number of icons is not determined)
6. Emergency Response Symbology from the Homeland Security Working Group, available from https://www.fgdc.gov/HSWG/ (209 icons)
7. "Freie Tonne" nautical icons, available from https://www.freietonne.de/index.php?site=31 (284 icons)
8. OCHA humanitarian icons, available from https://github.com/mapaction/ocha-humanitarian-icons-for-gis (361 icons)
9. Ordnance Survey 1:25 000 map symbols, documented at https://www.ordnancesurvey.co.uk/mapzone/assets/doc/Explorer-25k-Legend-en.pdf (39 tourist and leisure icons)
10. SJJB map icons, available from http://www.sjjb.co.uk/mapicons/contactsheet (300 icons)
11. European Road Signs, available at https://en.wikipedia.org/wiki/Comparison_of_European_road_signs (not counted)

Sets 1-3, 8 and 10 fulfilled all requirements stated above. Out of those, sets 1-3 were selected for further consideration in the study, because of the perceived relevance of icon sets 1 and 2 for online mapping, and the alternative theme of the NPS icons, which also include a large variety of concepts. In total, the three selected sets contain 731 icons designed for use on digital maps.

Figure 6.1 The 204 icons of the Maki icon set.

For the forced-choice design of the study, participants must be offered a limited number of alternative choices, preferably five to six to keep response times low and increase the chance of each icon to be shown at each size level. So which icons from the vast number of icons present in the three sets should be selected? In order to propose valid minimum dimensions at which the icons of the overall set can be discriminated, it would be desirable to present to participants the *most similar* icons from each set. If these can be reliably distinguished, it can be assumed that pairs of icons which are better differentiated can be distinguished as well. Therefore, a similarity analysis has been undertaken to identify subsets of similar icons from each set.

6.1.2. Icon similarity analysis

To identify sets of icons to use for the study, an analysis of the graphical similarity of pairs of icons within an overall set was required. A manual, subjective approach in which users rate the similarity of icons was briefly considered, but the idea was discarded due to the large number of icons to be analysed (for a set of n icons there are $n \cdot (n\text{-}1) / 2$ pairs of icons to analyse, which would result in 20 910 pairs for the smallest of the three sets), and the bias that human observers may have towards the icon's *semantic* similarity instead of *graphical* similarity. In any case, undertaking such manual rating of similarity and comparing with the automated approach described below would be a worthwhile endeavour for future research.

Multiple algorithmic approaches have been proposed for shape similarity analysis. Veltkamp and Latecki (2006) give an overview of different methods and their properties and conclude that a raster-based approach, in which shapes are rendered as a bitmap at a coarse resolution and compared cell-by-cell, performs well in comparison to other approaches, with the main drawbacks being its sensitivity to graphical transformations such as scaling, rotation and shear and the possibility of yielding a difference of zero for shapes which differ only by very small details. One main advantage of a raster-based approach is its robust applicability for complex and multi-part shapes. These properties appear favourable for the analysis of map icons, since the comparison should take into account only the default orientation, shapes are expected to be differentiated by sufficiently large geometric details if a high enough sampling resolution is chosen, and the icons are frequently made up of complex and multi-part shapes.

Lu and Sajjanhar (1999) present a particular raster-based approach to assess shape similarity, for which shapes are normalized for scale and rasterized to bitmaps of a pre-defined resolution (the authors report to use values of 12 and 24 pixels resolution), and the resulting bitmaps are compared by an XOR operation. For each pixel location, if the pixel at that location has a differing bit value for the two shapes, it contributes to the overall difference value computed for the two shapes. The difference value is then simply the number of pixels with an XOR value of 1.

The results of such raster-based shape difference analysis are shown in Figure 6.2 for three shapes from the TAO set. In an initial step, shapes are rasterized as bitmaps at low resolution – in this case, a resolution of 20 × 20 pixels was chosen. From these, further bitmaps can be derived by applying a binary operator on every pixel of the source bitmaps. In the example, bitmaps for subtractive difference (A-B and B-A) and absolute difference (A XOR B) have

been computed. A function *count(M)* returns the number of filled pixels for a bitmap *M*, which, when applied to the difference bitmaps, can be used as a simple metric for the difference of two shapes.
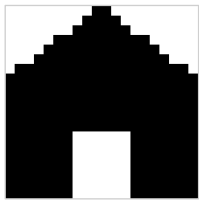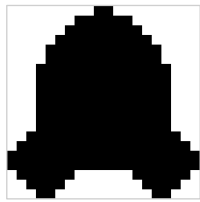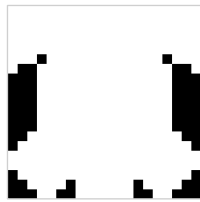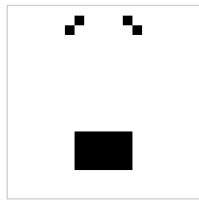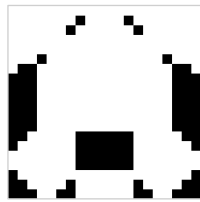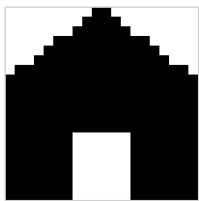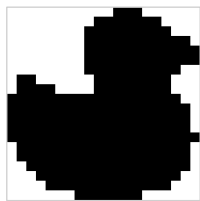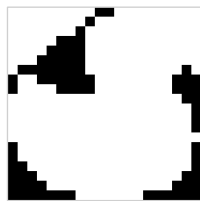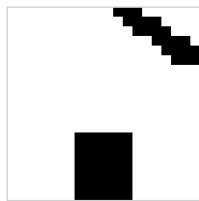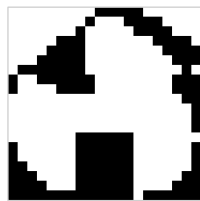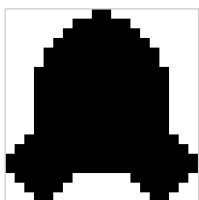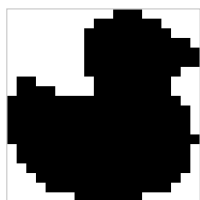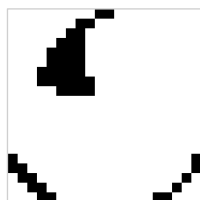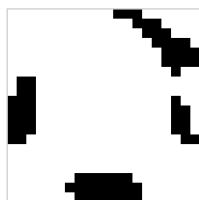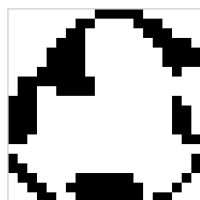
Shape A 20×20 bitmap	Shape B 20×20 bitmap	A – B (pixel count)	B – A (pixel count)	A XOR B (pixel count)
house	rocket	(66)	(28)	(94)
house	duck	(79)	(64)	(143)
rocket	duck	(47)	(70)	(117)

Figure 6.2 Raster-based shape difference analysis for pairings of three icons of the TAO set.

When applying raster-based shape analysis to arbitrary shapes, the alignment of the two shapes before rasterization crucially influences the result. Lu and Sajjanhar propose orienting the shapes in a way such that their major axis is parallel to the y-axis, scaling the shapes such that the two major axes are of equal length, and aligning the two shapes, which are now of equal width, along the top edge of their bounding box. For the purposes of comparing map icons, such approach appears suboptimal: if it is assumed that icons will be presented without rotation or non-uniform scaling, as seems reasonable for categorical map icons, the similarity of icon pairs should be computed without scaling or rotating them for a "best match", but with their original proportions and orientation as defined in the icon design. Furthermore, it can be observed that some map icons are composed of a base icon – e.g. the symbol for a "house" or a "car" – and an auxiliary icon, denoting some specific function or feature (see Figure 6.3 for examples). Such auxiliary icons can be placed below, above, to the side or inside the base icon. The most suitable alignment for assessing the similarity of two such composite icons is the one that aligns the base icons (because they are identical for both icons) and yields only the difference between the auxiliary icons (because these are the parts that are actually

different from each other). Depending on the placement of the auxiliary icons, the optimal alignment can vary for each icon pair.

Shape A	Shape B	XOR difference, aligned top-left (pixel count)	XOR difference, aligned bottom-centre (pixel count)
OSM / bicycle_parking	OSM / motorcycle_parking	(49 pixels)	(88 pixels)
Maki / car	Maki / car-rental	(203 pixels)	(68 pixels)

Figure 6.3 Examples for the effect of alignment on computed difference of composite icons. The upper icon pair yields smallest difference when aligned with the top-left corner, while the lower pair is optimally aligned at the bottom centre.

As a pragmatic approach that has proven well suitable for the icons in the selected test sets, the following procedure was implemented for determining the ideal alignment of an icon pair:

1. Compute the bounding box for both icons, based on a rasterized image of each icon at high resolution (e.g. 128×128 pixels).
2. Compute the offsets for both bounding boxes for the 9 cardinal alignments (left|centre|right horizontal alignment, top|centre|bottom vertical alignment)
3. For each of the 9 alignments, rasterize both shapes at coarse resolution at the given offsets, and count the number of filled cells for the XOR'ed raster images.
4. Use the alignment that yields the minimum difference as result.

Once the ideal alignment has been calculated for an icon pair using the procedure outlined above, the difference between the two shapes can be computed. The following requirements
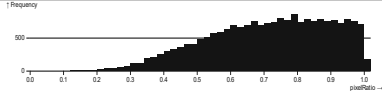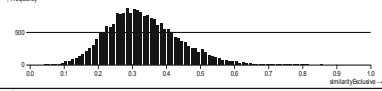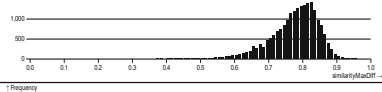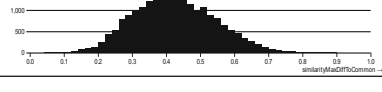
for a difference metric *d(A,B)* for two shapes *A* and *B* have been formulated (after Veltkamp & Latecki, 2006):

- Nonnegativity: $d(A,B) \geq 0$ (all differences are positive or zero)
- Normality: $0 \leq d(A,B) \leq 1$ (maximum difference value is 1)
- Identity: $d(A,A) = 0$ (the difference of a shape to itself is zero)
- Uniqueness: $d(A,B) = 0$ implies $A = B$ (if the difference is zero, the shapes are identical)
- Symmetry: $d(A,B) = d(B,A)$ (difference remains the same regardless of order of shapes in the pair)[47].
- Invariance under translation: For any geometric translations t_1 and t_2: $d(t_1(A),t_2(B)) = d(A,B)$

Given a pair of bitmaps *A* and *B*, basic raster operations (*A* - *B*, *A* AND *B*, *A* XOR *B*), a function *count(X)* that returns the number of pixels in a bitmap *X*, and optimal alignment of the two bitmaps as described above, the following basic metrics were computed for each icon pair:

- **commonPixels** = *count(A* AND *B)*
- **exclusivePixels** = *count(A* XOR *B)*
- **exclusiveArea** = **exclusivePixels** / resolution[2]
- **maxDiffPixels** = *max(count(A – B), count(B – A))*

From these basic metrics, the following normalized similarity metrics have been derived:

Similarity Metric	Formula	Histogram for Maki icon pairs
inkRatio	$\dfrac{min(count(A), count(B))}{max(count(A), count(B))}$	
similarityExclusive	$\dfrac{commonPixels}{commonPixels + exclusivePixels}$	
similarityMaxDiff	$1 - \dfrac{maxDiffPixels}{resolution^2}$	
similarityMaxDiffToCommon	$\dfrac{commonPixels}{commonPixels + maxDiffPixels}$	

For the purpose of finding sets of mutually similar icons that can be used as stimuli for the study, the similarity metrics for all icon pairs of each of the three icon sets were computed using a JavaScript application running in the browser. The SVG files of an icon set can be dragged onto the browser window, upon which the ideal alignments and similarity metrics are computed for all icon pairs and saved to a CSV file. Computing the similarity metrics for a set of approximately 200 icons (with ~20 000 icon pairs) for the best of 9 cardinal alignments of each icon pair took about 20 minutes on an Intel Core i5-6500 CPU with 3.2 GHz. The

---

[47] Veltkamp and Latecki note that symmetry of shape difference does not always reflect human perception, as has also been shown in study 1, where confusion matrices for TAO icons were not symmetrical. However, no clear and simple to implement model for computing asymmetric shape difference was found for the purpose of this study.

results were loaded into an online computational notebook for interactive exploration of icon similarities. In the notebook, all icons of a set were laid out using a force-directed graph layout, in which icons would repel each other, unless the pair met a similarity threshold, in which case they were connected by a line and subjected to an attractive force. The similarity metrics used and the threshold value was adjusted interactively in such a way that distinct clusters of similar icons would form in the graph. Based on this interactive visualization, clusters of similar icons were identified to be used as icon sets in the empirical study. Figure 6.4 shows the visualization for various threshold settings and some of the manually identified clusters of similar icons.

similarityExclusive ≥ 0.62, inkRatio ≥ 0.75, exclusiveArea ≤ 0.25
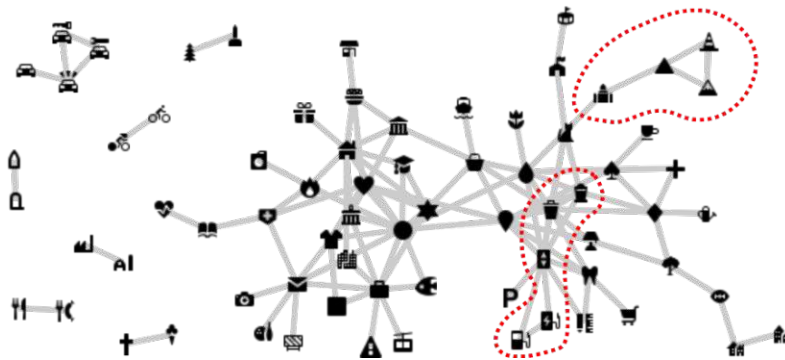
similarityMaxDiff ≥ 0.90, inkRatio ≥ 0.75, exclusiveArea ≤ 0.25
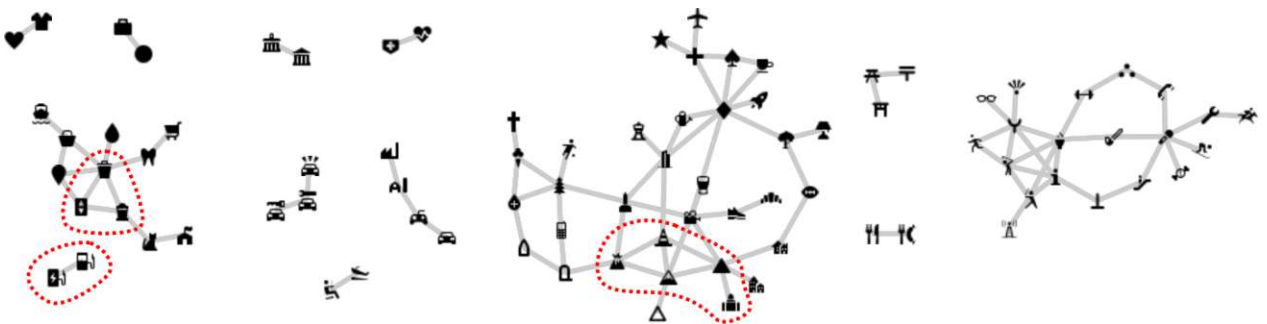
similarityMaxDiffToCommon ≥ 0.73, inkRatio ≥ 0.75, exclusiveArea ≤ 0.25
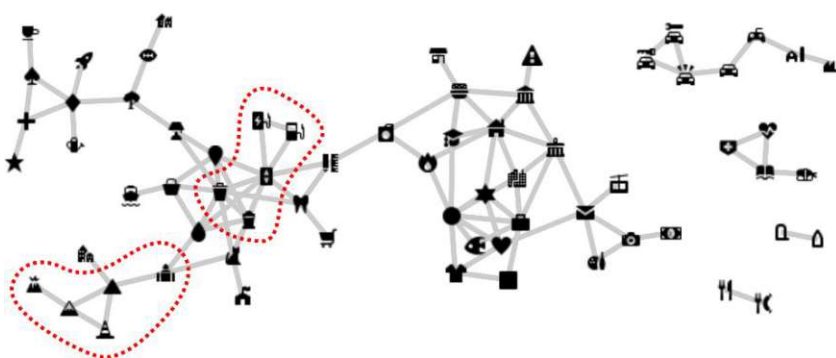
Figure 6.4 Clusters of similar icons for the "Maki" set, for different similarity metrics and thresholds. Pairs of icons above the similarity threshold are connected by a grey line and subjected to an attractive force in the force directed layout. Icons selected as stimuli for the study are marked by the red dotted outline.

An interactive approach involving subjective judgement of the study author was chosen firstly because it was not expected that the simple similarity metrics would perfectly reflect perceptual similarity, but also because the goal was to identify clusters of 4-6 similar icons to use as stimulus sets for the n-AFC tasks. Many clusters identified by the computational analysis containing fewer or more icon pairs, so a manual selection had to be done to yield usable sets of icons. Furthermore, icons should be different in their overall design and not only by a distinct part. For example, the icon pair ¶ (restaurant) / ¶ (restaurant-bbq) consistently shows up as a pair of highly similar icons for all metrics, but the two icons have no other icons with comparably high similarity in the set, and testing the discriminability of the two icons would arguably only test the discriminability of the right half of the icons – the left half is identical for both of them.

Using the interactive notebook, groups of icons that were clustered in close proximity for various settings of similarity metrics and thresholds were identified for inclusion in the study. Multiple candidate sets were identified for each icon set using the interactive process described above. Of these, four sets were selected for inclusion in the study, each containing 5 icons.

Name	Icons	Further icons considered
Maki-triangular *		
Maki-rectangular *		
Maki-solid		
NPS-vertical *		
NPS-rectangular		
NPS-slope		
NPS-figure-ground		
NPS-drinks		
OSM-castles *		
OSM-vertical		

Figure 6.5 Icon sets for potential use as stimuli for the study. The sets have been identified using the interactive procedure informed by computational analysis of shape similarity as described. Sets marked with an asterisk (*) were selected for inclusion in the study.

6.1.3. Pixel grid alignment and grid fitting

The previous section presented an approach for deriving a selection of map icons for the study using an interactive process informed by computational analysis. This section will discuss whether means of presentation can be developed that make optimal use of the high resolutions available in modern displays.

The first study has shown that digital displays can introduce sampling artifacts which are detrimental for the recognition of small stimuli. One strategy discussed there was to add a random offset in the range of ±0.5 pixels to the position of the stimulus, in order to avoid any systematic bias that may otherwise negatively affect the reproduction of stimulus geometry. However, one could also raise the question of whether the stimulus geometry can be aligned with the pixel grid in an ideal way for reproducing fine details, even if at the cost of precise geometric accuracy. This idea has a precedent in the realm of text rendering, where typographers have for a long time faced the question of how the shapes of letters can be fitted to the discrete grids of pixels on computer monitors and laser printers, as has been discussed in Section 3.2.3.2.

Similar principles can be applied to the optimization of icons for display on a discrete pixel grid. Even when created as vector graphics, which should theoretically result in resolution-independent geometry that can be scaled to arbitrary size without loss of detail or "smoothness", icons are often designed with the help of a discrete grid representing the intended pixel grid. For example, the "Maki" icons are designed on a grid of 15 × 15 units (of which only the inner 13 × 13 cells are used for most icons), the OSM icons are designed on a 14 × 14 grid, and the NPS icons are provided in different variants, of which the 22 × 22 grid variant was chosen for this study. Designers are free to use curved and oblique geometry elements, or orthogonal elements not precisely aligned with the grid, but are encouraged to keep the main graphical elements of an icon in alignment with the grid in order to ensure good reproduction at the nominal pixel size (Zhang, 2020a). Introducing an integer pixel multiplier will keep the grid in alignment with the pixel raster, but applying a non-integer multiplier (as is common on modern smartphones, see Section 3.2.2) or arbitrary scaling factor will break the alignment and result in aliasing artifacts.

The effect of scaling down an icon to a pixel size smaller than its nominal grid size, and how manual grid fitting can improve the contrast of fine details, is shown in Figure 6.6. When reproduced at nominal size, contrast is increased in the grid-fitted version, but the details of the design are retained also in the version rendered using antialiasing – because the grid of the icon's geometry aligns precisely with the pixel grid of the display. However, when an icon is scaled down to a smaller pixel size, fine details, such as the cable protruding from the "charging station" icon, appear at reduced contrast in the antialiased version. Manually fitting the scaled-down geometry to the pixel grid will distort the geometry as compared to the original version, but will help to retain graphical detail at maximum contrast (see Figure 6.6, rightmost image).
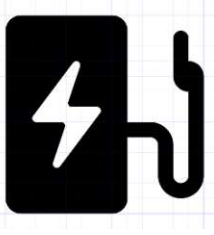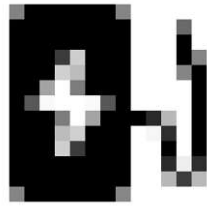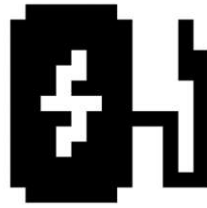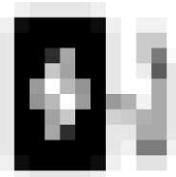
Figure 6.6 The "charging station" icon form the Maki icon set, rendered at nominal size of 15 × 15 pixels and scaled down to 12 × 12 pixels. The version manually fitted to the pixel grid retains detail at maximum contrast.

Producing icons in a range of discrete pixel sizes was common when icons had to be designed as bitmap graphics, but has become less common with the advent of vector graphics and high-resolution screens (J. Krtek, personal communication, May 2022). In the context of this study, the question arises whether the manual creation of bitmap icons optimized for various discrete pixel sizes can indeed improve legibility, particular at high-resolution screens with the potential of reproducing minute details at optimal contrast. Therefore, all icons of the "Maki-triangular" and "Maki-rectangular" subsets have been manually recreated as bitmaps in a range of discrete pixels sizes from 6 × 6 pixels up to 20 × 20 pixels, as shown in Figure 6.7.
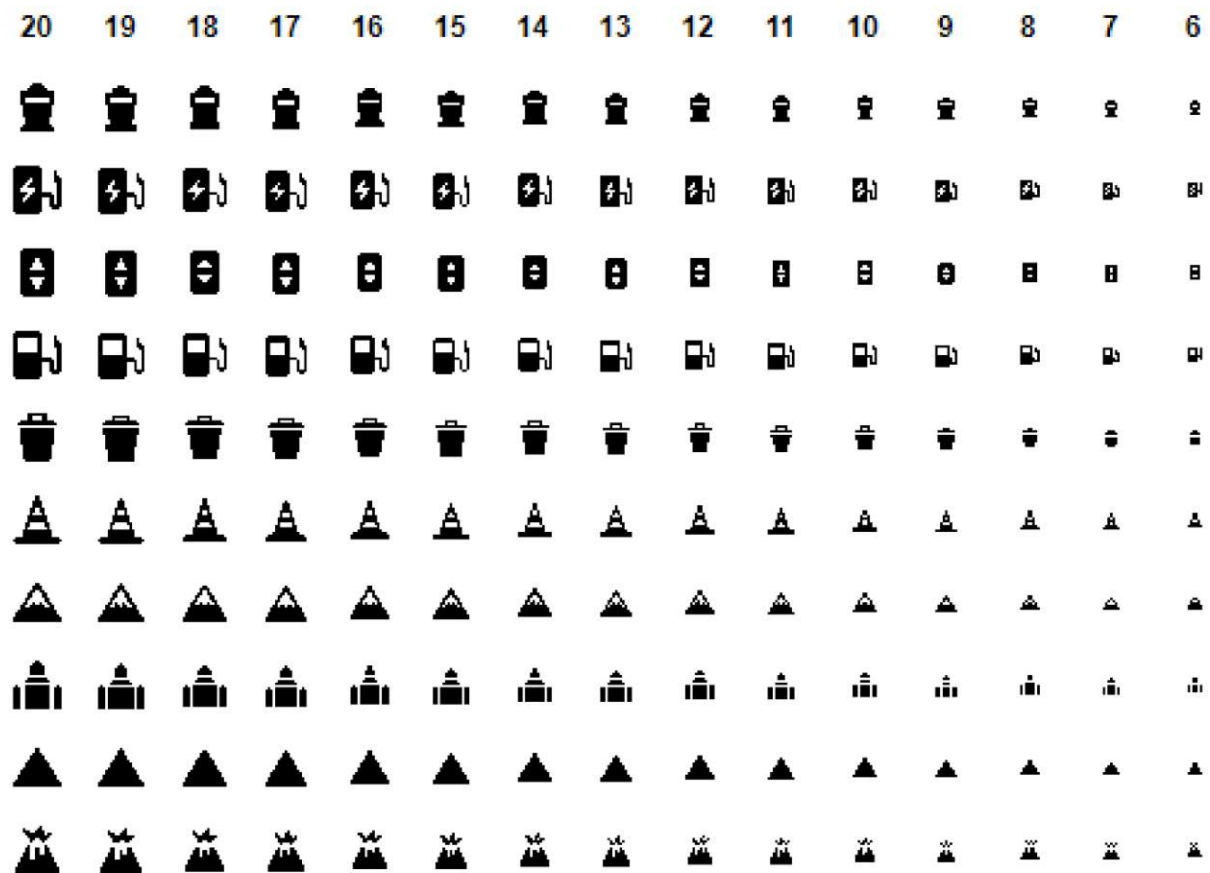
Figure 6.7 Icons from the maki-rectangular and maki-triangular subsets, manually created as bitmaps for a range of discrete pixel sizes from 6 × 6 to 20 × 20 pixels.

The process of pixel fitting leaves some room for subjective judgement when the original geometry is adjusted to fit a mismatching grid of discrete pixels. To ensure a consistent result with minimal deviation from the original, the following guidelines have been formulated to inform the pixel fitting process, which was performed manually in the GIMP raster graphics editor:

- The starting point is a bitmap image rendered from the SVG geometry at the target resolution.
- Features and gaps which have a size of ≥1 grid unit in the SVG geometry should be preserved as good as possible in the bitmap graphics.
- Dimensions of features or gaps can be increased to the next integer pixel size if necessary for preserving the feature, but overall size or area should not increase if possible.
- Geometric relationships and alignment between features should be preserved if possible.
- A feature present at a smaller size should also be present at all larger sizes.
- Reflectional and rotational symmetry of features should be preserved.

These principles can be illustrated in the example of the 12-pixel icon shown in Figure 6.6: The rotational symmetry of the "lightning bolt" feature has been preserved, while not increasing its overall size within the constraints of the discrete pixel grid. The rounded corners of the large rectangle have been removed, and should therefore not be re-introduced at smaller sizes. The width of the cable has been increased to 1 pixel, and the gaps between windings of the cable have been normalized to 1 pixel distance (shortening the distance between the main body of the icon and the first winding of the cable). The "handle" part has been widened to 1 pixel, but shortened to 2 pixels length to keep the vertical alignment below the upper border of the icon's main body.

To make use of any potential advantage of the grid-fitted icons, these must be used only at the precise dimensions, specified in *physical* display pixels, for which they were created. This implies that a direct comparison with arbitrary other sizes, potentially specified in metric units and converted to non-integer pixel sizes, will not generally be valid, and that the effective size of such a grid-fitted icons will depend on the pixel density of the display on which it is presented. However, performance of participants for discriminating the grid-fitted icons can be compared to the two closest sizes of continuously-scaled icons, to see if the manual effort involved in creating the icons pays off by facilitating more accurate discrimination.

6.1.4. Shape difference amplification

The previous section discussed improving the presentation of stimuli by creating bitmap graphics optimized for the pixel grid. In this section, an idea for improving the discriminability of icons within a set will be explored that is based on modifying the vector graphics representation of its shape. Section 6.1.2 discussed the automated analysis and quantification of shape differences for the purpose of selecting the "most similar" icons

within an overall collection. The idea of the method proposed here takes this approach for analysis and attempts to apply it to the *modification* of icons with the goal of increased readability: The geometry of one icon shall be modified in a way that the shape difference to the most similar icons is increased.

This section presents a first implementation of the idea, for the purpose of gathering initial data on whether such an approach delivers results that warrant further investigation in the future. For this prototypical implementation, the "NPS-vertical" subset of icons (see Figure 6.5 above) was chosen because of the icons being differentiated only by minute differences in horizontal extent, which the author felt were particularly prone to confusion and thus a good candidate for exploring improved methods of presentation. The procedure for shape contrast amplification was performed manually, but could be implemented into an automated system if promising result can be achieved, as follows (see Figure 6.9 for an illustration of the results of this method):

1. Rasterize the icons using their nominal grid sizes or a suitable pixel size, and produce a matrix of pairwise difference bitmaps, as described in Section 6.1.2 (see Figure 6.8).
2. For each icon I, consider the n most similar icons S_i, and process *in increasing order of similarity* as follows:
   a. At the centre point of each pixel present in I but not present in S_i (the pixels set in the difference bitmap $I-S_i$), add a *black* circle of radius r, $0.5 < r < 1$ pixels.
   b. At the centre point of each pixel present in S_i but not present in I (the pixels set in the difference bitmap S_i-I), add a *white* circle of radius r, $0.5 < r < 1$ pixels (or subtract a circle of this size from the geometry if a transparent is required).

This results in a vector-based icon design in which the filled parts that differentiate it from the most similar icons appear enlarged (due to the black circles used to "overdraw" those pixels, which extend beyond 1 pixel in size), and the parts in which a similar icon extends beyond that icons boundary appear reduced in size (due to the white circles used to "overdraw" the area in which the similar icon is larger, which extend beyond 1 pixel in size, thus covering parts of the icon). Processing the most similar icons of each icon in order of increasing similarity ensures that the most similar icons are processed later, and their differences more likely to influence the result.

This procedure has been performed manually for the five icons of the "NPS-vertical" set, with values of n=3 (number of similar icons to consider for each icon) and r=0.82 pixels (radius of circles, chosen subjectively to produce a noticeable effect). As a first step, the analysis of raster-based shape difference done at the icon set's nominal resolution of 22 pixels yields the matrix of differences shown in Figure 6.8.

Based on the difference matrix, the three most similar icons have been determined for each icon, and the geometry adjusted as described above to yield the "difference amplified" icons as shown in Figure 6.9.

While the icons modified for shape difference amplification look distorted at large sizes, the idea is that the modified geometry emphasizes those parts of the icon that are important for differentiation from other similar icons.
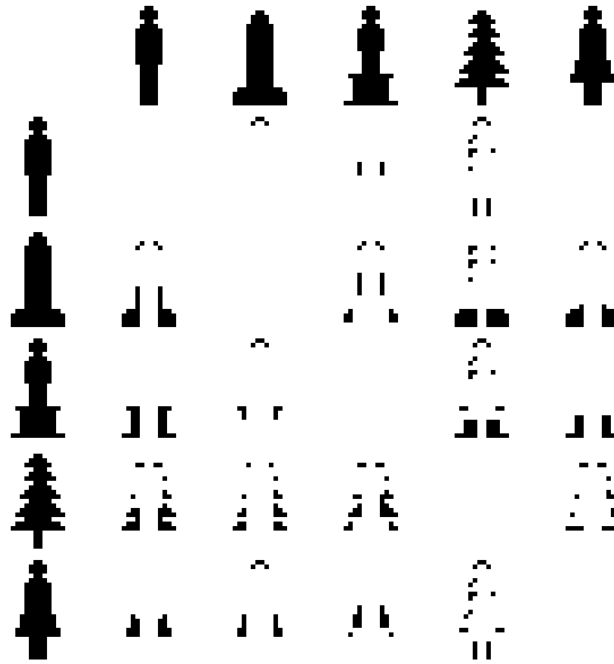
Figure 6.8 Matrix of pairwise raster differences of the "NPS-vertical" icon subset.

Original Shape 25mm / 5mm / 1.5mm	3 most similar shapes (XOR pixel difference)	"Difference Amplified" 25mm / 5mm / 1.5mm
	(22) (40) (48)	
	(36) (48) (54)	
	(36) (40) (42)	
	(45) (57) (73)	
	(22) (42) (45)	

Figure 6.9 Analysis and results for attempted "difference amplification" for the icons of the "NPS-vertical" subset.

6.2. Study design

The considerations made in Section 6.1 on stimulus selection and presentation will be used to inform the design of the second study, which will investigate point symbol legibility empirically.

6.2.1. Selection of stimulus sizes

The basic framework of n-alternative-forced-choice (*n*-AFC) tasks will be maintained, as this task design delivers data that is well suited for statistical analysis. However, the staircase method of stimulus adaption was seen as problematic after analysis of the first study for several reasons: (1) it results in irregular sampling of trial conditions across the group of participants, because stimuli will be presented at sizes near each participant's threshold, reducing the validity of statistical analysis of overall trial results; (2) the relevance of a "threshold" value at which a 50% or even 90% success rate is accomplished by a participant is of dubious relevance for cartography – certainly, cartographers would like map symbols to be correctly read at an accuracy higher than 90%. On the other hand, requiring a 100% success rate will give lapses, erroneous operation of the response device, or even user interface malfunctions a disproportionate influence on results; (3) when stimulus size is adjusted by a constant factor for each step (e.g. 1.2 or $\sqrt{1.2}$ in the previous study), threshold values will be floating-point numbers with an arbitrary number of decimal digits (e.g. 1.273493 mm), which is hardly useful as a simple guideline for practitioners – but the precise effect of rounding the number up or down for efficient communication (e.g. to 1.3 mm) is not known.

For the reasons given in the previous paragraph, stimuli in the second study will be presented using the method of constant stimuli (Cunningham & Wallraven, 2011), in which stimuli are simply presented at predefined sizes for a fixed number of repetitions to each participant. In the context of psychophysical research, other methods such as the staircase method or QUEST (Watson & Pelli, 1983) are usually considered superior for the purpose of rapidly converging towards a participant's threshold level with fewer trials, but do not deliver a robust estimate of error rates at higher levels. However, as has been argued above, for cartographic applications precise performance at larger stimulus sizes and corresponding high success rates is of primary interest. Also, stimuli shall be presented at numeric values that are rounded to a certain number of significant digits, and not determined by mathematical operations.

The method of constant stimuli presents stimuli at a pre-defined set of sizes and number of repetitions for each size to all participants. The fixed set of sizes at which symbols are presented has been chosen to mimic roughly the factor of 1.2 used in the first study, but sizes are rounded to at most two significant decimal digits. The values were chosen to include size levels at which the symbols can be identified with high reliability, down to sizes at which correct classification becomes unreliable, by an initial choice by the author's subjective impression which was subsequently refined in a pilot study involving three participants.

The sampling strategy described above is hoped to deliver a reliable sampling of overall success rates across participants, at distinct size levels which can be directly translated to recommended minimum dimensions. For example, a result could be that for a particular combination of stimulus class and display, the stimuli could be identified by participants with a 99% success rate at sizes of 1.5 mm and 1.25 mm, with 95% at 1.0 mm, 92% at 0.85 mm and 83% at 0.6 mm (these numbers are taken from the results of Task 1 on display D_5 – see Figure 6.14 on page 144 for the full results). In contrast to the reporting of just a single threshold value for each participant, such reporting can give a more representative insight into the impact of stimulus variation on success rate, and also allows for statistically sound analysis of trial results for specific levels or sub-groups of participants.

6.2.2. Devices and experiment configuration

The decision was made to keep a within-subject design in which each participant would perform all tasks on all displays. As described in the previous section, the method of constant stimuli was perceived to provide advantages over the adaptive staircase method used in the first study, at the cost of a potentially increase number of repetitions per participants per task. In order to keep overall experiment duration per participant under one hour on average, while investigating the various aspects discussed in Section 6.1, the decision was made to reduce the number of display stations to three. The first study had shown a clear disadvantage of the lowest-resolution device for five out of six tasks, but the display used (LG P-970) has a lower pixel density than any phone currently available for consumers (as of 2022). It was therefore replaced by the phone model LG K50S with a measured pixel density of 265 ppi, which reflects more accurately the lower bound of pixel densities commercially available today. This phone supports version 9 of the Android operating system as well as current versions of Google Chrome for Android, which removed the need to provide server-side rendering for the software configuration. A new adapter block was produced to mount the LG K50S in the way described in Section 4.2.2. The two phones of highest pixel density (D_3 and D_4 in the first study) were kept in use for the second study and will be referred to by the same designations for consistent identification throughout this thesis, while the newly introduced phone LG K50S will be referred to as D_5 – it needs to be noted, however, that D_5 is now the device with the lowest pixel density in the study (265 ppi), followed by D_3 (522 ppi) and D_4 (801 ppi). Thus, the devices will subsequently be listed in the order of their display resolutions: D_5, D_3, D_4. As in the first study, the Xiaomi Mi Mix 2 has been used as response device, with a pixel density of 403 ppi at a display size of 68 × 136 mm. Table 6.1 gives an overview of the devices used for stimulus display in this study.

ID	Model	Year	Display Technology	Subpixels	Display Dimensions	Pixel Density[48]
D_5	LG K50S	2019	LCD	RGB	69 × 149 mm	265 ppi
D_3	Samsung Galaxy Note 4	2014	OLED	PenTile	71 × 125 mm	522 ppi
D_4	Sony Xperia Z5 Premium	2015	LCD	RGB (ZigZag)	68 × 121 mm	801 ppi

Table 6.1 Devices used for stimulus display in study 2.

The screen size of D_5 is considerably larger than that of D_1 used in the previous study, therefore the bezels could be cut to a larger opening – 66 × 72 mm – to potentially accommodate larger stimuli. The bezels were produced from solid grey cardboard (300g/m²) with an outer dimension of 350 × 250 mm. The shelves described in Section 4.2.2 were used to hold the phones, with only stations A, B and C occupied by displays in the order D_3, D_4, D_5. The order in which each participant was assigned to individual stations was randomized. The shelves were positioned to result in a horizontal distance of 325 ± 1 mm between chin rail and display surfaces, resulting in a viewing distance of 34 - 35 cm to the centre point of each screen, at an angle of 16-20° down from horizontal.

Display brightness was calibrated for all displays to account for an oblique viewing angle as described in Section 4.3.3, which resulted in target values of D_5: 390 cd/m², D_3: 323 cd/m², D_4: 390 cd/m² as measured with the display calibration sensor. At the start of each lab session, the displays were verified to lie within ± 2% of these target values, and adjusted if needed by using the brightness adjustment method provided by the Android operating system.

All other aspects of the hardware and software configuration for the experiment remained unchanged from the general design described in Chapter 4 and the configuration of study 1 described in Chapter 5.

6.2.3. Stimulus and task design

The basic forced-choice design of the first study was maintained for this study, and stimuli were designed in accordance with the ideas described in Section 6.1. The first four tasks at each station demanded discrimination of the icons from some of the identified test sets, presented in isolation at maximum contrast. Subsequently, icons subjected to manual grid fitting and shape difference amplification as described in Section 6.1 were presented in tasks five to eight. The final two tasks presented participants with a simple map, on which icons from one of the subsets were places, with the task of counting one specific icon type.

6.2.3.1. Task 1: Discriminating icons from the set "Maki-rectangular"

For each trial of this task, a single icon from the "Maki-rectangular" set was presented in the centre of the display at maximum contrast, black on white background. The response device showed buttons for each of the five icons from the set, with each button showing the graphical icon at a size of 6 mm and a textual label. The response buttons were stacked vertically, in

---

[48] Pixel density values have been derived from measurements of the actual devices, and may deviate slightly from information found elsewhere.

alphabetical order of the textual labels (see Figure 6.10a for a screenshot of the response device showing the five choices). Before Task 1 started, a message displayed on the stimulus device reminded participants of the task, reading "Next Task: Press the button on the response device that best matches the shown graphics. Press «Continue» when you are ready."

The icons were randomly selected by shuffling an array containing two entries for each icon and picking the next entry from that array for each trial, thus ensuring an even number of presentations for each icon and preventing repetition of the same icon more than two times in a row. Stimuli were presented at the following sizes in order: 1.5, 1.25, 1.0, 0.85, 0.7, 0.6, 0.5 and 0.4 mm. For each stimulus size, four trials were run before advancing to the next smaller stimulus size.

To prevent aliasing effects influencing the legibility of icons in a consistent way, which has been seen in the first study to cause bias in the results, the icons were offset by a random amount in the range of -0.5 – +0.5 physical pixels.

(a)	(b)	(c)	(d)
Tasks 1, 5, 6	**Tasks 2, 7**	**Tasks 3, 8**	**Task 4**
Cemetery	Construction Site	Men's Restroom	Castle
Charging Station	Mountain	Monument	City Gate
Elevator	Temple	Statue	Fort
Petrol Station	Triangle	Large Tree	Fortress
Waste Basket	Volcano	Women's Restroom	Palace

Figure 6.10 Screenshots of the response device for tasks 1–8, scaled down to 75% of actual size.

6.2.3.2. Task 2: Discriminating icons from the set "Maki-triangular"

Task 2 was set up in an identical way as Task 1, with icons from the set "Maki-triangular" as stimulus choices. See Figure 6.10b for a screenshot of the response device showing the five choices for this task.

6.2.3.3. Task 3: Discriminating icons from the set "NPS-vertical"

Task 3 was set up in a similar way as Task 1, with icons from the set "NPS-vertical" as stimulus choices. See Figure 6.10c for a screenshot of the response device showing the five choices for this task. The number of repetitions for each size was reduced to 3 for this task.

After 2 participants, the smallest size of 0.4mm was excluded as a condition for this task, as it became apparent that success rates were below 50% even for the next larger size.

6.2.3.4. Task 4: Discriminating icons from the set "OSM-castles"

Task 4 was set up in a similar way as Task 1, with icons from the set "OSM- castles" as stimulus choices. See Figure 6.10d for a screenshot of the response device showing the five choices for this task. The number of repetitions for each size was reduced to 2 for this task.

After 2 participants, the smallest size of 0.4mm was excluded as a condition for this task, as it became apparent that success rates were below 50% even at the next larger size.

6.2.3.5. Task 5: Discriminating icons from the set "Maki-rectangular", rendered at maximum binary pixel contrast by using a threshold value

It has been explained in Section 6.1.3 that rendering icons at arbitrary sizes will result in non-optimal contrast when geometry is not precisely aligned with the pixel grid, because antialiased rendering results in pixels filled with intermediate grey values when a pixel is not fully covered by a filled shape. If maximum contrast is desired, rendering without antialiasing would achieve maximum black/white contrast, at the cost of perceived smoothness of the resulting image. Unfortunately, on modern systems antialiasing is activated by default and can in many cases not be deactivated. One approach to neutralize the effects of antialiasing and achieve full contrast on such systems is to render the shape with antialiasing, and then perform a pixel-by-pixel threshold operation, colouring any pixels which are below a given threshold fully black (e.g. pixels below a value of 128 for the average of its red, blue and green channels, representing the midpoint between black and white), and all pixels above the threshold fully white. It has to be noted that this approach will remove any detail of geometry that results in less than half of a pixel covered, and therefore will cause significant distortions of small shapes. However, particularly for the highest resolution phones, the benefits of achieving increased contrast with such a thresholding approach was assumed to potentially outweigh the detrimental effects of shape distortions, which may be below the threshold for being noticeable on these displays due to the small pixel size.

For Task 5, icons from the "Maki-rectangular" set were rendered using the default canvas rendering algorithm, and subsequently subjected to a thresholding pass with a threshold value of 128 (all physical pixels with an average RGB value larger than the threshold were coloured white, all others black).

Before rendering, icons were scaled to integer pixel sizes to correspond with the manually grid-fitted icons used in Task 6. For each station, a range of integer pixel sizes in the range from 6 to 20 pixels was identified to result in stimulus sizes within the range for which

successful discrimination can be reasonably expected. For this purpose, Table 6.2 was constructed, showing the metric size of integer pixel sizes for each display. From this table, suitable pixel values to result in stimuli in the range of 0.45 to 1.0 mm were selected for each display, resulting in the following choices of pixel dimensions:

D_5: 10, 9, 8, 7, 6 pixels
D_3: 20, 18, 16, 14, 13, 12, 10, 9 pixels
D_4: 20, 18, 15, 14 pixels

For Task 5, pixel sizes progressed in decreasing order at each station, with four trials being presented for each size. Icons were selected from a shuffled array in an identical manner as described for Task 1. On the response device, the icons were presented at a size of 6mm and rendered in the default way, at maximum contrast using anti-aliasing (see Figure 6.10a).

The purpose of this task was to verify whether the manually grid-fitted icons, which are used as stimuli in Task 6, result in better performance than the simple thresholding method performed on the stimuli of this task, and whether such thresholding for maximum contrast has the potential to increase performance on high-resolution displays as compared to the default rendering done for Task 1.

D_5		D_3		D_4	
265 ppi		522 ppi		801 ppi	
px	**mm**	**px**	**mm**	**px**	**mm**
1	0,10	**1**	0,05	**1**	0,03
12	1,18				
11	1,08				
10	0,96	**20**	0,97		
9	0,86	**18**	0,88		
8	0,77	**16**	0,78		
7	0,67	**14**	0,68		
		13	0,63	**20**	0,63
6	0,58	**12**	0,58	**18**	0,57
		11	0,54	17	0,54
		10	0,49	**15**	0,48
		9	0,44	**14**	0,44
		8	0,39	12	0,38

Table 6.2 Conversion of integer pixel sizes to metric dimensions for each display, with roughly corresponding sizes positioned on the same line of the table. It becomes apparent that for icon sizes in a range of 6 to 20 pixels, the metric sizes for D_5 and D_4 barely overlap. On the highest-resolution display D_4, even the largest size of 20 pixels results in an icon of less than 0.7 mm in size. Pixel sizes selected for tasks 5 and 6 for each display are shown in bold.

6.2.3.6. Task 6: Discriminating icons from the set "Maki-rectangular", manually grid-fitted

The manually grid-fitted icons for the "Maki-rectangular" set, as described in Section 6.1.3 and shown in Figure 6.7, were used as stimuli for Task 6. The pixel sizes to be presented on

each display were determined in an identical way to what has been described above for Task 5, based on the dimensions shown in Table 6.2. The following pixel dimensions were selected for each display:

D_5: 10, 9, 8, 7, 6 pixels
D_3: 20, 18, 16, 14, 13, 12, 10, 9 pixels
D_4: 20, 18, 15, 14 pixels

Other parameters of the task were identical to Task 5.

6.2.3.7. Task 7: Discriminating icons from the set "Maki- triangular", manually grid-fitted

The manually grid-fitted icons for the "Maki-triangular" set, as described in Section 6.1.3 and shown in Figure 6.7, were used as stimuli for Task 7. The sequence of pixel sizes for each station and other parameters of the task were identical to Task 6.

6.2.3.8. Task 8: Discriminating icons from the set "NPS-vertical", rendered using shape difference amplification

The icons from the "NPS-vertical" set with manual shape difference amplification applied as shown in Figure 6.9 on page 127 were presented as stimuli for Task 8. The parameters of presentation and sequence of sizes was identical to Task 3 (regular rendering of the icons from the same set), with four repetitions at each size level. On the response device, the icons on the response buttons were presented at a size of 6mm and rendered in the default way, without shape difference amplification (see Figure 6.10c).

6.2.3.9. Task 9: Counting icons from the set "Maki-triangular" on a map

In the tasks described so far, the stimuli consisted of single icons presented in the centre of the screen at maximum contrast. To gain insight into more realistic use cases and a slightly more complex map use task, tasks 9 and 10 asked the participant to count icons on a map.

Base map design

The base maps prepared for this purpose were created from GIS data of rural areas in Austria. Four regions were selected which contained a somewhat uniform density of roads, some rivers and road crossings, but no urban agglomerations or large waterbodies. The maps were created using data extracted from OpenStreetMap, and rendered using the Web Mercator projection (EPSG:3857).

The symbology of the base maps was designed to be similar to the styling of the Google Maps road maps layer, without textual labels or icons other than the target icons. The map background was a light grey (RBG code #eef0f2), and roads were depicted by white lines with a 80% grey outline. Road width was set to 1.3 mm, 1.0 mm and 0.9 mm for primary, secondary and tertiary roads, respectively, with the outline adding another 0.15 mm to either side (0.2 mm for primary roads). Waterways were rendered as 0.26 mm wide lines in light blue (#9ad1f3), and waterbody polygons were filled in the same colour. Figure 6.11 shows the four base maps designed for use in this task.

In each of the four regions identified as suitable as a base map, 12 placeholder locations for point symbols were defined in the vicinity of roads, covering a roughly square area on the projected map. These placeholders were positioned according to the following rules:

- One cluster of three icons was created at a street intersection or junction, with each icon being placed at the opposite side of the road of a previously placed icon.
- Another cluster of two icons was created at a different location, at opposite sides of the road.
- The remaining 7 icons were distributed somewhat evenly over an approximately square region, at plausible locations next to roads.

The layout of each map added a 20% margin to either side of the convex hull of placeholder points. The resulting maps were output to SVG files with a nominal size of 100 × 100 mm.
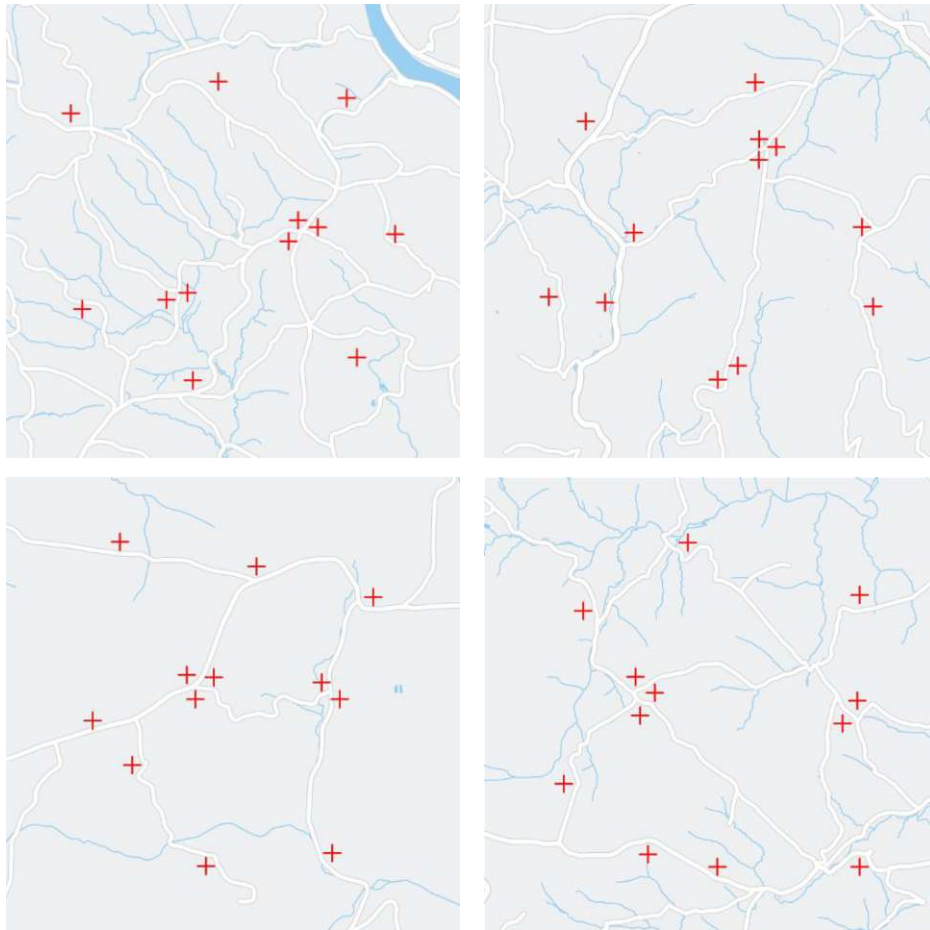
Figure 6.11 The four base maps used for tasks 9 and 10, reproduced at the size used in the experiment (60 × 60 mm). Each map contains 12 placeholders (red crosses) for placing icons, which were replaced by the icons from the target set.

Task design

For the "count on map" tasks 8 and 9, the base maps created in the way described above were rendered at a size of 60×60 mm against a 10% grey background (RGB code #e5e5e5) on the

stimulus display. One of the four base maps was chosen at random for each trial. The 12 placeholder elements on the map were replaced with icons from the target set (for Task 9: "Maki-triangular"), scaled to the target size level. To ensure the presence of the target icon (the icon that the user was prompted to count) as well as a sufficient number of the icons most similar to it, the following approach was implemented to determine the set of icons shown on the map for each trial:

1. The target icon for the trial was chosen randomly among the five icons of the set (e.g. "mountain"). For each icon of the set, a list of the other four icons, sorted by raster-based XOR similarity (see Section 6.1.2), was also stored in the experiment definition file (e.g. for the "mountain" icon: "triangle", "construction", "volcano", "place-of-worship").
2. A random integer number from 1 to 7 was chosen as the number of map locations assigned to the target icon.
3. The next most similar icon was assigned to half of the remaining map locations (rounded up), chosen at random. This step was performed repeatedly, until the last icon in the list was reached, to which the remaining map locations were assigned.

This results in the assignment of map locations to the target icon and the icons most similar to it according to one of the following frequency distributions for each trial: **1**-6-3-1-1, **2**-5-3-1-1, **3**-5-2-1-1, **4**-4-2-1-1, **5**-4-2-1-0, **6**-3-2-1-0, **7**-3-1-1-0. After determining the frequency of each icon in this way, icons were assigned in random order to the placeholder locations on the map.

During the task, the stimulus display showed the map with icons places at the placeholder locatons as described, and a label on top of the map, reading: "Count:", with the icon in graphical form and the pluralized name of the icon next to it (see Figure 6.12, left). The response device showed the text "Count the number of:", with the icon in graphical form and the pluralized name of the icon next to it, a "map legend" with all five icons and their names stacked vertically in alphabetical order, and a series of buttons with numbers ranging from 0 to 12 (see Figure 6.12, right). Participants were not made aware of the fact that the target icon was always present between 1 and 7 times, and could choose among all theoretically possible answers.
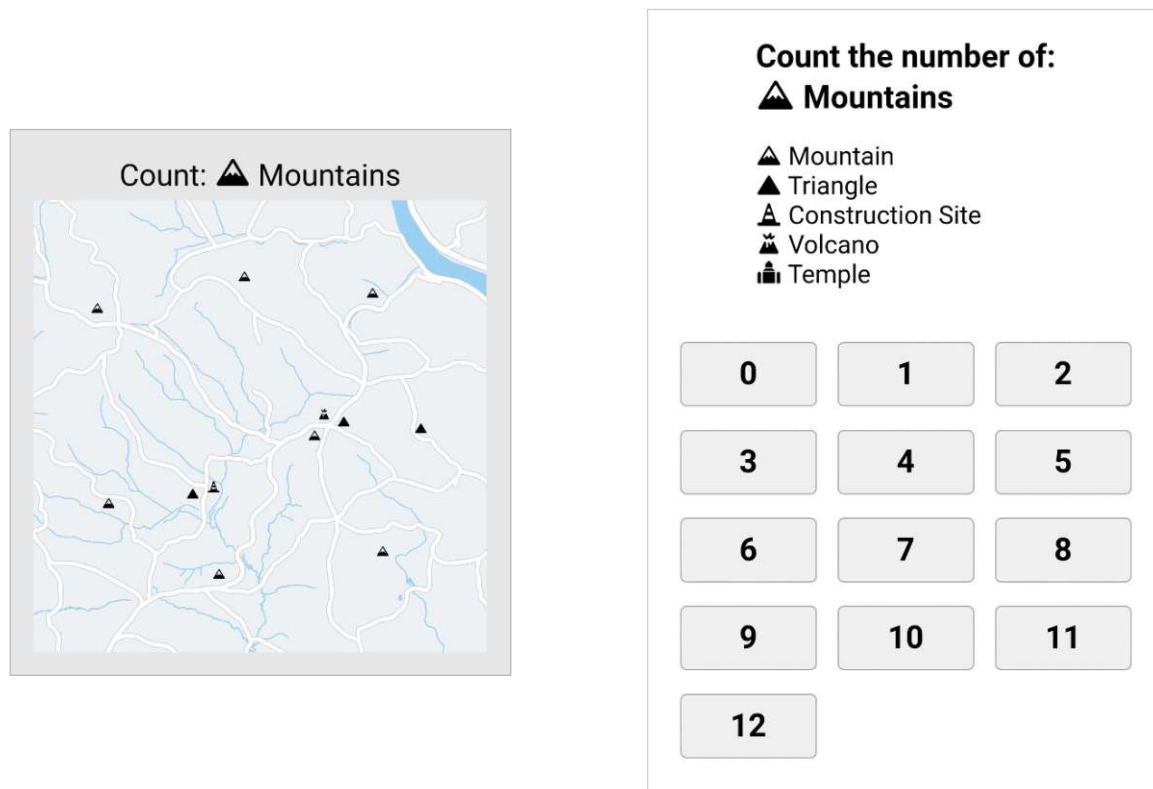
Figure 6.12 Example stimulus on display D_3 (left) and corresponding response interface (right) for Task 9, reproduced at original size. Icons on the map are shown at the initial size of 1.5 mm (referring to, in this case, the base of the triangle).

Before Task 9 was started, the text: "Next Task: Count the indicated icons on the map accurately, but also as fast as possible. Press «Continue» when you are ready." was displayed on the stimulus display.

Four trials were run (two with and two without the base map) at each of the following sizes: 1.5 mm, 1.25 mm, 1.0 mm, 0.85 mm, 0.7 mm, 0.6 mm. The size levels of 1.5 mm and 1.0 mm were introduced after the first four participants, when it became apparent that a drop in performance was already noticeable at the higher size levels.

6.2.3.10. Task 10: Counting icons from the set "Maki- rectangular" on a map

Task 10 was set up identical to Task 9, using the same base maps. For this task, the icon set "Maki-rectangular" was used as target icons.

6.2.4. Participants

21 participants for the second study were recruited in March 2022 among students of the course "Web Mapping" (summer term 2022). Bonus points amounting to 5% of the courses overall points were awarded as a compensation for the time spent travelling to and from the experiment. Another 9 participants were recruited among visitors to the cartography research group between June and September 2022. Before participation, it was made clear to all participants that participation in the experiment was entirely voluntary (any bonus points for the course were awarded as soon as a participant arrived at the experiment location), and

an alternative task to get the bonus points was offered to all students of the course. As mentioned in Section 4.3.4, participants were reminded again verbally as part of the experiment protocol that their participation in the experiment was entirely voluntary and that they could abort the experiment at any time and should do so if the felt unwell.

The session of one participant had to be abandoned due to a power outage in the building in which the lab was located; 29 participants completed the experiment. Of these, 15 participants identified as male and 14 as female on the questionnaire. 18 participants were aged 16-25 years, 8 were in the 26-35 years age group, 2 in the 36-45 age group and one participant was in the 56-65 years age group.

Participants were asked to arrive wearing any vision correction (glasses or contact lenses) as they would normally use for reading a book or looking at their phone for an extended period. The effective habitual visual acuity was assessed during the experiment using the "Tumbling E's" task on display D_4. Due to the diminished performance observed for this task on display D_3 in the first study, for this second study the task was only run on the highest-resolution display.

Figure 6.13 shows the visual acuity of participants as assessed by the Tumbling E's task. The results show that a single participant reached a measured logMAR score below -0.35, and another 7 participants reached a score below -0.30. Since the proportion of the population with a logMAR score of -0.30 or lower ("20/10 vision") is estimated to not exceed 1% (Durrie et al., 2019), it was concluded that the chosen method of assessment and calculation of the logMAR score most probably overestimates visual acuity (underestimates the logMAR value). This is in line with an assessment by Pointer (2008), who concluded that various methods of assessing visual acuity may diverge by up to 0.10 logMAR units. The values assessed should therefore not be taken for absolute values, or directly compared with other assessments of visual acuity. However, for the purpose of assessing relative acuity of participants the results should be reliable.

Visual acuity of participants, study 2
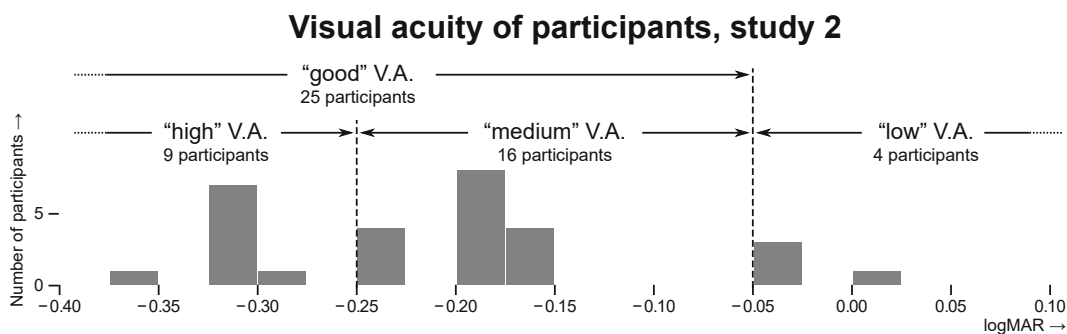
Figure 6.13 Distribution of visual acuity of participants and visual acuity groups used for the analysis for the second study. Visual acuity has been assessed with a "Tumbling E" task on display D_4.

Based on the histogram of logMAR scores, three classes of visual acuity have been defined: "high" visual acuity for logMAR < -0.25, "medium" visual acuity for logMAR ≥ -0.25 and logMAR < -0.05, and "low" visual acuity for logMAR ≥ -0.05. Furthermore, the class of "good" visual acuity as the union of "high" and "medium" acuity classes (logMAR < -0.05)

has been defined and will be used as the main target group for analysis of the results of the study. It is important to note that these qualifiers refer only to the relative performance of participants in relation to the other participants; in a medical context, a participant with a logMAR score of 0, putting them in the class of "low" habitual acuity and excluding them from the "good" acuity class for the purpose of this study, would most probably be assessed as still having a healthy, normal acuity, and may only need suitable vision correction to reach an even better score. Given the fact that the method of assessment probably underestimates logMAR scores as discussed above, and that therefore the chosen threshold of logMAR -0.05 roughly corresponds with the well-established criterion of "20/20 vision" (logMAR 0), and considering the overall distribution of participants' assessed logMAR scores, the choice of class boundaries was found to plausibly reflect demographic segments that may be realistically encountered in the real world.

6.2.5. Changes during the experiment

The size levels to be presented for each task were established and verified in informal pilot tests conducted among the author and colleagues at the cartography research group. In the early course of running the study with actual participants, a few cases were noticed in which the configured smallest sizes for a task were yielding very low recognition rates, and in two cases the initial levels proved too small for reliably establishing performance at large-enough sizes. Thus, the following changes were implemented during the course of the experiment:

- After 2 participants, the smallest size levels for tasks 3, 4 and 8 were removed, as they were resulting in very low success rates. The number of trials at larger levels for Task 3 was increased from 2 to 3.
- After 4 participants, the smallest size levels for tasks 9 and 10 were removed, as they were resulting in very low success rates. Additional size levels at 1.5 mm and 1.0 mm were introduced for these tasks, because a drop in performance was noticeable between the previous two largest levels of 0.85 mm and 1.25 mm.
- After 6 participants, the smallest size levels for tasks 5, 6 and 7 on displays D_3 and D_4 were removed, as they were resulting in very low success rates.

The removed levels will not be taken into account for the analysis of results, and should have had no effect besides a slight reduction of overall experiment duration. The levels which have been introduced after 4 participants will have a smaller number of trials in total, which will unfortunately reduce the significance of the statistical analysis. This will be taken into account in the results section.

6.3. Hypotheses

6.3.1. Hypotheses on the relationship between size and performance

One way to formalize the idea of a minimum dimension for a map symbol or a set of symbols is as an implicit hypothesis: that the legibility of the symbol is constant (at some to-be-

determined level) when the symbol is drawn at the minimum dimension or larger, and deteriorates when reduced to a smaller size. While other definitions are certainly possible, for example taking into account user comfort or complexity of the task, this model will be operationalized for the analysis for this study. The framework of constant stimuli, in which the same number of trials for each combination of parameters (in this case, size and display resolution) is presented to each participant and responses are classified by a binary scheme (correct/incorrect) is well suited for such analysis, as it will produce for each parameter combination the proportion of correct responses given by the participants. Detecting the minimum dimension as defined above is thus equivalent to testing, at each decrease in size, the hypothesis that the rate of correct responses at the smaller size is significantly lower than the rate of correct responses at larger sizes.

For identifying the minimum dimension, this implicit hypothesis will be tested for each decrease in size, starting with a comparison of the two largest sizes at which the stimuli were presented. In order to be confident that indeed the correct point at which a drop in performance can first be attributed to a reduction in size, the first two size levels should be chosen in a way that there is no significant difference in performance between them. Otherwise, it may well be that the largest level is already at a size level at which performance is diminished, but there is no way to find out whether that is the case and, if so, at which size level the true threshold lies, by analysing the data. This also motivated the changes described above, introduced when the study was already under way – for tasks 9 and 10, it was becoming apparent that the largest levels had not been chosen large enough to confidently establish the "base line" performance for the tasks. To establish the minimum dimension in the way proposed here, symbols must be presented at least at two larger size levels to establish the base line performance for the task.

The first significant drop in performance detected upon a reduction of size will be the most important one, as it will, by above definition, establish the minimum size under the given circumstances. With each further reduction in size, we can expect the performance to drop further. To gain additional insight into how reducing the size of a symbol below the minimum size affects the rate of correct identification, it is proposed to keep analysing the performance difference with every subsequent size reduction until the smallest size is reached, using the same framework: whether a significant drop can be detected as compared to the success rate since the last detected drop. The distribution of performance drops across the range of sizes, even beyond the minimum size as established by the first drop, will hopefully contribute additional insight.

6.3.2. Hypotheses derived from the one-arcminute and one-pixel models

As has been laid out in Chapter 3, there are two simple models of graphical resolution that have been widely used throughout the history of visual media: the model of one minute of arc as the smallest detail that can be reliably recognized by humans with normal visual acuity, and the model of one (physical) pixel as the smallest detail that can be reliably reproduced on digital screens. If these models are valid, it could be expected that, for graphics presented in isolation and in non-detrimental circumstances, a minimum size derived from such

theoretical assumptions should correspond with an experimentally established first drop in performance, as outlined above.

Table 6.3 shows a set of hypothetical minimum dimensions for the four icon sets used in this study, derived from these simple models of visual acuity and graphical detail. The values in the table are based on a viewing distance of 34cm and a smallest detail of one unit of the grid on which the icons were designed (e.g. $1/22^{nd}$ of the icon size for the NPS icons – see Section 6.1.3 for discussion). Each of the values in the table can be read to represent a hypothesis that the first drop in performance would occur close to that size level. At the given viewing distance, 1 minute of arc corresponds with a size of ~0.1 mm, and 1 CSS pixel corresponds with a size of ~0.125 mm.

Icon set	Model: 1 minute of arc	Model: 1 CSS pixel	Model: 1 pixel on D_5	Model: 1 pixel on D_3	Model: 1 pixel on D_4
Maki-rectangular	1.3 mm	1.65 mm	1.25 mm	0.65 mm	0.4 mm
Maki-triangular	1.3 mm	1.65 mm	1.25 mm	0.65 mm	0.4 mm
NPS-vertical	2.2 mm	2.75 mm	2.1 mm	1.1 mm	0.7 mm
OSM-castles	1.4 mm	1.75 mm	1.35 mm	0.7 mm	0.45 mm

Table 6.3 Hypothetical minimum dimensions derived from simple models of visual acuity and pixel size.

6.3.3. Further task- and display-specific hypotheses

Besides the hypotheses implicit in testing for the minimum dimension and comparing the results with simple models of graphical detail as outlined above, some further hypotheses relating to the performance for individual tasks or displays have been formulated before the start of the experiment:

- <u>HS2.1</u>: Because of the significant advantage found of displays D_3 and D_4 over D_2 for the icon-related tasks 4 and 5 of the first study, an advantage of D_3 and D_4 over the lower-resolution display D_5 is expected for the simple icon discrimination tasks 1–4. This would show as a significant difference in performance for participants with good visual acuity between those displays for some identical size levels.
- <u>HS2.2</u>: A significant advantage of the highest-resolution display D_4 over D_3 is not expected, because its resolution is generally considered to be higher than the resolution of human visual perception, and a significant advantage was not observed for the relevant tasks in the first study. Due to the random offset of ±0.5 pixels introduced for positioning icons on the display for tasks 1–4, it is not expected that any strong detrimental influence of aliasing, such as has been observed for Task 1 in the first study, will be observed.
- <u>HS 2.3</u>: An advantage of a lower-resolution display over one of higher resolution is not expected for tasks 1–4, 9 and 10. Despite the results of Task 1 of the first study, which contradicted a similar hypothesis, the specific circumstances of a detrimental alignment of the stimulus geometry with the pixel grid is not expected for the stimuli of this study, and should be mitigated by the random offset of ±0.5 pixels introduced for positioning icons on the display.

- **HS2.4**: For the icons subjected to automated thresholding (Task 5) a generally worse performance is expected than for the default rendering (Task 1) on the lower-resolution displays D_5 and D_3 due to the distortions of shape introduced by the thresholding process. For the manually grid-fitted icons (tasks 6 and 7), better performance than the default rendering is expected on these displays, due to the increased contrast and exact alignment with the pixel grid. For the display of highest resolution D_4, the interventions of tasks 5, 6 and 7 are expected to result in no significant difference in performance due to its high resolution.

- **HS2.5**: The intervention of shape difference amplification (Task 8) is expected to negatively impact performance at large sizes due to the noticeable distortion of the original shapes. At smaller sizes at or below the first significant drop in performance, better performance than the baseline task (Task 3) is expected at least for some displays and size levels.

- **HS2.6**: Display resolution is not expected to have much impact for the map counting tasks (tasks 9 and 10), since the resolution of peripheral vision is considered far below the resolution of any of the displays, and identification of the individual icons will not be affected by display resolution at the larger sizes expected to be required to reliably complete the task.

- **HS2.7**: For the map counting tasks (tasks 9 and 10), a significant drop in performance is expected at larger sizes than for the corresponding task of identifying individual icons (tasks 2 and 1, respectively).

6.4. Results

The method of constant stimuli adopted for this second study, resulting in stimuli of all levels being presented at equal frequency to all participants, allows for a valid analysis of performance at specific levels across participants or subgroups of participants (because all stimulus sizes are presented to all participants, even if well above or below the individual threshold for the participant). For each trial, the result is a binary categorical one: the response is either correct or incorrect. For any set of trials, e.g. all trials at a specific size level on a specific display, the result is the frequency (absolute counts) or percentage rate (relative proportion) of correct responses in proportion to the overall number of trials.

Field (A. Field, 2017, p. 839) recommends using Fisher's exact test for testing whether categorical outcomes are likely to have originated from the same distribution of probabilities, in particular when expected frequencies of any result category may be smaller than 5 (which is the case for the data collected for this study, since trials at large sizes are expected to yield only a very low number of incorrect responses). The test calculates the exact probability p at which two sets of results (e.g. trials at two different size levels at a particular display) originate from the same probability distribution. The null hypothesis (no significant difference) is rejected when p is smaller than a certain threshold, commonly 0.05 (less than 5% probability of originating from the same distribution) or 0.01 (1% probability).

As explained in Section 6.3.1 above, the method proposed to identify minimum dimensions is to test, for a sequence of size reductions, at which point the first significant drop in the rate of correct responses occurs. To operationalize this procedure and aid in discussing the results, the following terminology and procedural details are proposed:

- The *rate of correct responses* at a given size level will also be referred to as the *recognition rate* or *performance*.
- A *drop* of recognition rate is observed when the recognition rate at a smaller size is lower than at a larger size. A *significant drop* is observed when the reduction in performance is considered significant compared to the performance at larger sizes since the last significant drop, according to Fisher's exact test with *p<0.05*, with drops meeting the stronger criterion of *p<0.01* marked separately. The *initial drop* is the first significant drop detected when analysing performance at decreasing size levels. A *consistent drop* in performance is defined as any drop in performance (significant or not) that is followed up by equal or lower performance at all smaller sizes. *Jumps* in recognition rate are defined in analogous manner for *increases* in performance at smaller sizes, and are expected to be rarely observed, as intuition mandates that recognition rates decrease with stimulus size.
- A *lapse* is defined as any non-significant drop of recognition rate at a smaller size that is followed up by a subsequent increase in recognition rate for at least one yet smaller size. A *singular lapse* is defined as a single incorrect response at a given pairing of size and display. A *significant lapse* is defined as a significant drop in performance followed by a significant jump in performance at a smaller size (this is not expected to be found in the results, and would certainly warrant further investigation).

The results will be presented in groups of related tasks. For each task, the results will be presented in a similar manner according to the following principles. First, the performance of participants with good visual acuity will be reported and presented in graphical form. For each display, significant drops in performance will be identified, as well as any significant differences in performance between displays at any particular size level. A subsequent inspection of the data for each of the three groups of visual acuity (high, medium and low) will be undertaken, and any significant deviations from the overall results will be reported. It needs to be pointed out that due to the small number of trials completed by the group with "low" visual acuity, and due to the participant recruitment strategy, which focused on recruiting young, healthy participants, the results of this group are not to be read as being representative for map users with diminished visual capability, but are provided as hints for potential further investigations into performance at diminished visual acuity, which was not the focus of this study.

An alternative, simplified way of establishing the minimum size of a set of symbols would be to specify the smallest size at which a certain high success rate is achieved by a group of participants. Therefore, limits L_{98} at 98%, L_{95} at 95% and/or L_{90} at 90% recognition rate, ignoring lapses as defined above, will be reported for tasks where this is sensible, for each of the three visual acuity subgroups and for participants with good visual acuity. Which one of these limits can be reported will depend on the overall performance for the specific task (it does not make sense to report a 95% limit for a task which was only completed at 90%

accuracy at the largest size), and considerations of statistical significance (a smaller group size requires a larger threshold difference, as a smaller difference would not be statistically significant).

6.4.1. Tasks 1-4: Discriminating icons from various icons sets, in default rendering

Figure 6.14 shows the results for Task 1, discriminating icons from the set "Maki-rectangular", for participants with good visual acuity. A total of 100 trials were completed by this group for each size level on each display. Recognition rates were high (at about 98% on average) on all displays for sizes above 1 mm. A first significant drop is detected for a decrease of size to 1.0 mm on display D_5, with recognition rates remaining high at this level on the other two displays. For a size of 0.85 mm, performance drops further to 92% on D_5, and around 95% on D_3 and D_4, which is, however, not detected as a significant drop for these displays ($p=0.10$ for D_4). Indeed, performance improves back to 99% at a size of 0.7 mm on D_4, while dropping significantly to 92% on D_3. At this size level, the recognition rate on D_4 is significantly higher than on both other displays. At smaller sizes, recognition rates are below 90% on all displays, with consistently higher performance on D_4, with a significant difference to both other displays again detected at the smallest size of 0.4 mm. However, at this size, recognition rates even on the highest-resolution display are below 70%, with around 50% for the other two displays.

Figure 6.14 Recognition rates for Task 1, participants with good visual acuity (logMAR < -0.05). Vertical black lines indicate a significant drop in performance between size levels, vertical black brackets indicate a significant difference in performance between displays for the same stimulus parameters, (*) *p < 0.05* (**) *p < 0.01*

Participants with high visual acuity showed generally high performance down to a size of 0.7 mm, with lowest recognition rates of 94.4% at sizes of 0.7 mm and 0.85 mm on display D_3. Performance for this group remained comparably high on D_4 down to the smallest size of 0.4 mm, at which a recognition rate of 86.1% was accomplished on this display. For participants with medium visual acuity, a significant drop of performance to 92.2% was observed for sizes below 1.25 mm on D_5, while performance remained high down to a size of 0.85 mm on D_3. On D_4, a recognition rate of 98.4% was accomplished at a size level of 0.7 mm by this group, however there is a significant lapse in performance to be reported with only 92.2% at the larger size of 0.85 mm.

	$L_{98,goodVA}$	L_{95}		
		highVA	*mediumVA*	*lowVA**
D_5	1.25 mm	0.7 mm	1.25 mm	1.5 mm*
D_3	1.0 mm	1.0 mm	0.85 mm	0.85 mm*
D_4	0.7 mm	0.7 mm	1.0 mm	1.5 mm*

Table 6.4 Limits L_{98} (98% recognition accuracy) and L_{95} (95% recognition accuracy) for Task 1. (*) Due to the small group size, results for this group are not statistically valid and are included only for informational purposes.

Results for the second icon set, Maki-triangular, tested in Task 2, are shown in Figure 6.15. A total of 100 trials were completed by participants with good visual acuity for each size level on each display for this task. Performance remained high on all displays down to a size of 0.7 mm, notwithstanding non-significant lapses, with performance below 95% for sizes smaller than that on all displays. At a size of 0.5 mm, performance on D_4 remained significantly higher (at 87%) than on D_3 (at 77%).

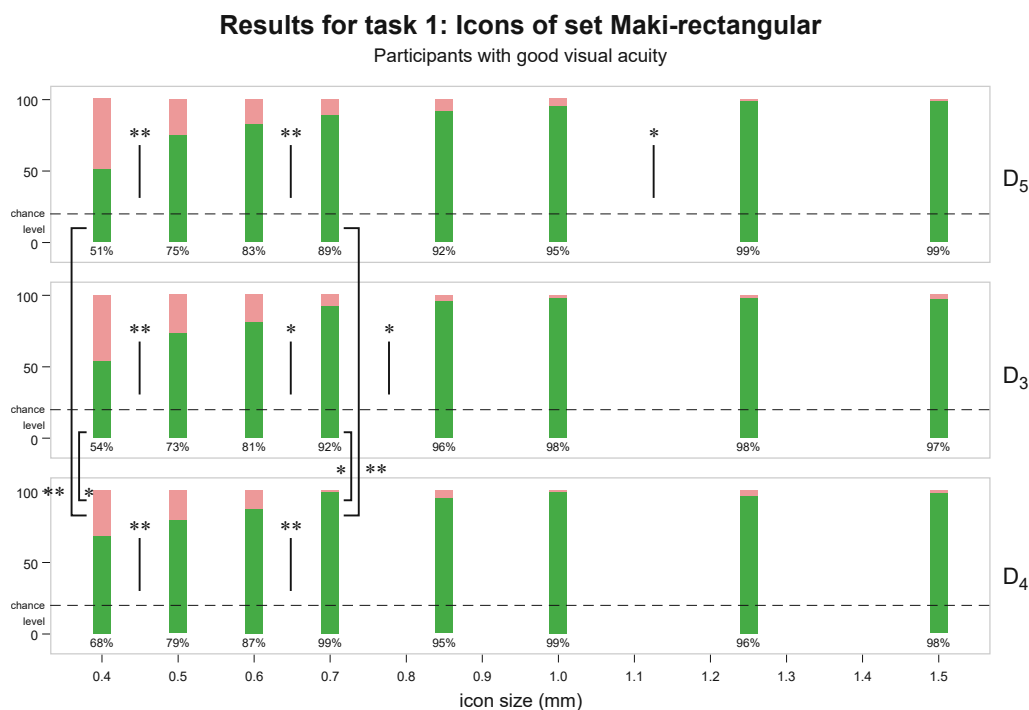**Results for task 2: Icons of set Maki-triangular**

Participants with good visual acuity

Figure 6.15 Recognition rates for Task 2, participants with good visual acuity (logMAR < -0.05). Vertical black lines indicate a significant drop in performance between size levels, vertical black brackets indicate a significant difference in performance between displays for the same stimulus parameters, (*) $p < 0.05$ (**) $p < 0.01$

Participants of highest visual acuity maintained recognition rates at or above 95% for this task down to a size of 0.7 mm at D_5, 0.6 mm at D_3, and 0.5 mm at D_4. Performance at that size on D_4 (97.2%) was significantly higher than on both other displays (80.6% and 83.3% for D_5 and D_3, respectively). For the medium acuity group, significant lapses to 90-92% performance for sizes of 1.0 mm and 1.25 mm were observed on D_5, with performance recovering to above 95% for smaller sizes down to 0.7 mm. On D_3, a non-significant lapse (93.8%) was detected for a size of 0.85 mm, with performance recovering to 95.3% for the size level below. On D_4, performance was consistent down to the size of 0.7 mm, dropping to 93.8% and below for sizes smaller than that. Participants with low visual acuity generally showed only singular lapses down to a size of 1.0 mm on all displays.

	$L_{98,goodVA}$	L_{95}		
		highVA	*mediumVA*	*lowVA**
D_5	0.7 mm	0.7 mm	0.7 mm	1.0 mm*
D_3	1.25 mm	0.6 mm	0.7 mm	0.85 mm*
D_4	0.7 mm	0.5 mm	0.7 mm	0.85 mm*

Table 6.5 Limits L_{98} (98% recognition accuracy) and L_{95} (95% recognition accuracy) for Task 2. (*) Due to the small group size, results for this group are not statistically valid and are included only for informational purposes.

A third set of icons, NPS-vertical, was tested in Task 3, for which the results are shown in Figure 6.16. A total of 73 trials were completed by participants with good visual acuity for each size level on each display for this task (for the first two participants, only 2 trials were shown for each level, which was increased to 3 for the subsequent 23 participants from this

group). A non-significant but consistent drop of recognition rate was observed already for the first decrease of size from 1.5 mm to 1.25 mm on all displays. The further reduction of stimulus size to 1.0 mm resulted in a significant drop in performance to below 90% on D_5 and D_3, while remaining higher, at 93.2%, on D_4. However, the performance difference between D_4 and D_3 at this size level just misses the threshold for significance (*p=0.059*). A further decrease to 0.85 mm resulted in a significant drop of performance also on that display, with no significant differences between displays detected for sizes smaller than 1.0 mm.
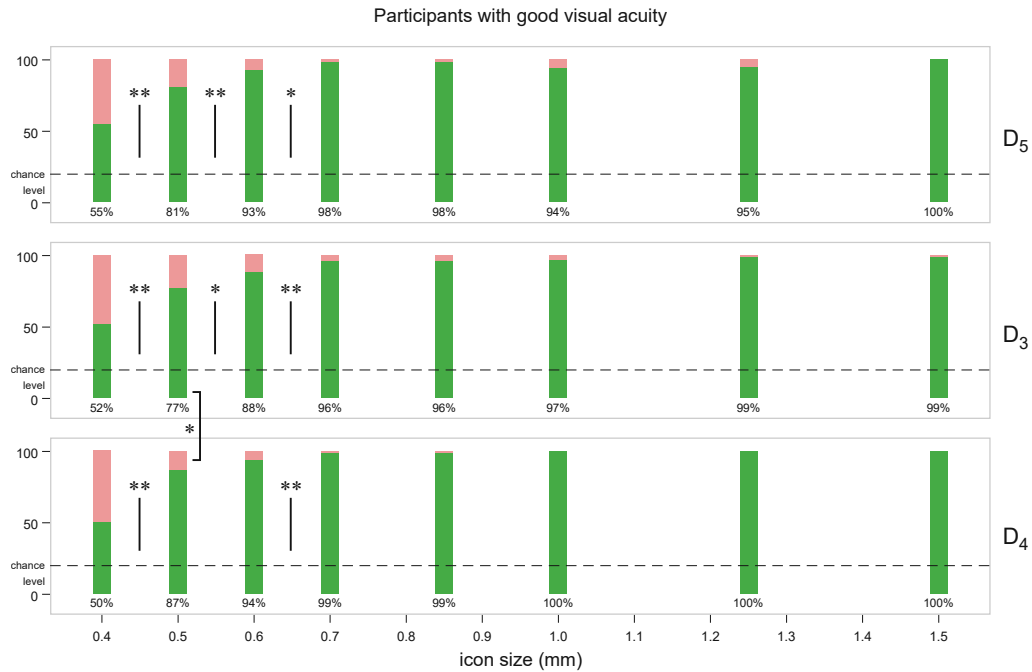

Figure 6.16 Recognition rates for Task 3, participants with good visual acuity (logMAR < -0.05). Vertical black lines indicate a significant drop in performance between size levels, (*) *p < 0.05* (**) *p < 0.01*

The participant group with highest visual acuity did not accomplish recognition rates above 90% for sizes smaller than 1.25 mm on any display, with the exception of a non-significant jump in performance to 92.6% at a size of 0.85 mm on D_3. Participants with medium visual acuity showed more differentiated performance across stations, with recognition rates on D_5 above 90% only at the largest size of 1.5 mm (97.8%), on D_3 down to a size of 1.25 mm (93.5%), and on D_4 down to a size of 1.0 mm (95.7%). On D_5 and D_4, the first drop becomes significant only for a decrease in size from 1.0 mm to 0.85 mm for this group, while on D_3 it remains at below 1.25 mm. The results for the group with lowest visual acuity for this task are not suitable for even informal analysis, due to the low number of trials (12 on each display) at each size level. Success rate across the 3 displays for this group was 91.7% at the largest size of 1.5 mm and dropped to below 70% for sizes of 1.25 mm and smaller.

| | $L_{98,goodVA}$ | L_{95} | | |
		highVA	*mediumVA*	*lowVA**
D$_5$	1.5 mm	1.25 mm	1.5 mm	—*
D$_3$	1.5 mm	1.25 mm	1.5 mm	—*
D$_4$	1.5 mm	1.25 mm	1.0 mm	—*

Table 6.6 Limits L_{98} (98% recognition accuracy) and L_{95} (95% recognition accuracy) for Task 3. (*) Due to the small group size, results for this group are not statistically valid and did not exceed 95% at the largest size of 1.5 mm across all three displays.

Results for the last set of icons used in the study, OSM-castles, are shown in Figure 6.17. Only two trials at each size level were run for this task on each display, which reduces the statistical significance of the results. No significant drop in performance was detected on any station for sizes of 1.25 mm or larger. For a further decrease of size, a significant drop was detected for D$_5$ and D$_3$, with a recognition rate of 88% on D$_4$ not being detected as a significant drop ($p=0.08$) due to the lower level of performance at larger sizes at this station. A consistent decrease of performance is observed on all stations for sizes of 0.85 mm and smaller.
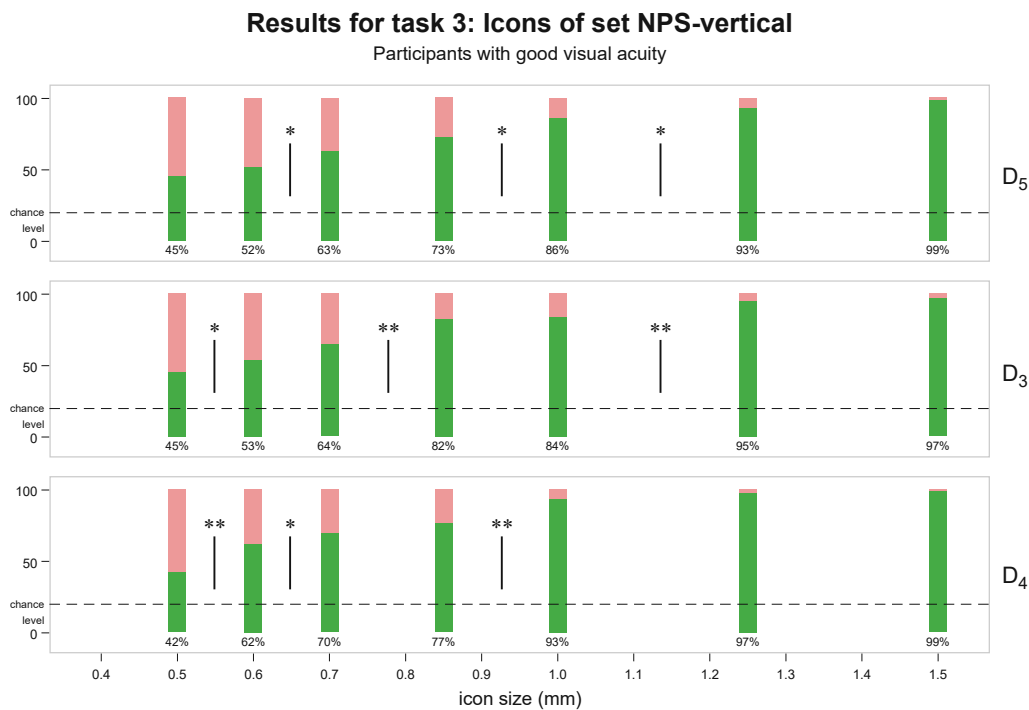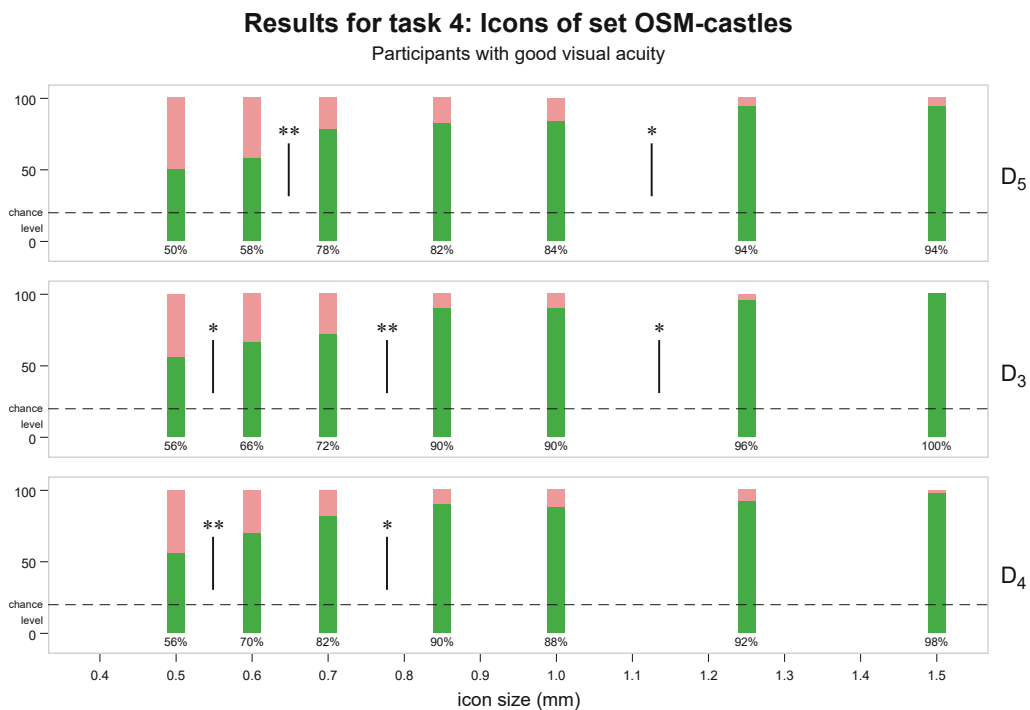

Figure 6.17 Recognition rates for Task 4, participants with good visual acuity (logMAR < -0.05). Vertical black lines indicate a significant drop in performance between size levels, vertical black brackets indicate a significant difference in performance between displays for the same stimulus parameters, (*) $p < 0.05$ (**) $p < 0.01$

The small number of trials significantly limits the validity of any analysis by sub-groups. For the group with highest visual acuity, a consistent decrease in performance can be observed for size levels below 1.0 mm, with only non-significant lapses for larger sizes. The group of medium acuity shows a drop to performance below 85% for a size of 1.0 mm on all three stations, with an increase to higher levels at 0.85 mm for D$_3$ and D$_4$. For the group of lowest

acuity, only a total of 8 trials has been run for each combination of size and display. Only individual lapses have been recorded for this group at size levels of 1.25 mm and 1.5 mm.

6.4.2. Tasks 5, 6, 7: Threshold and grid-fitted icons

As an attempted intervention to improve legibility, icons were rendered as binary black-white bitmaps aligned with the pixel grid for tasks 5-7. For icons from the "Maki-rectangular" set, icons were subjected to automated threshold computation for Task 5, while a version manually fitted to the pixel grid for various integer grid sizes was used for Task 6 (see Section 6.1.3 for a detailed discussion). Because these interventions are related to the physical pixel grid, the analysis of the results will be done separately for each station, with a focus on whether any significant differences compared to the normal rendering of the stimulus (Task 1) can be detected.

Figure 6.18 shows the results of the three tasks for D_5, the display of lowest resolution in the study, for participants with good visual acuity. The stimuli created by automated thresholding led to consistently worse results than stimuli of the next comparable smaller size, significantly so for 2 of the 5 sizes tested (7 and 9 physical pixels). For the icons created manually by adjusting to the pixel grid, performance was very similar to the default rendering, with no significant differences detected. Even the significant drop in performance between 0.7 mm and 0.6 mm (from 89% to 83%) was also detected in the corresponding pixel sizes of 7 and 6 physical pixels (from 89% to 82%). This overall tendency is also reflected in the results of each participant group of high, medium and low visual acuity.

Figure 6.18 Recognition rates on display D_5 (265 ppi) for tasks 1, 5 and 6, for participants with good visual acuity (logMAR < -0.05). Vertical black lines indicate a significant drop in performance between size levels, black brackets indicate a significant difference in performance between stimulus types for similar sizes on the same display, (*) *p < 0.05* (**) *p < 0.01*

Figure 6.19 shows the same comparison for display D_3. Due to the higher resolution, a wider range of pixel sizes (8 to 20 physical pixels) fell within the size range that was considered relevant for this study. For the icons rendered with automated thresholding, a significant jump in performance can be detected when decreasing from a size of 16 pixels to 14 pixels, with performance of the threshold-based rendering at 16 pixels being significantly worse than the comparable size of 0.85 mm with default rendering. However, the threshold-based icons performed significantly better at a size of 13 pixels (91%) than at the comparable size of 0.6 mm with default rendering (81%). For the manually grid-fitted icons, a significant improvement of performance was also detected for a size of 13 pixels (93%) as compared to the default rendering at the comparable size of 0.6 mm (81%). At 12 pixels, the manually grid-fitted icons outperform (89%) the larger size of 0.6 mm as well, but this difference misses the significance mark. At other sizes, performance of the grid-fitted icons was comparable to the default rendering. A significant increase in performance of the icons rendered with automated thresholds over the manually grid fitted icons was detected for very small sizes of 8 and 9 pixels, however at already low success rates of 63% and 74%, respectively.
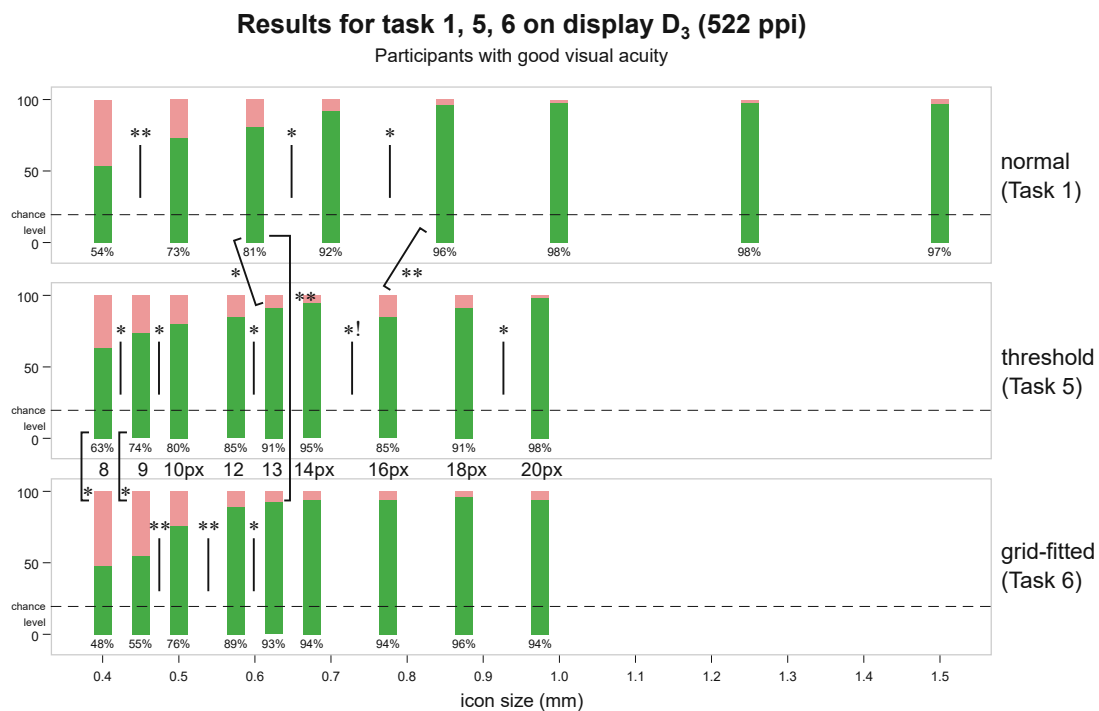
Figure 6.19 Recognition rates on display D_3 (522 ppi) for tasks 1, 5 and 6, for participants with good visual acuity (logMAR < -0.05). Vertical black lines indicate a significant drop in performance between size levels, black brackets indicate a significant difference in performance between stimulus types for similar sizes on the same display, (*) *p < 0.05* (**) *p < 0.01*

Figure 6.20 shows the comparison for the display of highest pixel density, D_4. Due to the small size of physical pixels, the sizes of the grid-fitted icons cover only a small range of approximately 0.35 – 0.65 mm, already below the first significant drop in performance detected for Task 1. Within this range, performance of the grid-fitted icons showed no significant deviations from neither the icons rendered with automated thresholding, nor from the default rendering.
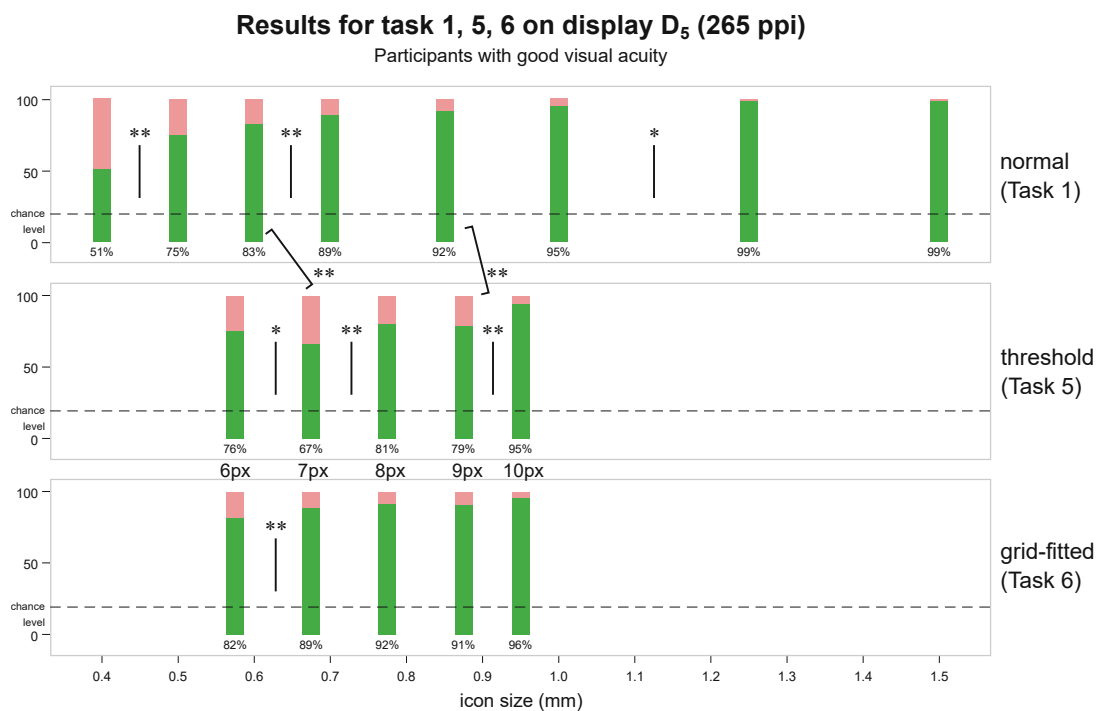
Figure 6.20 Recognition rates on display D_4 (801 ppi) for tasks 1, 5 and 6, for participants with good visual acuity (logMAR < -0.05). Vertical black lines indicate a significant drop in performance between size levels, (*) *p < 0.05* (**) *p < 0.01*

Manually grid-fitted icons were also tested for the icons from the set Maki-triangular in Task 7, for which the results are shown in Figure 6.21. The results are generally consistent with the results of the default rendering for the same icon set (Task 2, shown in Figure 6.15 on page 146), with the exception of a significant drop in performance to 88% for the size of 8 pixels on the lowest-resolution display D_5 (*p=0.002*). This drop is followed by a subsequent increase in performance to 95% for the size of 7 pixels (*p=0.063*), and 91% for the size of 6 pixels on that display, so the drop is qualified as a (significant) lapse for that specific size.

Analysis of participants grouped by visual acuity shows results consistent with the overall trend for this task. The lapse at a size of 8 pixels is detected only for participants of medium visual acuity, of which multiple participants gave incorrect responses at this level. This makes an explanation of the drop by a software or network malfunction unlikely.

Results for task 7: Icons of set Maki-triangular, manually grid-fitted

Participants with good visual acuity

D_5

	6px	7px	8px	9px	10px

** ↓

**

100 / 50 / chance level / 0

91.0%   95.0%   88.0%   97.0%   97.0%

D_3

8px 9px 10px 12px 13px 14px 16px 18px 20px

** **

*

53.0% 64.0% 71.0% 92.0% 92.0% 95.0% 94.0% 93.0% 100.0%

D_4

12px 14px 15px 18px 20px

* **

62.0% 70.0% 77.0% 86.0% 90.0%

0.4   0.5   0.6   0.7   0.8   0.9   1.0   1.1   1.2   1.3   1.4   1.5

icon size (mm)

Figure 6.21 Recognition rates for Task 7, participants with good visual acuity (logMAR < -0.05). Vertical black lines indicate a significant difference in performance between size levels; dashed frame indicates a significant decrease in performance compared to the default rendering (Task 2) at similar size, (*) *p < 0.05* (**) *p < 0.01*

6.4.3. Task 8: Shape contrast amplification

Task 8 presented icons modified with experimental shape contrast amplification to participants (see Section 6.1.4 for explanation). Figure 6.22 presents the results for this task for participants with good visual acuity. At the largest size of 1.5 mm, performance was generally worse than for the unmodified symbols, with a significant decrease on D_5 and D_4, and missing the significance threshold on D_3 (*p=0.085*). Performance for a size of 1.25 mm was comparable to the unmodified symbols. At a size of 1.0 mm there are mixed results, with increased recognition rates when compared to the unmodified stimulus of Task 3 for D_5 and D_3 (only being detected as significant for D_3, *p=0.043*), and a non-significant decrease of performance on D_4. Recognition rates are consistently higher than the ones for the unmodified icons for sizes below 1.0 mm, and meet thresholds for significance when grouped across displays for the sizes of 0.85 mm and 0.5 mm.

Results for task 8: Icons of set NPS-vertical, shape contrast enhancement

Participants with good visual acuity
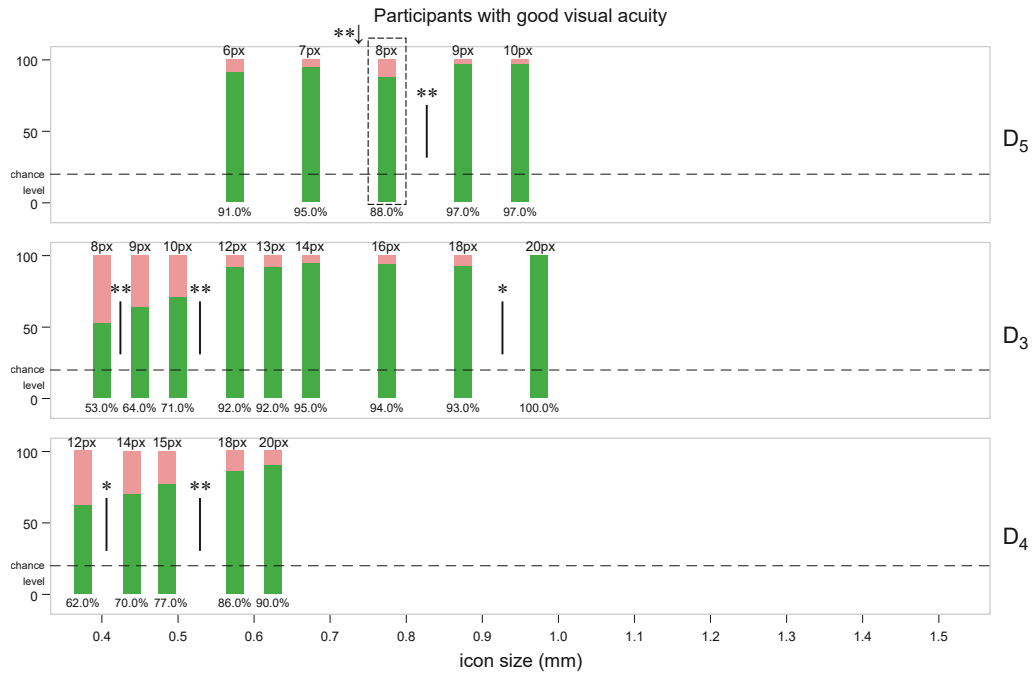

Figure 6.22 Recognition rates for Task 8, participants with good visual acuity (logMAR < -0.05). Vertical black lines indicate a significant drop in performance between size levels, vertical black brackets indicate a significant difference in performance between displays for the same stimulus parameters, dashed boxes indicate a significant increase (↑) or decrease (↓) in groupwise performance over the default rendering (Task 3), (*) $p < 0.05$ (**) $p < 0.01$

6.4.4. Tasks 9, 10: Count icons on a map

Results for Task 9, counting icons from the "Maki-triangular" set on a simple map, are shown in Figure 6.23 for participants with good visual acuity. Due to the introduction of size levels 1.5 mm and 1.0 mm after 4 participants (see Section 6.2.5), only 21 of 25 participants with good visual acuity completed the task at these levels. The results of participants who completed all levels are, however, consistent with the results of the other 4 participants, so the change should not influence the outcome of the analysis, besides slightly reduced statistical significance at the levels introduced later. At the largest size of 1.5 mm, the rate of correct responses was at 95% for all three displays. A significant drop is registered already for the first reduction of size to 1.25 mm on the lowest-resolution display D_5, with a correctness rate of 85% and a significant difference to performance on the two other displays combined. This pattern continues, with consistently lower recognition rates than on the other displays observed for D_5 for all sizes below 1.5 mm, significant differences to D_3 for all icon sizes of 1.0 mm and smaller, and a significant difference to D_4 for the size level of 0.7 mm. D_4 shows significantly worse performance than D_3 for the smallest size of 0.6 mm, with no significant differences between the two higher-resolution displays for larger icon sizes.

Results for task 9: Counting icons from set Maki-triangular
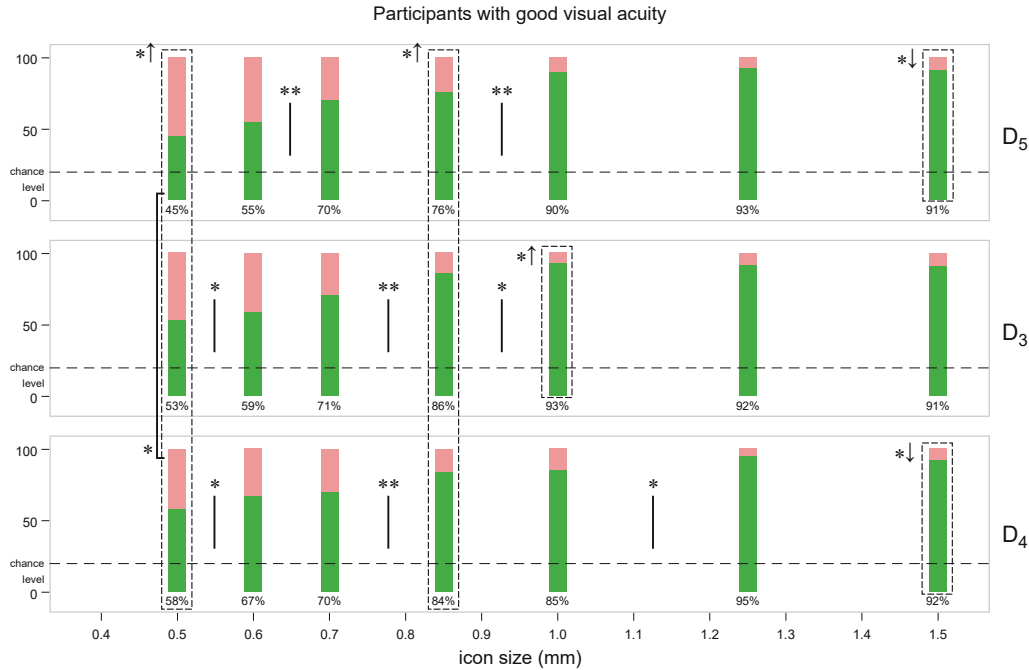
Participants with good visual acuity

Figure 6.23 Rates of correct responses for Task 9, participants with good visual acuity (logMAR < -0.05). Vertical black lines indicate a significant drop in performance between size levels, vertical black brackets indicate a significant difference in performance between displays for the same icon size, (*) $p < 0.05$, (**) $p < 0.01$

The overall trend of reduced performance on D_5 is also observed when participants are grouped by visual acuity. Performance is consistently lower on this display at all sizes below 1.5 mm for participants with high and medium acuity, and also the advantage of D_3 for the smallest stimulus size can be observed for both groups. The small number of trials run for the group of low visual acuity prevents valid statistical analysis, but also for this group the rate of correct answers was at 100% at the two largest sizes for D_3 and D_4, and at 87.5% for each of those sizes at D_5.

	L_{90}			
	goodVA	*highVA*	*mediumVA*	*lowVA*
D_5	1.5 mm	1.5 mm	1.5 mm	— **
D_3	1.0 mm	1.25 mm	1.0 mm	(1.25 mm)*
D_4	1.25 mm	1.25 mm	1.25 mm	(1.25 mm)*

Table 6.7 Limits L_{90} (90% accuracy) for Task 9. *) Too few participants for reliable results. **) Performance for this group was below 90% for the starting size of 1.5 mm.

A similar overall trend can be seen for the results of Task 10, shown in Figure 6.24, for which icons of the set Maki-rectangular had to be counted. Performance on the lowest-resolution display D_5 is consistently lower than on both other displays for all size levels, with a significant difference to D_3 registered for the sizes of 1.0 mm, and a significant difference to D_4 at 0.7 mm. Performance on D_3 and D_4 is largely similar, with the largest difference apparent at an icon size of 1.0 mm (with a success rate of 90.5% at D_3 and 81% at D_4, *p=0.060*), with similar levels of performance for other sizes.

Results for task 10: Counting icons from set Maki-rectangular
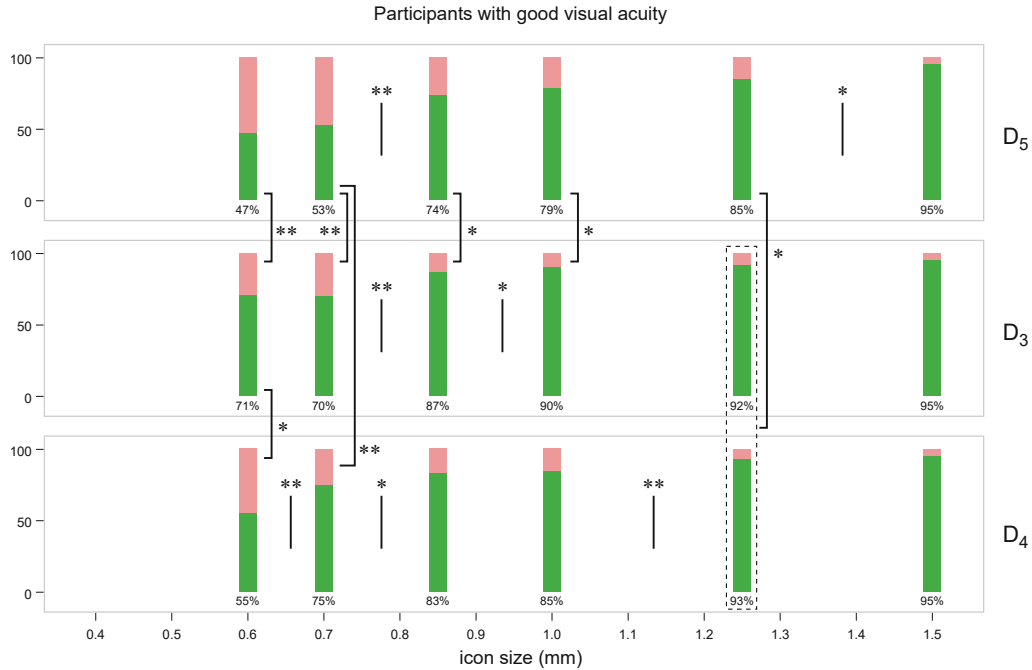
Participants with good visual acuity
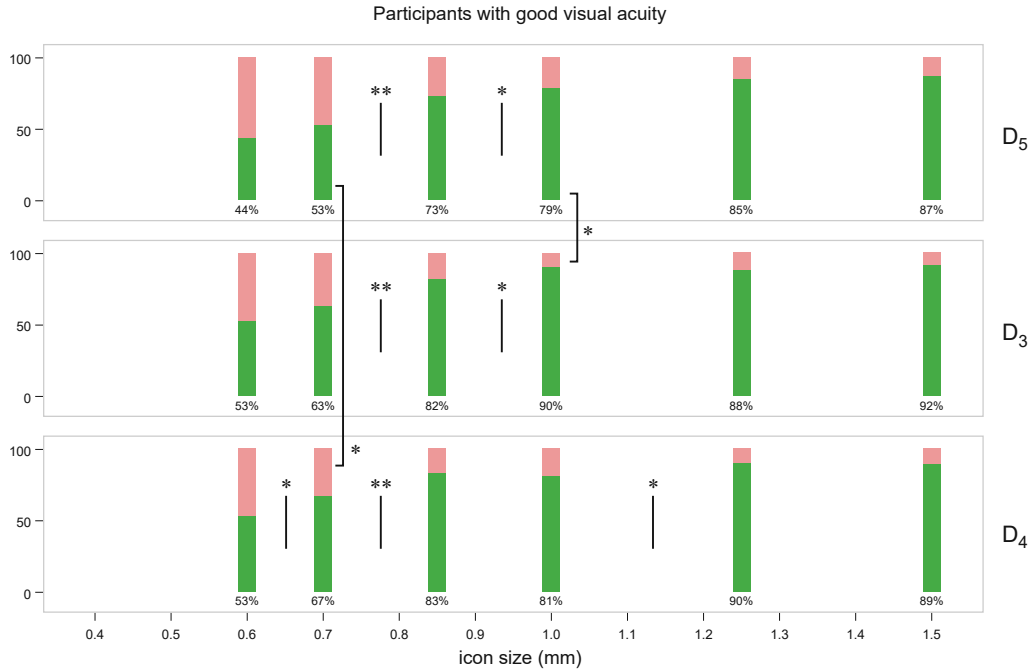
Figure 6.24 Rates of correct responses for Task 10, participants with good visual acuity (logMAR < -0.05). Vertical black lines indicate a significant drop in performance between size levels, vertical black brackets indicate a significant difference in performance between displays for the same icon size, (*) *p < 0.05*, (**) *p < 0.01*

Analysis of participant groups is complicated by the fact that no group accomplished performance above 90% consistently for the three displays for this task. The group of highest visual acuity shows a pattern of *increasing* performance with decrease of size for the first three size levels on D_5, and for the first four levels on D_3, with performance increasing from around 85% at the largest size to 96.4% at a size of 1.0 mm on D_5, and 91.7% on D_3. This effect can be observed independent of station order, making an explanation with effects of learning or fatigue less plausible. This group did not perform worse on D_5 than on the other two displays down to a size of 0.85 mm, when performance starts to drop more sharply than on the other two displays. For the medium acuity group, lower performance on D_5 can be consistently observed at all sizes, for the group with low visual acuity this difference can be observed for sizes below 1.0 mm.

Due to the low overall success rates, with a rate of 90% not reached consistently on any display for the largest two sizes, the threshold for reporting performance limit has been reduced to 85% in order to be able to report tabular results. The values for L_{85} can be found in Table 6.8.

	L_{85}			
	goodVA	*highVA*	*mediumVA*	*lowVA*
D_5	1.25 mm	1.0 mm	1.5 mm	(1.25 mm)*
D_3	1.0 mm	0.85 mm	1.0 mm	(1.5 mm)*
D_4	1.25 mm	1.25 mm	1.25 mm	(1.0 mm)*

Table 6.8 Limits L_{85} (85% accuracy) for Task 10. *) Too few participants for reliable results.

6.5. Discussion

The proposed operationalization for establishing the "minimum size" of a cartographic symbol by detecting the first significant drop in recognition rate when presented at a sequence of decreasing size has shown to deliver useful results when a) the baseline performance at large size levels can be reliably established during the experiment, and b) performance is at a consistent level across secondary variables (in the case of this study: display resolutions) for those larger sizes. This was the case for the results of tasks 1, 2, 3, 8 and 9 in this study. Tasks 5, 6 and 7 were designed with the goal of comparison to the results of other tasks, therefore the establishing of a baseline performance was not relevant in these cases. The approach showed to produce results of limited clarity for tasks 4 and 10, for which baseline performance varied between displays already at the largest size (Task 4), or for which inconsistent performance at larger levels prevented a clear picture from emerging from the statistical analysis (Task 10).

The results for all individual tasks in the context of the hypotheses formulated in Section 6.3 will be discussed in the following paragraphs. For Task 1 (default rendering of "Maki-rectangular" icons, see Figure 6.14 on page 144), a high baseline recognition rate of 97.8% could be established for the two largest sizes (1.5 mm and 1.25 mm) across the three displays. While differences between individual stations do not become significant until the much smaller size of 0.7 mm, the first significant drop for the lowest-resolution display D_5 is detected for the reduction below 1.25 mm, and drops for D_3 and D_4 are detected below the smaller sizes of 0.85 mm and 0.7 mm, respectively. While this is consistent with the overall trend of the data, it also shows a problem with the chosen approach upon detailed analysis: the recognition rate of 95% for a size of 1.0 mm on D_5 is registered as a significant drop, while the same performance at a size of 0.85 mm on D_4 is not considered as a significant drop, because of the slightly lower performance at larger sizes on that display. In this particular case this coincides with a recovery of performance to 99% at the smaller size of 0.7 mm on D_4, so the reduced performance can be classified as a lapse – however, it shows that establishing the baseline performance separately for each display, and requiring a static threshold of $p < 0.05$ for the detection of drops may introduce artefacts in the analysis that may lead to ambiguous conclusions in some cases. This could be mitigated by a) establishing baseline performance at larger levels across all three displays; b) considering the performance not only at larger sizes, but also at smaller sizes for the detection of drops (thus taking the concept of a "lapse" into account when determining baseline performance); and/or c) abandoning the static testing for significance against an (arbitrary) probability level of 0.05 in favour of an approach that quantifies each reduction in performance by specifying its p value on a continuous scale, possibly accompanied by some appropriate means of visualization. For this study, the decision was made to continue using the framework of analysis initially committed to, instead of inventing a new framework post hoc that better fits the results. However, for future studies, a further refinement of the framework of analysis should be considered.

Adopting above ideas for the analysis of Task 1 would have potentially moved the initial drop on D_5 one level down to below 1.0 mm, but would not have changed the overall picture that performance on D_5 was worse than on D_3 (with an initial drop at larger size, and consistently lower performance at the critical sizes of 1.0 mm–0.7 mm) and D_4 (with an initial

drop at a larger size, a significantly worse performance at a size of 0.7 mm, and consistently lower performance at sizes of 1.0 mm or smaller), which supports hypothesis HS2.1. Also, an advantage of D_4 over both other displays is expressed by the initial drop at a smaller size, the significantly better performance at a level of 0.7 mm, and a consistently better performance at sizes at or below 0.7 mm, which *contradicts* hypothesis HS2.2 (positing no significant advantage of D_4 over D_3). HS 2.3 is supported by the fact that no lower-resolution display showed a smaller initial drop or a significant difference at any level over a higher-resolution display.

Task 2 (default rendering of "Maki-triangular" icons) showed fewer pronounced differences between displays, with the initial drop for each display detected for the reduction of size from 0.7 mm to 0.6 mm. No significant difference was detected between D_5 and another display for any size level, retaining the null hypothesis for HS2.1. A significantly higher performance on D_4 over D_3 was detected at the size of 0.5 mm, two levels below the initial drop, at recognition rates of 87% versus 77%. While this *contradicts* HS2.2, a performance of 87% will hardly be acceptable for real-world applications, so this effect is not highly relevant for real-world applications. No advantage of a lower-resolution display over one with a higher resolution has been detected, supporting HS2.3.

Task 3 (default rendering of "NPS-vertical" icons) shows a consistent decrease in performance already between the two largest levels on all displays, which is, however, not significant. Performance at the largest level of 1.5 mm across stations (98.3%) is high enough to suggest that baseline performance has indeed been achieved at this level. The initial drop to below 90% recognition rate was detected for the decrease from 1.25 mm to 1.0 mm for displays D_5 and D_3, and at one level smaller for the decrease from 1.0 mm to 0.85 mm for display D_4. No significant differences have been found between displays for any size level. These results are ambiguous with respect to HS2.1 (only D_4 shows indication for better performance than D_5), *contradict* HS2.2 (with the later drop in performance on D_4 as compared to D_3, which can be read as indication for better performance), and support HS 2.3 (with no lower-resolution display showing an advantage over a higher-resolution one).

As discussed above, the results for Task 4 (default rendering of the "OSM-castles" icons) are problematic, since performance on D_5 and D_4 is consistently lower than on D_3 already for the two largest levels, which casts doubt on whether an accurate baseline performance has been established. This initial performance difference causes a drop in performance to 90% being detected for D_3 below 1.25 mm, while the lower recognition rate of 88% on D_4 is not considered a drop because of the lower performance at higher levels at this display. Statistical significance is also reduced by the fact that only two trials for each size were run for this task. Had the alternate method of establishing baseline performance by considering all three displays been adopted for the analysis, the drop would have been identified as significant for D_4, and not for D_3. In the light of these objections, any conclusions drawn from the results should be treated with caution. In any case, it can be said that there are no large differences apparent between displays, but performance on D_3 (medium resolution) is consistently higher than on both other displays for the largest three size levels, and performance of D_4 at all levels is consistently higher than on D_5. This may be related to the fine white lines present in the icons of the OSM-castles set, which may appear more clearly on the medium resolution

display, and fail to reproduce with sufficient clarity on the lower resolution display. Reading the results in this way would support hypotheses <u>HS2.1</u> and <u>HS2.2</u>, while being in *contradiction* to <u>HS2.3</u>, with the lower resolution display D_3 showing better performance than the higher-resolution display D_4.

Tasks 5 and 6 attempted to increase stimulus contrast and crispness of icons from the "Maki-rectangular" set by rendering at maximum binary contrast, generating only fully white or fully black pixels. Task 5 did so by applying an automated threshold to the rendered image, which was hypothesized to result in diminished performance on displays D_5 and D_3, while not affecting the results much for the highest-resolution display D_4. Hypothesis <u>H2.4</u> is largely supported by the data shown in Figure 6.17–Figure 6.19, where a significantly worse performance at comparable, slightly larger size than with the reference stimulus of Task 1 has been detected for two size levels at display D_5 (with consistently lower performance for Task 5 at all size levels below the initial size of 10 pixels), and for one size level for display D_3. *Contradicting* hypothesis <u>H2.4</u> is the better performance of the threshold icons at the size of 13 pixels for this task, and also the slightly better performance at 14 pixels for Task 5, which may indicate a singular "sweet spot" for which this method of contrast amplification does indeed provide a benefit. However, this effect is observable only at a size already below the initial drop of recognition rate, so a real-world application in which icons of 0.7 mm size are subjected to a threshold operation to boost recognition rates from ~80% to ~90% for a potentially narrow band of display resolutions seems of dubious value. As expected, the thresholding did not result in a detectable difference of performance on D_4 at any size level, which supports hypothesis <u>H2.4</u> where this was anticipated.

For Task 6, icons from the Maki-rectangular set were manually grid-fitted for pixel sizes of 20 down to 6 pixels. The results for this task on display D_5 show very little difference in performance compared to similar size levels of Task 1, including the occurrence of a significant drop in performance at around 0.65 mm. This *contradicts* hypothesis <u>H2.4</u> in which the expectation was expressed that the manually grid-fitted icons would provide a benefit especially on lower-resolution screen. Hypothesis <u>H2.4</u> is supported by a single size level for D_3, where the grid fitted icon at 13 pixels provided a significant advantage over the comparable size of 0.6 mm. For all other size levels, recognition rates are very similar than the ones observed for Task 1. Again, although one size level has been found for which a significant improvement could be shown, it can be doubted that it is of any practical relevance – designers will rarely want to go through the effort of manually creating pixel-based icons, when the promised effect is to improve recognition rates from ~80% to ~90% for a size of 0.6 mm for a potentially narrow band of display resolutions. The results on D_4 for Task 6 again support hypothesis <u>H2.4</u>, as no significant improvement has been detected for the grid-fitted icons over the results of the default rendering of Task 1 on the highest-resolution display.

Task 7 repeated the experimenting with manually grid-fitted icons for icons of the set "Maki-triangular", with no significant advantage over the results of default rendering (Task 2) detected for any size level on any display, again *contradicting* hypothesis <u>H2.4</u> for which a positive effect of the intervention had been expected for the two lower-resolution displays.

An arguably creative intervention was attempted for Task 8 by modifying the geometry of icons from the set "NPS-vertical" with the "shape difference amplification" approach described in Section 6.1.4. As expressed in hypothesis HS2.5, this intervention was expected to have a negative impact on legibility at larger sizes, due to the distortions introduced which may confuse the viewer, and was expected to have a positive impact on legibility at smaller sizes. It was hoped that in the ideal case, the intervention could delay the initial drop in performance to a lower size level. The assumptions of HS2.5 were largely supported by the results of Task 8, where for the largest size level a significantly lower recognition rate than for Task 3 was detected for displays D_5 and D_4, with non-significantly lower performance also observed on D_3. Performance on D_3 was significantly improved as compared to Task 3 for a size of 1.0 mm, with a non-significant improvement observed for D_5 and decrease of performance observed for D_4. At a size of 0.85 mm, a consistent improvement of performance over Task 3 was observed for all three displays, which is shown to be significant when the results from the three stations at that size are grouped together. At lower sizes, a consistent improvement was observed for all size levels (significant again at a size of 0.5 mm). However, even at the largest size for which a significant improvement was observed for all three displays (0.85 mm), the recognition rate of Task 8 was at an average of 82% (85% at stations D_3 and D_4), compared to 77% without the intervention (79.5% at stations D_3 and D_4). While the improvement at smaller sizes appears more consistent for the intervention of "shape difference amplification" than for the grid-fitted icons, again the effect becomes only noticeable at sizes for which the recognition rate would probably be considered unacceptable for most real-world applications. Nevertheless, the consistent improvement at small sizes and the accurate prediction of results by hypothesis HS2.5 (the negative effect on performance at larger sizes could be mitigated by only activating the method when very small icons need to be presented) suggest an opportunity for future research into the novel approach attempted here.

Task 9 asked participants to count icons from the "Maki-triangular" set on the map. The results show a success rate of 95% at the largest size, which unfortunately drops already for the second level (significantly on D_5), so that it cannot be confidently assumed that the 95% establish a valid baseline performance for the task. The significant and consistent drop on D_5 – performance at all sizes but the largest one is significantly lower than on at least one other display – strongly *contradicts* hypothesis H2.6, which assumed that display resolution would not make a significant difference for the map-counting tasks. It certainly supports hypothesis H2.3, for which it was assumed that a lower resolution would not provide an advantage for this type of task. Hypothesis H2.7 is also strongly supported, with performance at every size level on every display being lower than the simpler task of identifying the icons of the same set presented in isolation (Task 2). Initial drops for Task 2 were detected below 0.7 mm for all three displays, whereas for Task 9, initial drops (from an already lower initial performance of 95%) were detected between levels of 1.5 mm (for D_5) and 1.0 mm (for D_3).

The analysis of the results for Task 10 (counting icons from the "Maki-rectangular" set on a map) suffers from the difficulties discussed at the beginning of this section. Initial performance at the largest size level is below 90% for displays D_5 and D_4, and only slightly higher for D_3, thus the confidence in having established a valid baseline performance for the

task is low (there is no reason to believe baseline performance would be lower than the 95% of Task 9) and the detection of initial drops with the chosen approach must be considered unreliable. Despite those limitations, the conclusions for Task 9 are supported also by the results of this task, with significantly diminished performance on the lowest-resolution display D_5 for two distinct levels *contradicting* hypothesis H2.6 which assumed that no such difference would be encountered. Hypothesis H2.3 is again supported, as the lower resolution was certainly of no advantage for the task. Hypothesis H2.7 is technically only supported by the results for D_3 and D_4, for which the initial drops occur at larger levels than the ones for the simple icon discrimination task (Task 1). The initial drop detected for D_5 at the small size of 1.0 mm must be questioned because of the low starting performance for that display, and is therefore not a strong argument against H2.7.

Section 6.3.2 presented a set of hypotheses for minimum dimensions derived from simple models of visual acuity (the one-arcminute-model) and display fidelity (models of 1 CSS pixel or 1 physical pixel as minimum size for graphical detail), which can now be contrasted with the empirical results. The models of one physical pixel as the minimum size for graphical detail can be rejected, as they would result in sizes well below acceptable levels of performance for displays D_3 and D_4 due to the small size of physical pixels for these displays (for example, Table 6.3 calculates a minimum size of 0.65 mm for D_3 and 0.4 mm for D_4 for tasks 1 and 2 for these displays, which the interested reader can compare with the results for these sizes and displays themselves). The one arcminute model and the CSS pixel model would command a minimum size of 1.3 mm and 1.65 mm, respectively, for icons from the Maki set, which both are considerably above the initial drop detected on any display for tasks 1 and 2. Adapting the model of smallest graphical detail to ½ CSS pixel would result in a mandated minimum size of the Maki icons of 0.8 mm, which is above the initial drop for all displays for Task 2 (Maki-triangular), and above the initial drop for D_3 and D_4 for Task 1 (Maki-rectangular). For the NPS-vertical icons (Task 3), this would result in a mandated minimum size of 1.4 mm, which is also above the initial drop for all three displays. For the OSM-castles icons (Task 4), which are designed on a 14 × 14 units grid, a minimum dimension of ½ CSS pixel per grid unit would result in a size of 0.9 mm, which is below the initial drop for 2 of 3 displays for the task, at a level at which recognition rates between 80% and 90% were accomplished by participants with good visual acuity. This would be considered too unreliable for most real-world applications, even for information of low importance. However, it must be kept in mind that the OSM-castles icons proved most difficult to discriminate for the participants of this study, which may also be caused by their semantic ambiguity. Future research on the impact of semantic difference in addition to purely perceptual difference on the performance of icon discrimination tasks would certainly be justified.

Chapter 8 will present the attempt to derive a set of practical guidelines for map designers from the results of this study and Study 3, which will be presented in the next chapter.

7. Study 3: Minimum dimensions of cartographic line symbols on smartphone displays

The first study provided some initial insights into the discriminability of basic symbology types for straight line segments – lateral and longitudinal line patterns. A general tendency of a significant disadvantage of the lowest-resolution display, but only minor differences among the three phones of higher pixel density, became apparent in the results. Similar to study 2, which investigated the legibility of point symbols in more detail and tested additional ways in which the graphical fidelity of higher-resolution displays could be exploited, the third study attempts a closer investigation of symbology for line segments.

It needs to be noted that the study presented in this chapter was designed and implemented at the same time as the second study, and was conducted before the analysis of the earlier studies' results was undertaken. Therefore, the findings presented in the previous chapter did not yet influence the design of this study and the methodology for analysis.

7.1. Study Design

7.1.1. Devices and experiment configuration

The overall experiment configuration and the devices used for this study remained identical to study 2. Also, the method of constant stimuli described in Section 6.2.1, presenting a fixed number of stimuli at discrete, hand-picked sizes was adopted again for this study.

7.1.2. Stimulus and task design

7.1.2.1. Task 1: Detect orientation of arrows embedded in line

The first stimulus type used for this study explores the idea that on high-resolution displays, additional geometry may be embedded within the boundaries of a solid line to communicate additional information. Arrows are a common symbol to denote the direction of flow or movement on a map, and directional information may be an important component for unidirectional pathways such as bicycle lanes or one-way streets on road maps congested with other information. Thus, the space-saving embedding of arrow symbols would be a desirable form for conveying this information without using additional space on the map.

For the stimulus of Task 1, a black straight line of length 50mm was rendered with repeated white triangles placed along the centre axis of the line (see Figure 7.1). The geometry of the stimulus was controlled with the following parameters (values in brackets were used for Task 1): *arrow width*, as a proportion of the line (2/3), *arrow length*, as multiple of arrow width (2.0), *gap length* between the arrows, as multiple of arrow length (1.0), *angle*, the angle at which the line is drawn. The space remaining at the beginning and the end of the line was distributed equally between both ends. Figure 7.1, left, illustrates how the geometry of the stimulus was constructed from these parameters.
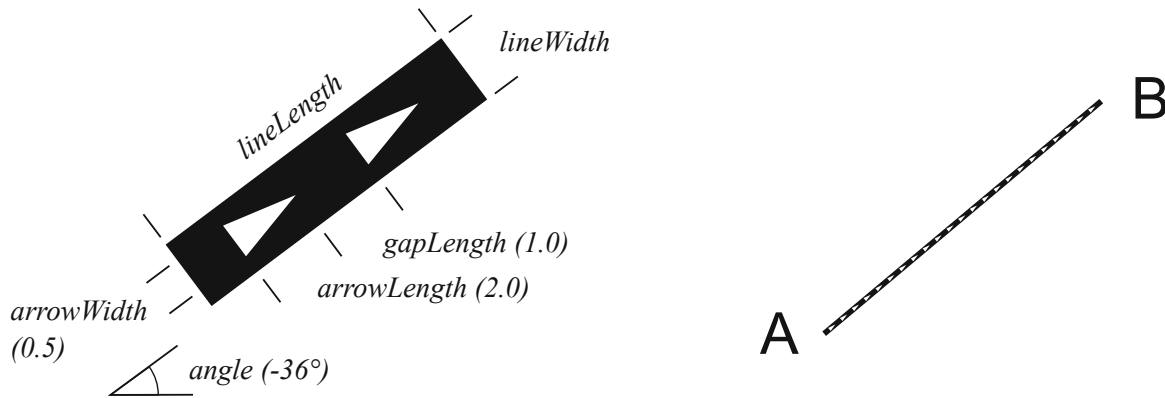
Figure 7.1 Left: Parameters controlling the geometry of the embedded arrows, with example values in brackets. Right: Stimulus with the parameters chosen for Task 1, at initial width (0.8 mm), reproduced as vector graphics.

The letter "A" was rendered next to the leftmost end point, and the letter "B" was rendered next to the right most end point. On the response device, two buttons were provided, labelled "A ►►► B" and "A ◄◄◄ B", respectively.

Before the task was run, a short instruction was displayed on the stimulus display, reading "Next Task: Press the button on the response device that matches the orientation of the arrows. Press «Continue» when you are ready.". The angle of the line was oriented at an angle chosen at random between 10° and 70° counter-clockwise rotation from the horizontal orientation (angles close to a vertical orientation were not used to ensure a clear left-right ordering of the lines end points). Orientation of the arrows was chosen at random for each trial between the "forward" orientation (A → B) and the "reverse" orientation (B ← A). Four trials each were run at the following progressively smaller line width values at each display: 0.8 mm, 0.6 mm, 0.5 mm, 0.4 mm, 0.3 mm, 0.25 mm.

7.1.2.2. Task 2: Detect orientation of arrows embedded in line, amplified with green dots

One idea for a "contrast amplification" approach for the stimulus used in Task 1 would be to emphasize the arrow corners with high-contrast graphical elements. As has been discussed in sections 3.1 and 3.2, green subpixels are particularly suited for this purpose, because out of the three primary pixel colours they will result in the strongest contrast sensation and are matched with the highest density of photoreceptors in the fovea. The approach for Task 2 attempts such contrast enhancement by purely geometrical means, without knowledge of the relationship of the stimulus geometry to the pixel grid or the display's subpixel structure, by placing green (RGB code #00ff00) circular discs with a diameter of 1.1 physical pixels at each corner of the embedded arrows (see Figure 7.2 for an illustration). This should, in theory, result in only green subpixels being lit and the green subpixels near the arrow corners being lit close to their maximum intensity.

It has to be remarked that a) the stimulus geometry constructed in this way is not identical for all displays for a given line width, since the width of the green discs is specified in physical pixels in an attempt to use the smallest high-intensity light emission available on each display, and b) that is could be argued that this modification enlarges the overall shape of the
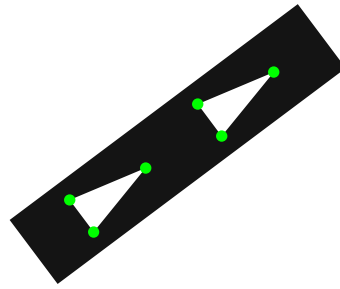
Figure 7.2 Attempted contrast amplification by placing green circular discs at the corners of each arrow for Task 2. The discs have a constant diameter of 1.1 physical pixels at each display station. See Figure 7.3, middle row, for a microscopic image of the actual stimuli rendered in this manner.
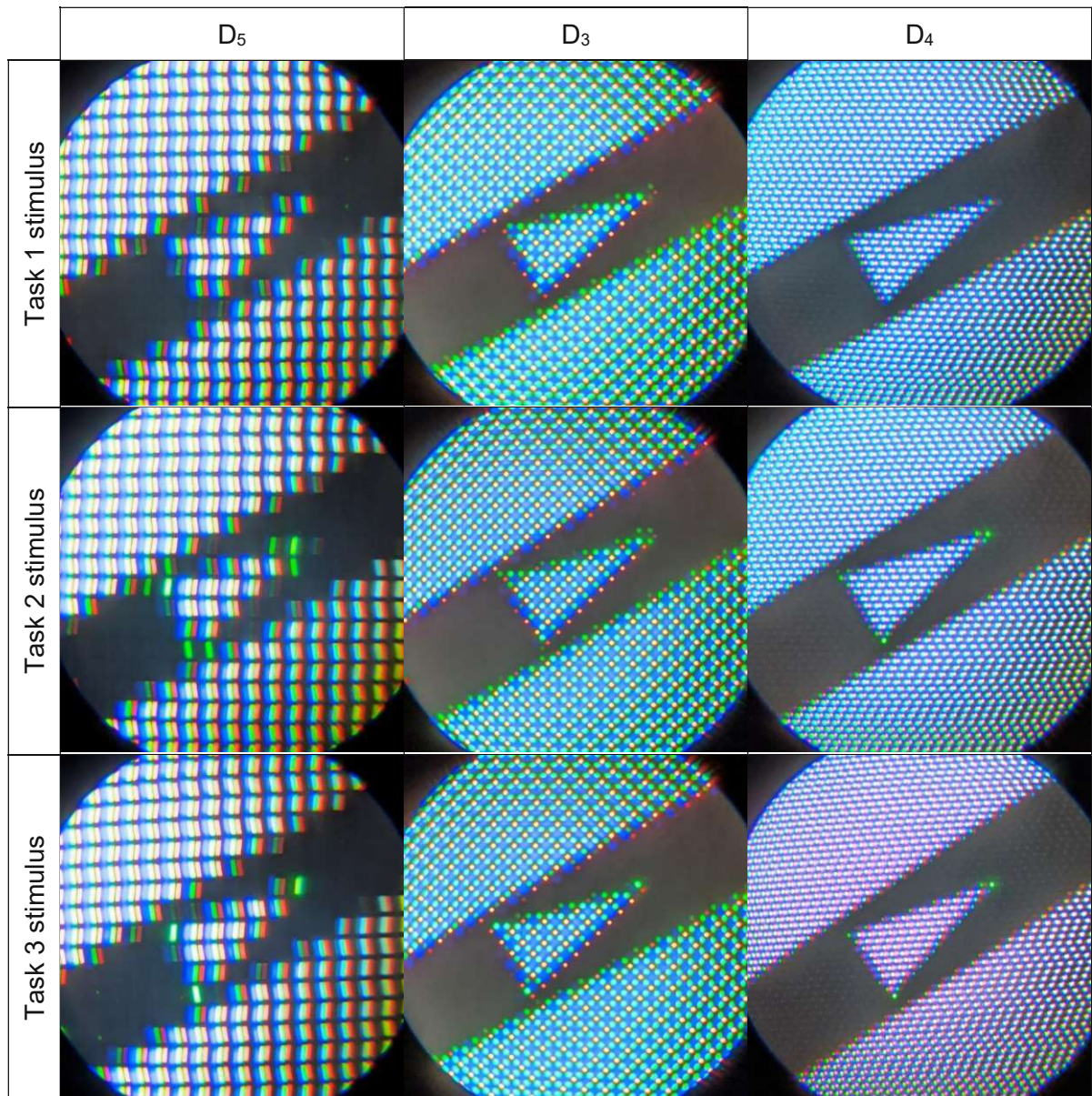
Figure 7.3 Microscopic images, magnified approximately 30×, of stimuli for tasks 1, 2 and 3 at a line width of 0.5 mm. Top row: unmodified geometry used for Task 1. Middle row: green discs with a diameter of 1.1 pixels placed at corner points, used for Task 2. Bottom row: Green pixels placed at corner points, aligned with the physical pixel grid, used for Task 3. (These images are best viewed on a screen for accurate reproduction of RGB colours)

arrow, since the green discs extend beyond the boundaries of the arrow. These aspects have to be considered when assessing the results for this task.

Besides the rendering of the stimuli, all parameters of tasks 2 (size progression, arrow dimensions, line angle, response interface etc.) were identical to Task 1.

7.1.2.3. Task 3: Detect orientation of arrows embedded in line, amplified with pixel-aligned green dots

In a variant of the contrast amplification idea described for Task 2, the stimuli for Task 3 had green dots added at the corners of each triangle, but these dots were aligned with the physical pixel grid. In the implementation, the coordinates of each triangle's corners were calculated, and rounded to the nearest full physical pixel, which was set to RGB colour #00ff00. While this procedure causes the highlight dots to not be perfectly aligned with the arrow geometry, it should ensure that only the green subpixel(s) of a single pixel are lit for each of the three corners. Figure 7.3 shows microscopic images of the stimuli created in this way (bottom row) and a comparison to the stimuli for tasks 1 and 2.

Again, the parameters for Task 3 (size progression, arrows dimensions, line angle, response interface etc.) were kept identical to Task 1 to allow comparison of results.

7.1.2.4. Task 4: Discriminate between lines of variable width

Figure 7.4 Stimulus display (left) and response interface (right) for Task 4, for a base line width of 0.5 mm and a width factor of 1.5.

Task 4 was designed with the goal to investigate whether increased pixel density has an impact on the discriminability of lines of different widths. For this purpose, the stimulus display was divided vertically into two regions. On the left half, four solid black lines of increasing width were displayed, labelled A, B, C, D, at a constant angle of 10° and a length of 16 mm. On the right half, a single solid line matching in width one of the four candidate lines was displayed. The response device showed four buttons, labelled A, B, C and D. Before the task was run, the instructions reading "Next Task: Select the line on the left which best

matches the line shown on the right. Press «Continue» when you are ready." were displayed on the stimulus display.

A constant factor by which the candidate lines differed in width was chosen sequentially from the progression 1.5, 1.33 and 1.25. For each of these width factors, nine trials were run with three trials each for the width of the narrowest line (line A) progressing through 0.5 mm, 0.2 mm and 0.1 mm. This resulted in the widths of the four candidate lines to progress for three trials each through the sequence: 0.5-0.75-1.12-1.69 mm, 0.2-0.3-0.45-0.68 mm, 0.1-0.15-0.23-0.34 mm, 0.5-0.67-0.88-1.18 mm, 0.2-0.27-0.35-0.47 mm … and so on. One of the four candidate widths was chosen at random and used to render the stimulus line on the right, at a length of 40mm and a random angle between 50 and 75 degrees from horizontal.

7.1.2.5. Task 5: Detect orientation of internal partial hachure, narrow spacing

A further exploration of the possibilities of differentiating line symbology within the space occupied by the line symbol was undertaken with tasks 5 and 6. The stimuli were composed of a white line with black outline, with a regular sequence of short black lines (dashes) protruding from the outline inwards at a specific angle on both sides of the line (see Figure 7.5 for illustration).
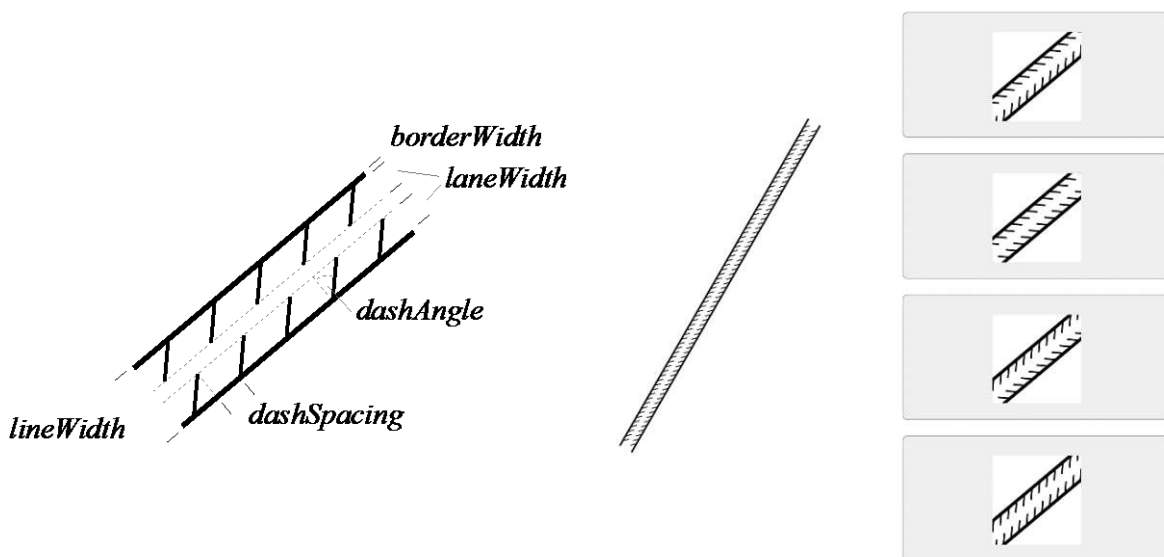

Figure 7.5 Internally structured lines for tasks 5 and 6. Left: parameters for specifying line geometry (dashed lines are construction lines and not part of the stimulus). Middle: screenshot of stimulus for Task 5 on D_4, at the initial line width of 2 mm. Right: response choices offered to the participant (scaled down to 75%).

Conceptually, the overall line was divided into three lanes (left, middle, right), of which the dashes occupied the outer lanes. The following parameters were available to define the geometry of the overall line symbol (constant values for all stimuli in brackets): width of the overall line (variable), width of the outer boundary lines as a fraction of overall line width (1/9), width of the right and left lane as a fraction of overall line width (8/27), width of the dash lines as a fraction of overall line width (1/18), spacing factor between the dashes (1), angle of the dashes in the right and left lanes (±45°).

For each trial of Task 5, orientation of the dashes on each side was randomized between positive and negative 45 degrees. The overall angle of the line was randomized between 30° and 60° from horizontal orientation. The stimuli were presented for three trials at each of the following overall line widths: 2 mm, 1.5 mm, 1 mm, 0.8 mm, 0.6 mm.

The response interface presented four choices as buttons for the four permutations of dash orientations: [45°,45°], [45°,-45°], [-45°,45°], [-45°,-45°] (see Figure 7.5, right). On each button, an example of the corresponding stimulus line was displayed, at a width of 5 mm, length of 20 mm and an angle of 40 degrees. Before Task 5 was started, the following instructions were displayed on the stimulus display: "Next Task: Press the button on the response device that best matches the type of line shown. Press «Continue» when you are ready."

7.1.2.6. Task 6: Detect orientation of internal partial hachure, wide spacing

Task 6 was designed identical to Task 5, with the exception of the spacing of dashes, which was increased by a factor of 1.7. Stimuli were presented for three trials at each of the following overall line widths: 1.5 mm, 1 mm, 0.8 mm, 0.6 mm, 0.5 mm.
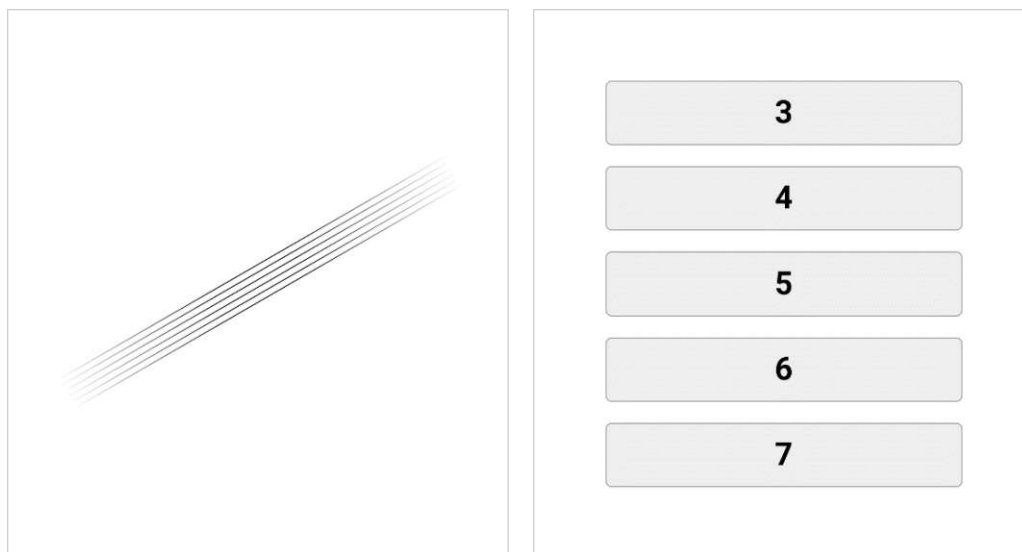
7.1.2.7. Task 7: Count parallel lines

Figure 7.6 Screenshots of stimulus on D_4 (left) and response interface for Task 7. The stimulus shown here has a line width of 0.12 mm and a gap width of 0.75 mm.

Some real-world map use tasks require the counting of narrowly-spaced parallel lines, such as estimating the steepness of terrain from contour lines (Arnberger & Kretschmer, 1975). The goal of Task 7 was to verify whether display pixel density has a significant impact on the performance of participants for such tasks. For this purpose, a number of parallel lines were displayed as stimulus, at different line widths and inter-line gaps. All permutations of line widths of 0.12, 0.08, 0.05 and 0.03 mm and inter-line gaps of 0.75, 0.5 and 0.25 mm were computed and randomly shuffled, with two entries for each permutation. To prevent participants using the end points of the lines as cues for counting, the lines were faded out to gradually lower intensity at both ends. This was implemented by defining a gradient which

transitioned from full white at both ends to full black at 40% of the line length from both ends, and stroking each of the parallel lines with that gradient.

For each trial of Task 7, a random combination of line width and inter-line gap was drawn from the shuffled permutations, and a random number of 4 to 6 parallel black lines was drawn at maximum contrast, at a random angle in the range from 15 to 70 degrees from horizontal orientation. The response interface offered five choices, labelled with numbers "3" to "7". Before the start of Task 7, the following instructions were displayed on the stimulus display: "Next Task: Count the number of parallel lines. Press «Continue» when you are ready."

Because of the randomized sequence of stimulus dimensions, a single dummy trial was run at maximum line width (0.12 mm) and an inter-line gap of 1 mm before the regular trials, in order to assure that the first trial presented the participant with a stimulus of comparably low difficulty.

7.1.2.8. Task 8: Count lines of specific type on pseudo-map

The tasks presented so far have presented stimuli in isolation at maximum contrast. Similar to the approach of study 2 for point symbols, it appeared desirable to verify the ecological validity of discriminability of fundamental line symbols in a setting that more closely resembles real-world applications. Like for tasks 9 and 10 of study 2, initially it was planned to use real-world GIS data to construct the base maps for the task. However, in contrast to point symbols, which can be arranged somewhat freely on an existing base map, line symbols have to be integrated in a way consistent with the overall spatial structure of the map while maintaining consistent and roughly equal dominance of individual features (e.g. line length) to avoid bias. Furthermore, for a task that asks the user to count the number of lines, the lines have to be clearly distinct from each other, and ambiguous situations like Y-junctions at flat angles etc. should be avoided. An informal and subjective review of GIS data for Austria, derived from OpenStreetMap, did not yield a map area that contained a suitable arrangement of roads to be used as a base map for the task. Therefore, a synthetic pseudo-map was created, in which the linear features and their junction points could be arranged freely and in a more consistent way.

Figure 7.7 shows the two pseudo-maps that have been constructed for the task. Starting point for constructing the maps was a square of 60 × 60 mm filled with 5% grey (RGB code #f2f2f2), on which a regular 3 × 3 grid of points was placed, inset 7 mm from the edges. These points were slightly displaced manually for a more organic appearance. Each point in the grid was connected to its closest neighbours in the cardinal directions with a freely drawn polyline. For constructing these 12 polylines, the following guideline were formulated: 5 lines should follow an only slightly curved path, 5 should contain more pronounced curves, and 2 lines should contain sharp hairpin bends. At the junctions (the 9 points as described previously), the tangent of each line should be sufficiently different such that the appearance of continuation of one line beyond the connection of two neighbouring points is avoided. At each point, a circular disc with a diameter of 3 mm, with solid white fill and a solid black outline of 0.3 mm width, is placed "over" the lines (later in the drawing order). A second

pseudo-map was created in a similar way, re-using some of the elements from the first map in a translated or rotated arrangement.
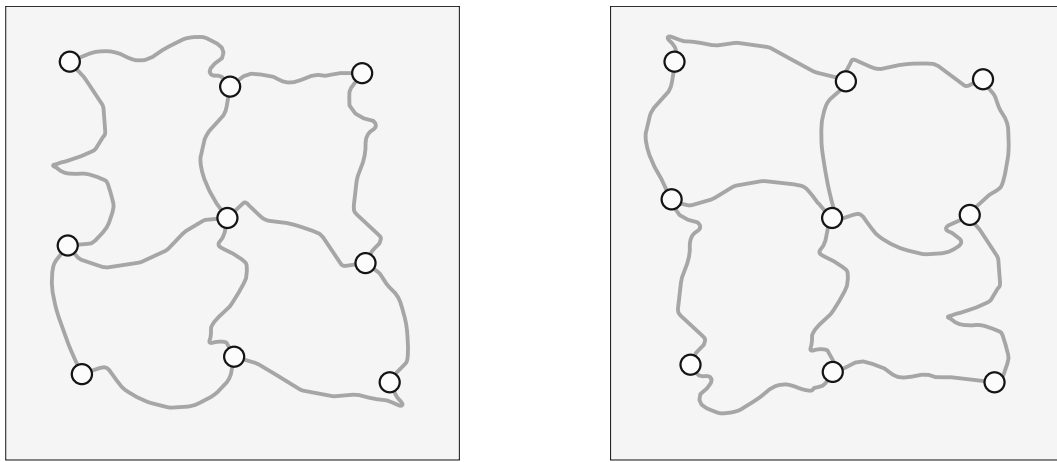

Figure 7.7 Pseudo-maps used for Task 8, reproduced as vector graphics at stimulus size. Grey lines were replaced by actual stimulus lines for each trial.

For constructing the stimuli, one of three types of lines was assigned to each line segment of the pseudo-map: double lines, composed from a black line of total line width and a white line of $1/3^{rd}$ of the line width drawn on top of that; triple lines, composed from a black line of total line width, a white line of $2/3^{rd}$ of the line width, and a black line of $1/5^{th}$ of the line width, stacked on top of each other; and a solid grey line of total line width, filled with a 40% grey value (RGB code # 989898). Each line segment on the map was replaced with one of these line types, at identical line width for all segments, for each trial.

For each trial, a target line type (double, triple or solid grey) and a number of occurrences ranging from 1 to 7 was chosen at random. The other two line types were assigned to the remaining line segments in equal proportion. The response interface offered buttons labelled from 0 to 12. Four trials each were run at the following line widths, in randomized order: 0.75 mm, 0.6 mm, 0.5 mm, 0.4 mm, 0.3 mm.
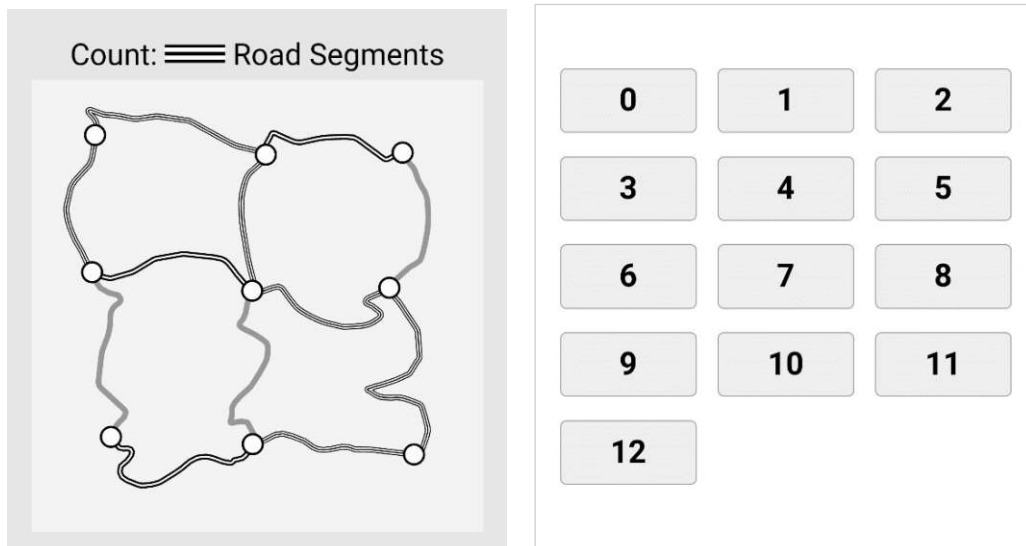
Figure 7.8 Example stimulus (left) and response interface (right) for Task 8, for the starting line width of 0.75 mm.

In order to explain how line segments are supposed to be counted, and to clarify ambiguous cases in which participants may erroneously perceive two line segments of the same type on either side of a junction point as a single continuous segment, the trials for Task 8 were preceded by two dummy trials with a line width of 1 mm and a predefined arrangement of lines, including such an ambiguous case. Before these dummy trials, the following instructions were displayed at the stimulus display: "Next Task: Count the line segments of the type indicated on top. (The first two correct answers are "4") Press «Continue» when you are ready."

7.1.3. Participants

For the third study, 30 participants were recruited in April and May 2022 among students of the course "Web Mapping" (summer term 2022), with different students recruited for this study than for the second study. Bonus points amounting to 5% of the courses overall points were awarded as a compensation for the time spent travelling to and from the experiment. The same protocol as described in Section 6.2.4 for the second study was kept for the third, including the verbal reminder to participants that their participation was fully voluntary and that any bonus points for the course were awarded as soon as a participant arrived at the experiment location.

Of the 30 recruited participants, 28 showed up for participation and completed the experiment. 13 participants identified as male and 15 as female on the questionnaire. 15 participants were aged 16-25 years, 12 were in the 26-35 years age group, and one in the 36-45 years age group. Again, participants were asked to arrive wearing any vision correction as they would normally use, and the effective habitual visual acuity was assessed during the experiment using the "Tumbling E's" task on display D_4.
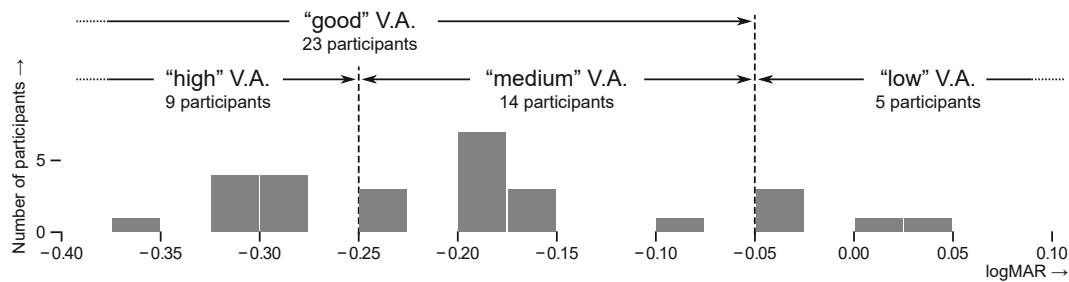
Visual acuity of participants, study 3

Figure 7.9 Distribution of visual acuity of participants and visual acuity groups used for the analysis for the second study. Visual acuity has been assessed with a "Tumbling E" task on display D_4.

Figure 7.9 shows the visual acuity of participants as assessed by the Tumbling E's task. The results show that, like in the assessment for the second study, one participant reached a measured logMAR score below -0.35, and another 8 participants reached a score around the -0.30 mark, supporting the hypothesis that the chosen method for assessment underestimates logMAR scores, as discussed in Section 6.2.4. The overall distribution of logMAR scores was similar to the one observed for study 2, and the same thresholds for dividing participants into visual acuity groups was kept for consistency. Again, the main focus of the study shall lie on the group with "good" visual acuity, supplemented by analysis of the other groups for a more differentiated insight. The small size of the "low" acuity group which, at a configuration of four trials per condition, will result in only about 20 trials per condition being run for this group will again limit the statistical significance of any findings for this group. The results of this group are included as initial information here, and any conclusions about the population segment with visual acuity below "20/20 vision" will require more research with a suitable recruiting strategy to assess a representative sample of the intended target demographics.

7.1.4. Changes during the experiment

In the course of running the study with actual participants, it became apparent that the smallest size level for the map search task (Task 8) yielded very few correct responses for all three displays, and that the initial two levels were separated by a too large gap. Therefore, after 10 participants the smallest line width of 0.25 mm was removed and a width of 0.6 mm was introduced. The removed level will not be taken into account for the analysis of results, and the removal should have had no effect on the overall experiment. The level which has been introduced will have a smaller number of trials in total, which will reduce the significance of the statistical analysis. This will be taken into account in the results section.

7.2. Hypotheses

7.2.1. Hypotheses on the relationship between size and performance

The framework for the detection of significant drops in performance established for the second study in Section 6.3.1 will be applied in analogous fashion for this study. For each

reduction of size, an implicit hypothesis will be tested that performance at the reduced size is significantly lower than performance at larger sizes.

7.2.2. Hypotheses derived from the one-arcminute and one-pixel models

Similar to the second study, an attempt will be made here to apply the simple models of a "minimum graphical unit" of 1 minute of arc, and 1 CSS pixel to the stimuli for this study. The model of one physical pixel as the minimum unit of graphical detail have already been rejected in the discussion of the second study, as this resulted in estimates that were significantly too low for D_3 and D_4 due to the small dimensions of individual pixels on these higher-resolution displays. Therefore, only the two models independent of physical pixel size will be considered here.

For Task 1, the assumption is made that the arrows embedded in the lines need to be 3 units wide at minimum, which results in a predicted minimum line width of 0.45 mm for the one-arcminute model and 0.55 mm for the CSS-pixel model. It is unclear how the interventions of tasks 2 and 3 affect legibility, so no prediction is made for these tasks.

For the discrimination of line width (Task 4), the simple models would mandate that the difference in width approaches one graphical unit (0.1 mm for the one-arcminute model and 0.13 mm for the CSS-pixel model). This would predict that for a base line width of 0.5 mm, lines will be successfully discriminated at the smallest width factor of 1.25; for a base line width of 0.2 mm, the one-arcminute model would predict reduced performance for the factors below 1.5, while the CSS-pixel model would predict reduced performance for all width factors; for a base line width of 0.1 mm, both models would predict reduced performance, since the difference in width between candidate lines is well below the graphical unit mandated by the models.

For the internal hachures used for tasks 5 and 6 it is unclear how a simple model would predict performance of such a complex stimulus. As an initial estimate, it will be assumed that each of the two "lanes" has to be 3 units wide, in order to reliably detect the orientation of the short hachure lines. This results in a predicted minimum overall line width of 1.0 mm for the one-arcminute model, and 1.25 mm for the CSS-pixel model. It is unclear how the wider spacing of hachure lines of Task 6 will affect performance, therefore an identical assumption is used as a working hypothesis for this task.

For the more complex tasks involving counting of lines (tasks 7 and 8), it is expected that raw resolution capability will not be the decisive factor for success, and drops in performance will be encountered at larger sizes than predicted by any simple model, so no predictions are made for these tasks.

7.2.3. Further task- and display-specific hypotheses

The following further hypotheses relating to the performance for individual tasks or displays have been formulated before the start of the experiment:

- <u>HS3.1</u>: The first study found some advantage of the higher display resolution of D_3 over D_2 for tasks related to the legibility of lines (tasks 2 and 3 of the first study), but not as

pronounced as for the icon-related tasks. An advantage of D_3 over the lower-resolution display D_5 is therefore expected to show for most of the tasks of this study, with the exception of tasks 2 and 3 where the larger pixel size is expected to result in a more pronounced effect of the intervention. This would show as a significant difference in performance for participants with good visual acuity between those displays for some identical size levels. For Task 7 (counting parallel lines) the lines are separated by gaps much larger than one pixel, so the increased resolution is not expected to have a significant effect in this case.

- HS3.2: The highest resolution display D_4 did not provide an advantage for the line-related tasks in the first study, and correlated with slightly lower performance for the lateral line pattern task (discriminating parallel lines) in the earlier study, although not significantly so. It will be interesting to see whether such detrimental effect of higher resolution can be reproduced for any of the tasks of the third study. Generally, a significant advantage of the highest-resolution display D_4 over D_3 is not expected, because its resolution is generally considered beyond the visual capabilities of humans. It may provide an advantage for the fine structures of tasks 5 and 6, which are therefore exempt from the hypothesis.

- HS 3.3: An advantage of a lower-resolution display over one of higher resolution is not expected for tasks 1 and 4-8. Despite the results of Task 1 of the first study, which contradicted this hypothesis, the specific circumstances of a detrimental alignment of the stimulus geometry with the pixel grid is not expected for the stimuli of this study. The interventions of tasks 2 and 3, however, may be more pronounced on the lower-resolution displays, so these tasks are exempt from the hypothesis.

- HS3.4: The intervention of amplifying the directional arrows with green dots of a diameter of 1.1 physical pixels (Task 2) will be most visible on the lowest-resolution display D_5, and is expected to compensate for any shortcomings in fidelity on that display. It would therefore be expected that this intervention has the strongest effect on improving performance on D_5. It is unclear how performance on the other two displays is affected by the intervention.

- HS3.5: The pixel-grid aligned green dots (Task 3) are designed to create the smallest possible visual sensation on each display by lighting only a single subpixel. It is therefore expected that performance is improved on D_5 when compared to Task 1. The subpixels on D_3 and D_4 are considered too small to create an effect that impacts performance for this task.

- HS3.6: For discriminating fine lines of different width (Task 4), while the higher resolution of D_4 could be of advantage for accurate reproduction of the stimulus geometry, the larger pixels of lower-resolution displays in combination with antialiasing have shown to be able to compensate for limited spatial resolution. A significant difference of performance across displays is therefore not expected.

- HS3.7: For the lines with internal hachures (tasks 5 and 6), particularly for the narrowly spaced variant (Task 5), an advantage of the display with highest resolution D_4 is expected, particularly for participants with high visual acuity.

- HS3.8: The wider spacing of the internal hachures of Task 6 is expected to result in better overall recognition rates, and mitigate any disadvantages of lower-resolution displays that may have been apparent for Task 5.

- <u>HS3.9</u>: For the map counting tasks (Task 8), display resolution is not expected to have much impact, since the resolution of peripheral vision is considered far below the resolution of any displays, and the more complex task of counting will begin to be affected at sizes at which the resolution is not a limiting factor.

7.3. Results

The same framework of analysis and terminology than in the second study will be applied for the results of this study. Readers are referred to Section 6.4 for a detailed discussion.

7.3.1. Task 1-3: Detect orientation of arrows embedded in line

Recognition rates for Task 1 were above 98% for line widths of 0.5 mm or larger at all displays. A significant drop in performance can be observed for the 0.4 mm width for the lower-resolution displays D_5 and D_3, while at D_4 recognition rate remained at 99% at this level and a drop is only observed one level smaller at 0.3 mm. Results are shown in Figure 7.10.
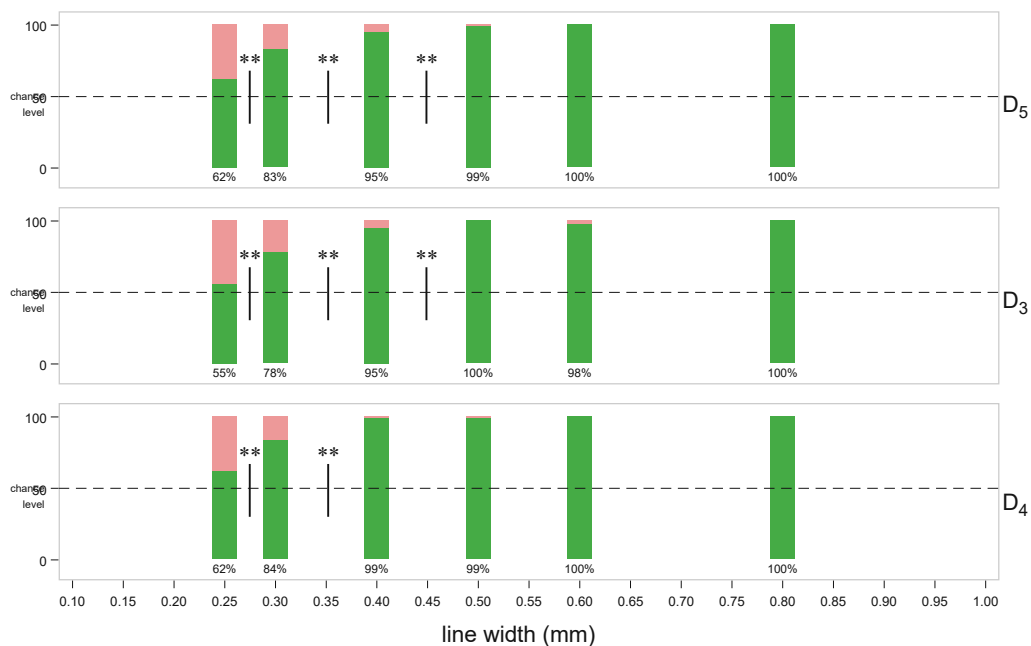

Figure 7.10 Recognition rates for Task 1, participants with good visual acuity (logMAR < -0.05). Vertical black lines indicate a significant drop in performance between stimulus sizes (**) *p < 0.01*

A similar pattern can be observed for both subgroups with high and medium visual acuity. The group of low visual acuity shows a significant drop to 85% recognition rate at the level of 0.4 mm also for display D_4 (from 100% at larger levels), with a slightly higher performance at this level, at 90% recognition rate, for D_3. At 0.5 mm or larger, all responses of this group were correct with the exception of singular lapses.

	$L_{98,goodVA}$	L_{95}		
		highVA	*mediumVA*	*lowVA*
D₅	0.5 mm	0.5 mm	0.4 mm	0.5 mm
D₃	0.5 mm	0.4 mm	0.5 mm	0.5 mm
D₄	0.4 mm	0.4 mm	0.4 mm	0.5 mm

Table 7.1 Limits L₉₈ (98% recognition accuracy) and L₉₅ (95% recognition accuracy) for Task 1.

Task 2 placed green dots with a diameter of 1.1 pixels at the corners of the embedded arrows as an intervention to increase the contrast of the arrow shapes. As has been pointed out, due to the differences in pixel size, the intervention results in a larger change for lower-resolution displays, thus a larger effect for lower resolutions has been expected. The results for Task 2 are shown in Figure 7.11 for participants with good visual acuity. At the lowest resolution display D_5, recognition rates remained at 99% for line widths down to 0.4 mm, with a significant drop to 95% for a width of 0.3 mm. Recognition rates for D_3 and D_4 remained high down to a size of 0.4 mm, with a significant drop at 0.3 mm. Recognition rates at line widths of 0.3 and 0.25 mm were significantly higher on the lowest-resolution display D_5 when compared to both D_3 and D_4, and when compared to the performance on that same display for Task 1.
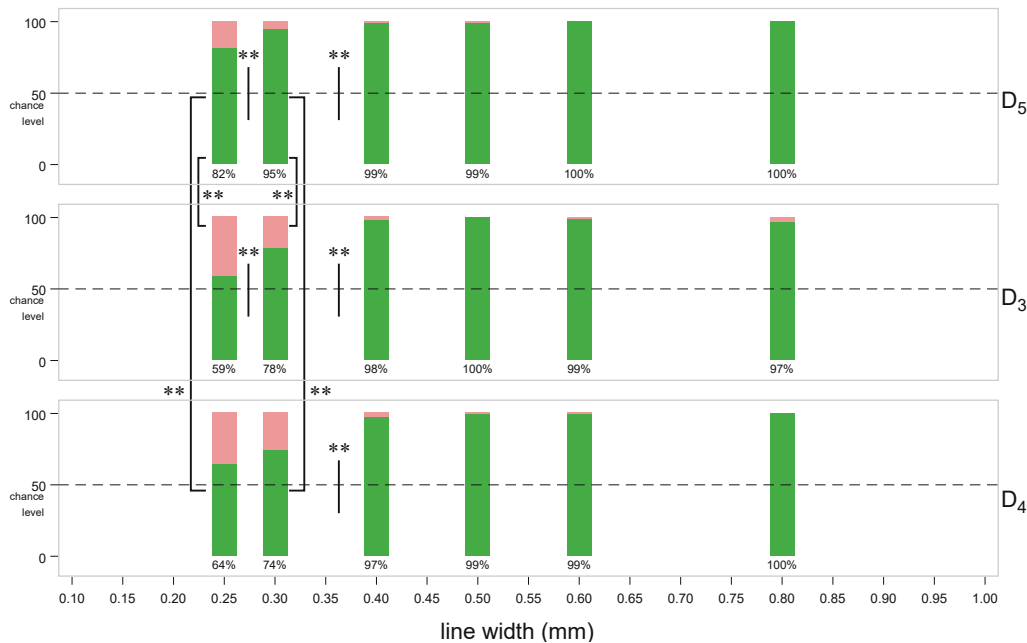

Figure 7.11 Recognition rates for Task 2, participants with good visual acuity (logMAR < -0.05). Vertical black lines indicate a significant drop in performance between stimulus sizes, black brackets indicate a significant difference in performance between displays at the same size, (**) *p < 0.01*

For the user group with high visual acuity, recognition rates remained at 100% down to a width of 0.3 mm on display D_5, and down to a width of 0.4 mm on D_3 and D_4, dropping below 95% for smaller sizes at each of the displays. Users with medium acuity maintained a recognition rate above 98% (only singular lapses) for line widths of 0.5 mm or larger, and for

0.4 mm on D_5, with a drop to ~95% at this size for the two higher-resolution displays. In the low acuity group, only singular lapses occurred down to a line width of 0.5 mm, and down to 0.4 mm on D_5, confirming the trend of a strong impact of the intervention for the lowest-resolution device also for this group.

	$L_{98,goodVA}$	L_{95}		
		highVA	*mediumVA*	*lowVA*
D_5	0.4 mm	0.3 mm	0.4 mm	0.4 mm
D_3	0.4 mm	0.4 mm	0.4 mm	0.5 mm
D_4	0.5 mm	0.4 mm	0.5 mm	0.5 mm

Table 7.2 Limits L_{98} (98% recognition accuracy) and L_{95} (95% recognition accuracy) for Task 2.

The intervention was modified for Task 3 by bringing the green dots into alignment with the physical pixel grid and colouring precisely the one green subpixel closest to each corner of the triangles. Figure 7.12 shows the results for this task for participants with good visual acuity.

Recognition rates do not drop significantly for sizes at or above 0.4 mm for D_5 and D_3. On D_4, a significant drop is detected for a line width of 0.4 mm, to a recognition rate of 95.7% for that level. It has to be noted, however, that the detected significance of this drop is also an artefact of the perfect recognition throughout the larger sizes on this display – had 3 lapses been registered for larger sizes on this display (as is, for example the case for the results on D_3 for this task), the p-value for this drop would have been calculated at 0.07 and therefore classified as non-significant. At a line width of 0.3 mm, performance on D_5 and D_3 is better than on D_4, with a significant difference when combined as a group (*p=0.047*).
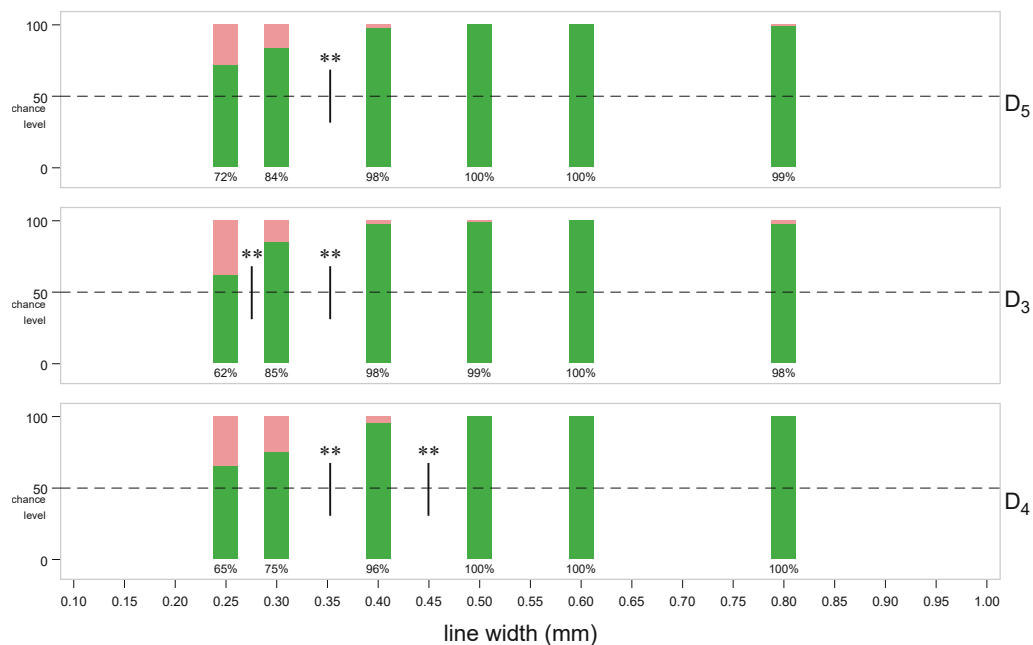

Figure 7.12 Recognition rates for Task 3, participants with good visual acuity (logMAR < -0.05). Vertical black lines indicate a significant drop in performance between stimulus sizes, black brackets indicate a significant difference in performance between displays at the same size, (*) *p < 0.05* (**) *p < 0.01*

The drop in performance for the 0.4 mm stimulus on D_4 also shows in the results for the participant group with high visual acuity (94.4% compared to a perfect 100% for larger sizes), with a higher recognition rate of 97.2% (corresponding to only a single incorrect response out of 36 trials) at the next smaller size of 0.3 mm on this display. For displays D_5 and D_3, recognition rates remain at 100% down to a line width of 0.4 mm, dropping to 94% on D_5 and 92% on D_3 for a line width of 0.3 mm for this participant group. For participants with medium visual acuity, recognition rates show only singular lapses for line widths down to 0.5 mm, dropping to a rate of 96.4% on all three displays for a size of 0.4 mm, and below 80% for smaller sizes. A notable drop of performance is apparent for this participant group at a size of 0.3 mm on D_4, with a recognition rate of 60.7% only slightly above chance level. For this display and size, the inter-group difference between high and medium visual acuity is significant at *p < 0.001* – see Figure 7.13 for a visualization of the results for these two participant groups on display D_4.
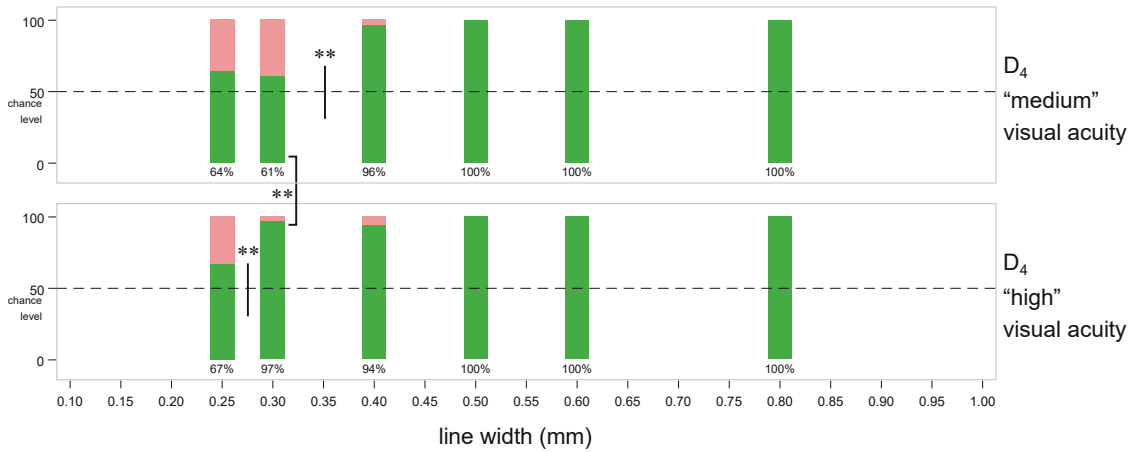

Figure 7.13 Recognition rates for Task 3 on display D_4 for participant groups with "medium" and "high" visual acuity. A significant drop at a line width of 0.3 mm can be observed for participants with lower visual acuity. (**) *p < 0.01*

For participants with low visual acuity, a drop in performance can be observed on all displays for line widths below 0.5 mm. For line widths below 0.4 mm, recognition rates drop to chance level for this group.

	$L_{98,goodVA}$	L_{95}		
		highVA	*mediumVA*	*lowVA*
D_5	0.5 mm	0.4 mm	0.4 mm	0.5 mm
D_3	0.5 mm	0.4 mm	0.4 mm	0.6 mm
D_4	0.5 mm	0.3 mm	0.4 mm	0.5 mm

Table 7.3 Limits L_{98} (98% recognition accuracy) and L_{95} (95% recognition accuracy) for Task 3.

7.3.2. Task 4: Discriminate between lines of variable width

Task 4 asked participants to judge the width of a line by comparing it to four candidate lines. The stimuli for this task varied by two controlled parameters: the width of the thinnest line (candidate A), and the factor by which the width was multiplied to yield subsequent line widths for the other three candidates. For the overall analysis, results of participants with good visual acuity have been grouped by display and width factor, and within each group arranged by base line width (see Figure 7.14). Within each group, drops in performance at subsequent base line widths have been tested for significance. Additionally, differences between displays for the same stimulus parameters (base line width and width factor), groupwise differences for different width factors on the same display, and groupwise differences for identical width factors across displays have been tested for significance.
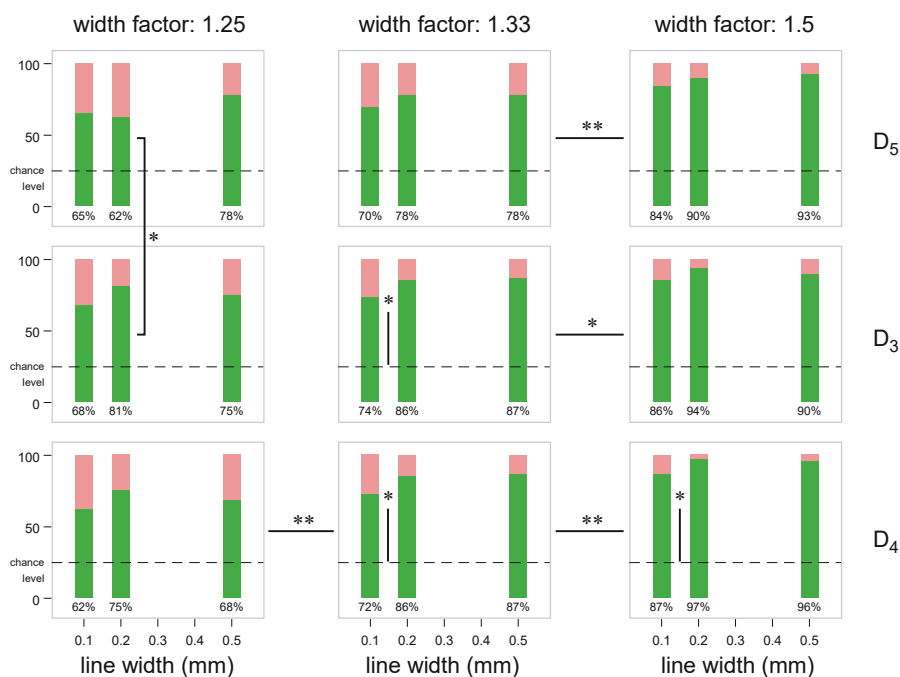

Figure 7.14 Recognition rates for Task 4, grouped by width factor, participants with good visual acuity (logMAR < -0.05). "Line width" refers to the width of the narrowest candidate line. Vertical black lines indicate a significant drop in performance between line widths within a group, vertical black brackets indicate a significant difference in performance between displays for the same stimulus parameters, and horizontal black lines indicate significant difference in performance between groups across width factors, (*) $p < 0.05$ (**) $p < 0.01$

The analysis shows few significant differences between displays – only a base line width of 0.2 mm with the smallest width factor of 1.25 showed a significant difference between D_5 and D_6, with recognition rates at 62% and 81%, respectively. For the larger width factors of 1.33 and 1.5, base line widths of 0.5 mm and 0.2 mm showed similar performance, with a drop in performance for the 0.1 mm base line width which was present in all six groups at these factors, but registered as significant in only three of them. On all three displays, overall performance for a width factor of 1.5 was significantly better than for a factor of 1.33. A further decease of the width factor to 1.25 shows a decrease of average performance on all three displays, but is tested as significant only for D_4. D_4, the highest-resolution display, is the only display for which a significant difference in performance is registered for each

decrease of width factor, and for the narrowest base line width as compared to the two wider ones for width factors 1.33 and 1.5.

Recognition rates for all participants for this task are around 95% for the combination of highest resolution display (D_4), largest width factor (1.5) and the two larger base line widths (0.2 mm and 0.5 mm), with rates general below 90% for smaller width factors and the smallest base line width of 0.1 mm. Participants with high visual acuity accomplished recognition rates above 90% for the following six combinations of display and width factor/baseline width: D_5: 1.5/0.2 mm; D_3: 1.5/0.5 mm, 1.5/0.2 mm; D_4: 1.5/0.5 mm, 1.5/0.2 mm, 1.33/0.2 mm. Participants with medium visual acuity accomplished recognition rates above 90% on five combinations: D_5: 1.5/0.5 mm; D_3: 1.5/0.2 mm, 1.33/0.5 mm; D_4: 1.5/0.5 mm, 1.5/0.2 mm. Participants with low visual acuity accomplished such rate of correct recognition on four combinations: D_3: 1.5/0.2 mm, 1.5/0.1 mm, 1.33/0.5 mm; D_4: 1.5/0.2 mm.

7.3.3. Tasks 5-6: Detect orientation of internal partial hachure

Task 5 presented participants with a line with internal hachures on the left and right side, each at an angle of plus or minus 45 degrees to the line orientation. For participants with good visual acuity, recognition rates were above 98% for an overall line width of 1.5 mm and 2.0 mm, and for a line width of 1.0 mm on D_4. Performance for this size dropped to around 95% on the other two displays, with a significant drop detected only for D_3. Below 1.0 mm, a significant drop in performance was detected for all three displays, with recognition rates on D_4 (91.3%) remaining significantly higher than on D_5 (75.4%).
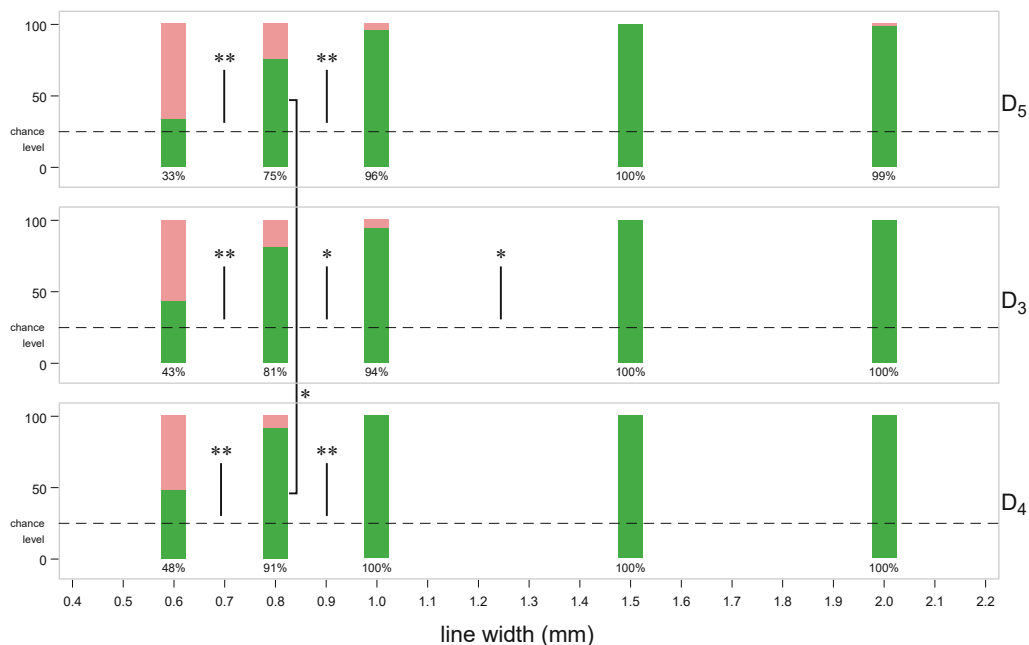

Figure 7.15 Recognition rates for Task 5, participants with good visual acuity (logMAR < -0.05). Vertical black lines indicate a significant drop in performance between stimulus sizes, black brackets indicate a significant difference in performance between displays at the same size, (*) $p < 0.05$ (**) $p < 0.01$

Participants with high visual acuity maintained a perfect 100% recognition rate on D_4 down to a line width of 1.0 mm, with only singular lapses on the other two displays at that size. At

a line width of 0.8 mm, recognition rates were above 95% for this group on the two higher-resolution displays D_3 and D_4, and below 90% on D_5. For the group with medium visual acuity, recognition rates were above 95% at a line width of 1.0 mm for displays D_5 and D_4, with a rate of 92.9% at that size for the medium-resolution display D_3. Performance dropped below 90% on all three displays for sizes under 1.0 mm for this group. Participants with low visual acuity showed only individual lapses at sizes of 1.5 mm or larger, and down to 1.0 mm on D_4. Performance drops below 90% for smaller sizes for this group.

| | $L_{98,goodVA}$ | L_{95} | | |
		highVA	mediumVA	lowVA
D_5	1.5 mm	1.0 mm	1.0 mm	(1.5 mm)*
D_3	1.5 mm	0.8 mm	1.5 mm	(1.5 mm)*
D_4	1.0 mm	0.8 mm	1.0 mm	(1.0 mm)*

Table 7.4 Limits L_{98} (98% recognition accuracy) and L_{95} (95% recognition accuracy) for Task 5. *) a singular lapse at each of these levels resulted in an effective accuracy value of 93.3% due to the small number of trials.

Task 6 used the same stimulus, at a wider spacing between the individual lines of the hachures. A similar pattern than the one for Task 5 can be seen for this task, with perfect recognition or only singular lapses for the group with good visual acuity at a size of 1.5 mm (note that a line width of 2.0 mm was not presented for this task in contrast to Task 5), recognition rates between 92.8% (D_5) and 100% (D_4) at 1.0 mm, and around 90% on all displays for a line width of 0.8 mm. Display D_4 is the only display that facilitated a 100% recognition rate at a line width of 1.0 mm for this user group for tasks 5 and 6 combined.
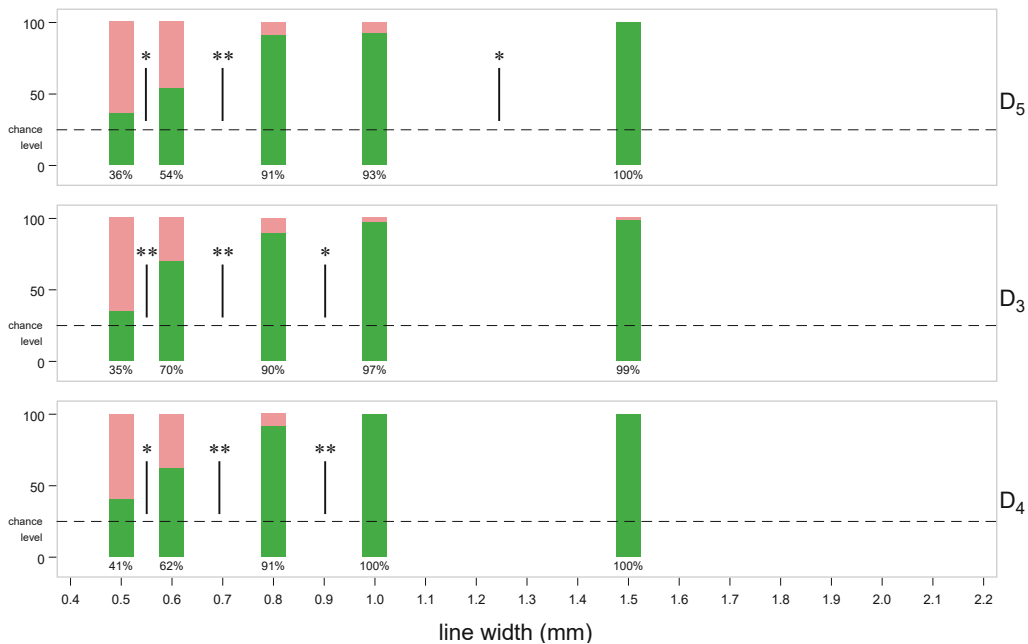

Figure 7.16 Recognition rates for Task 6, participants with good visual acuity (logMAR < -0.05). Vertical black lines indicate a significant drop in performance between stimulus sizes. (*) *p < 0.05* (**) *p < 0.01*

Users with high visual acuity maintained a 100% recognition rate down to 0.8 mm on all displays for Task 6, with a significant drop for sizes below that. Users with medium acuity exhibited a significant drop on D_5, compared to the other two displays, for a size of 1.0 mm to a rate of only 88.1%, with perfect recognition or only a singular lapse being recorded on the other displays for that size. Users with low acuity achieved a 100% recognition rate on D_4 down to a line width of 1.0 mm, with only a singular lapse on D_3 and two incorrect responses (out of 15 overall) on D_4 for that size.

| | $L_{98,goodVA}$ | L_{95} | | |
		highVA	mediumVA	lowVA
D_5	1.5 mm	0.8 mm	1.5 mm	1.5 mm
D_3	1.5 mm	0.8 mm	1.0 mm	(1.0 mm)*
D_4	1.0 mm	0.8 mm	1.0 mm	1.0 mm

Table 7.5 Limits L_{98} (98% recognition accuracy) and L_{95} (95% recognition accuracy) for Task 5. *) A singular lapse at this level resulted in an effective accuracy value of 93.3% due to the small number of trials.

The intervention of a wider spacing of hachure lines did not significantly impact the results at sizes of 1.0 mm or larger, for which all participants achieved recognition rates of above 90%. The only significant difference detected was for a line width of 0.8 mm on display D_5, which was correctly identified in 69.0% of all trials for Task 5 (narrower hachure spacing), and 84.5% for Task 6 (wider spacing). No significant differences were detected between results of Task 5 and Task 6 for the other two displays at this size.

7.3.4. Task 7: Count parallel lines

Task 7 presented participants with the challenge to count parallel lines. Two parameters of the stimulus were controlled for this task: the width of each line and the gap between lines. Therefore, analysis is done by grouping results by display and gap width value (0.25, 0.5 or 0.75 mm), and analysing results within each group and group differences at equal gap with across displays, and for the same display at different gap width values. Results for participants with good visual acuity are shown in Figure 7.17.

Success rates for this task were at around 90% for a gap width of 0.75 mm, with no significant differences detected between different line widths on any display for this gap size. Average performance for this gap width was highest on D_5, but not significantly different from other displays. Performance for a gap width of 0.5 mm was significantly lower only on D_5, and significantly lower for all displays for a further reduction of gap width to 0.25 mm. Within each group, a significant drop of performance between line widths were detected in two cases for a reduction from 0.08 mm to 0.05 mm, and one case for a further reduction to 0.03 mm. Remarkably, for a gap width of 0.25 mm on D_5, a reduction of line width from 0.05 to 0.03 mm resulted in a significant improvement of performance, although at low levels from 43.5% to 67.4%.

The overall trend is reflected in the results of participants with high and medium visual acuity, neither of which showed a significant difference of performance on different displays.

Participants with low acuity performed significantly better at the lowest-resolution display D_5 (90%) than on D_4 (72.5%) at the largest gap size of 0.75 mm ($p=0.042$), and better on D_4 (80%) than D_5 (52.5%) at the medium gap size of 0.5 mm ($p=0.009$).
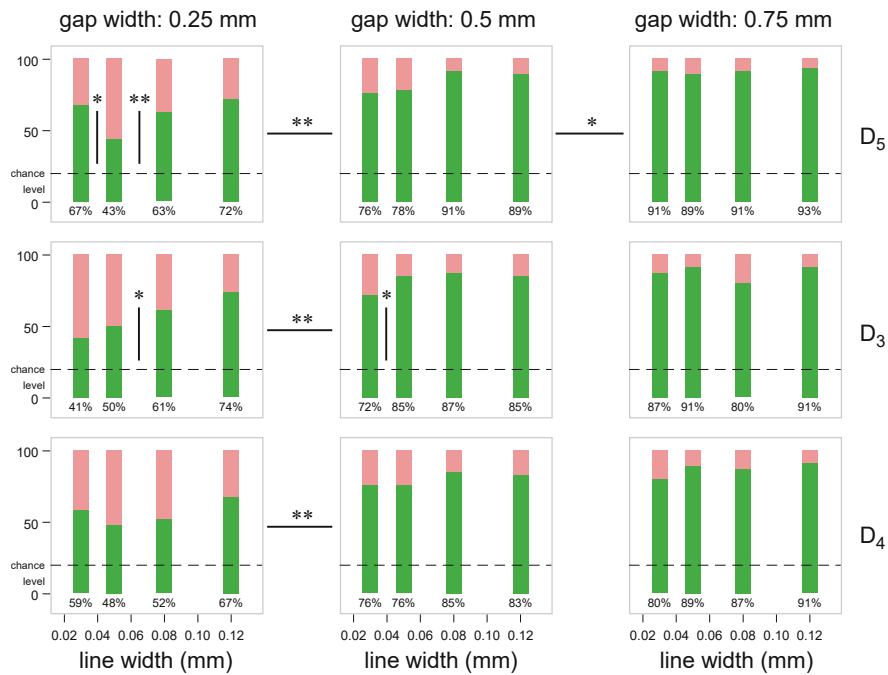

Figure 7.17 Recognition rates for Task 7, grouped by width gap width, participants with good visual acuity (logMAR < -0.05). Vertical black lines indicate a significant difference in performance between line widths within a group (notice the significant *increase* for a line width of 0.03 mm for the smallest gap size on D_5), horizontal black lines indicate significant difference in performance between groups across width factors. No significant differences between groups or individual results have been detected between displays. (*) *p < 0.05* (**) *p < 0.01*

7.3.5. Task 8: Count lines on pseudo-map

Task 8 asked users to count a specific type of line on a pseudo-map. After 10 participants the set of stimulus sizes was adjusted, as a drop in performance became apparent already at the second largest line width of 0.50 mm. An additional level of 0.60 mm was thus introduced. Unfortunately, this affects analysis of results, due to the smaller sample size and therefore reduced significance of results at this level. The results for participant #5 showed an unusually high rate of incorrect responses even at the largest sizes on all stations, so it may be assumed that the participant did not understand the task correctly. The results for this participant were disregarded for the analysis.

Figure 7.18 shows the results for the user group with good visual acuity for this task. Notably, performance for the initial size of 0.75 mm on display D_4 (92.0%) is already significantly lower than the combined performance at the two other displays (97.7%, $p=0.035$). While a further significant drop in performance on D_4 from this lower initial level is not detected until a line width below 0.5 mm, this again shows that the chosen approach for detecting an initial drop in performance may fail to deliver meaningful results if the starting size is chosen at a level at which performance is already diminished. Across all line widths, the rate of correct responses on D_4 is consistently lower than on the other displays (with the exception of equal performance as D_3 for a line width of 0.5 mm). On D_3, the first significant drop in performance

occurs below a line width of 0.6 mm. On D_5, the first significant drop is registered below 0.5 mm, making the lowest-resolution display the one with the best performance at relevant line widths.
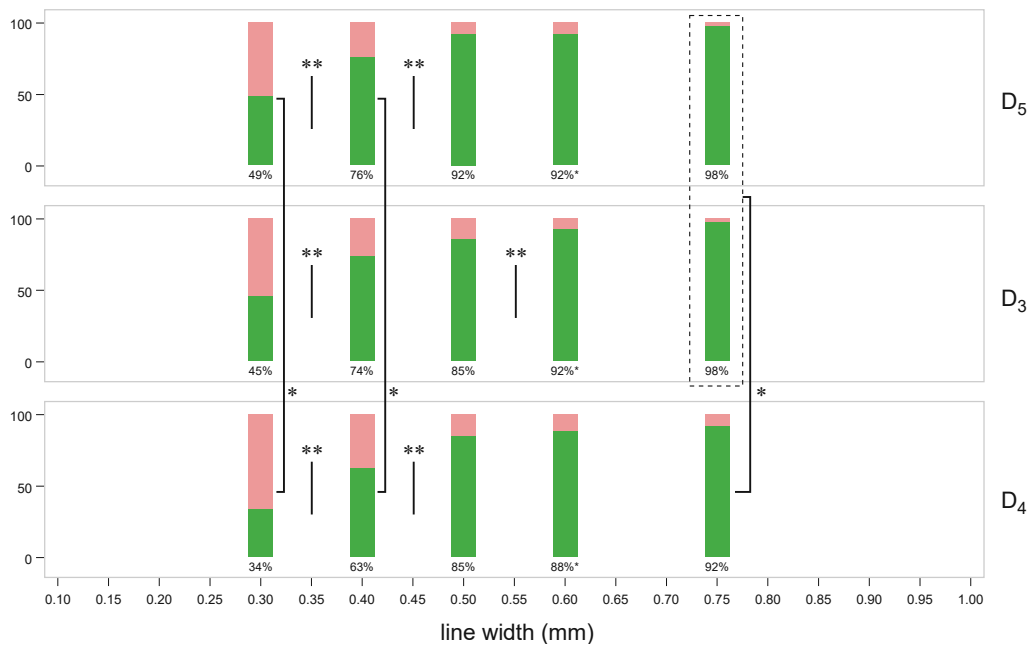

Figure 7.18 Proportions of correct answers for Task 8, participants with good visual acuity (logMAR < -0.05). Vertical black lines indicate a significant drop in performance between stimulus sizes, black brackets indicate a significant difference in performance between displays at the same size. Combined performance on D_5 and D_3 for the largest size of 0.75 mm is significantly higher than for D_4, (*) $p < 0.05$ (**) $p < 0.01$ (Stimulus size 0.6 mm was not presented to 9 of the 22 participants whose results are visualized here)

This overall trend is also apparent for participants grouped by visual acuity. Participants in the high acuity group show a performance above 90% at levels of 0.6 mm or larger, dropping to 88.9% on D_4 for a line width of 0.5 mm while performance remains above 90% (D_3: 91.7%; D_5: 97.2%) for the other two displays. The highest-resolution display D_4 is the worst-performing display for all sizes below 0.6 mm for this group. For participants with medium acuity, performance on D_5 does not reach 90% for any size, with the highest value of 88.5% for the largest line width of 0.75 mm, with a rate of correct responses of 96.2% for D_5 and 98.1% for D_3 at this level. For this group, performance on D_4 is worst among the three displays for 4 out of 5 levels in the range of 0.3-0.75 mm. For participants with low visual acuity, performance on D_4 is at 75% for a line width of 0.75 mm, improves to 93.8% for a line width of 0.6 mm, and drops to 50% or lower for smaller line widths. Table 7.6 compiles the L_{90} values for combinations of participant groups and displays for which a success rate of 90% has been reached.

	L_{90}			
	goodVA	*highVA*	*mediumVA*	*lowVA*
D_5	0.50 mm	0.50 mm	0.60 mm	0.75 mm
D_3	0.60 mm	0.50 mm	0.60 mm	0.75 mm
D_4	0.75 mm	0.60 mm	— *	— *

Table 7.6 Limits L_{90} (90% accuracy) for Task 8. *) Performance was below 90% for the starting width of 0.75 mm.

7.4. Discussion

The caveat discussed in Section 6.5 in the context of the second study also applies for some of the results of this study, namely that the chosen method of identifying initial drops in performance fails to deliver unambiguous results if the presented size levels were not large enough to reliably establish a "baseline" performance at larger levels. For this study, particularly Task 8 (counting lines on a pseudo-map) was affected by this limitation, for which a significant difference in performance between stations has been detected already for the largest size level. For other tasks, the establishment of a baseline performance level against which to quantify drops in performance has largely been successful.

The results for Task 1 show an initial drop at 0.4 mm for D_5 and D_3, delayed to one level smaller at 0.3 mm for D_4. This *contradicts* hypotheses HS3.1 (advantage of D_3 over D_5) and HS3.2 (no advantage of D_4 over D_3). However, no significant differences in performance at identical size levels on different displays have been detected, so the evidence in this case rests on a non-significant performance difference of 95% versus 99%. Hypothesis HS3.3 (no advantage of lowest-resolution display D_5) is supported by the results, which are very similar between D_5 and D_3. A minimum size for the "line with embedded arrows" symbol of Task 1 of 0.5 mm can be confidently asserted, and correlates with recognition rates above 95% even for the small group of participants with low visual acuity.

The intervention of placing green dots with a diameter of 1.1 pixels on the corners of the arrows for Task 2 resulted in an initial drop at lower sizes for D_5 and D_3. At the two smallest size levels of 0.3 mm and 0.2 mm, performance on the lowest-resolution display D_5 is significantly higher than on both other displays. For the size level of 0.3 mm performance on D_5 is still at 95%, indicating a relevant potential of the proposed technique to selectively increase performance on a low-resolution device and thus equalize the performance across display resolutions. The results strongly support hypothesis HS3.4 which assumed that the effect of the intervention would be most noticeable on the lowest-resolution display.

Interestingly, the results for Task 3, for which the green dots were aligned with the pixel grid to light precisely one green subpixel closest to each corner, did not produce such an unambiguous improvement. While the initial drop is shifted to one lower level as compared to Task 1 for displays D_5 and D_3, overall performance at small sizes on these displays is very similar to the results of the reference task. For the highest-resolution display D_4, the intervention was detrimental to performance as compared to Task 1, shifting the initial drop

one level up to 0.4 mm. The overall results *contradict* hypothesis HS3.5 to some extent, since the intervention proved detrimental to performance at the highest resolution. An interesting detail is the performance of participants with high visual acuity, shown in Figure 7.13, which accomplished recognition rates of 97% at a line width of 0.3 mm on the highest-resolution display for Task 3 (compared to 81% without the intervention in Task 1), which may indicate potential for this method in specific situations where high-resolution displays and high visual acuity of map users may be reliably assumed.

Task 4, for which users were asked to discriminate lines of different width, showed little difference between displays, supporting hypothesis HS3.6. Only for a single combination of base line width and width factor (0.2 mm / 1.25), a significant advantage of D_3 over the lowest-resolution display D_5 was shown. Hypotheses HS3.1 (advantage of D_3 over D_5), HS3.2 (no advantage of D_4 over D_3) and HS3.3 (no advantage of D_5 over the higher-resolution displays) are all supported by the results of this task. Generally, a width factor of 1.5 has shown to produce significantly higher recognition rates, and a minimum line width of 0.2 mm (corresponding to ~2 physical pixels on the lowest-resolution display D_5, with an increase of at least 1 pixel for each increase in width) seems advisable for reaching recognition rates of 90% or higher, with no consistent improvement for the much larger base line width of 0.5 mm.

Tasks 5 and 6 asked participants to identify lines with internal hachures, which participants of all acuity groups completed with high accuracy down to an overall line width of 1.5 mm. At the lower sizes of 1.0 mm and 0.8 mm, performance on the highest-resolution display D_4 was consistently higher than on the two other displays, with a significant difference identified compared to D_5 for the line width of 0.8 mm. Also, for D_4, a perfect recognition rate of 100% was accomplished by participants with good visual acuity down to a line width of 1.0 mm, and even by participants with low visual acuity for Task 6. This supports hypothesis HS3.7 for which an advantage of D_4 was assumed for this task. However, the advantage is not significant for the hachures with larger gaps of Task 6, where performance of D_4 and D_3 are at similar levels. The results of both tasks also do not contradict HS3.1 and HS3.3. It is unfortunate that a size level of 1.25 mm overall line width was not included in the selection of sizes, as this would have potentially allowed for a finer-grained estimate of the threshold. With the results as they are, a general minimum width of 1.5 mm has to be assumed for lines with fine-grained internal structure like the ones used in tasks 5 and 6, with good indication that if an ultra-high-resolution display like D_4 is available, this may be confidently reduced to considerably smaller size of 1.0 mm.

Task 7 asked participants to count parallel lines. At the largest gap width of 0.75 mm performance rates between 80% and 93% were observed, with the variation not clearly correlating with display resolution or line width. A significant drop in performance was observed at all three displays for the smallest gap with of 0.25 mm. The results support HS3.2 (no advantage of D_4 over D_3) and HS3.3 (no advantage of D_5 over higher-resolution displays). The task of counting parallel lines on a map seems much less relevant today than it was in the age of printed maps, where the counting of bundles of contour lines was an important technique to assess the steepness of terrain – nowadays, users would probably expect explicit interaction modalities to be provided for such assessment. Nevertheless, there may be use

cases for which the visual identification of individual lines in a bundle is of relevance. For such cases, a gap width of 0.5 mm in combination with a line width of 0.08 mm or wider, or a gap width of 0.75 mm in combination with finer lines down to 0.025 mm can be derived as a minimum dimension for acceptable accuracy from the results.

For Task 8, which asked participants to count lines of different types on a pseudo-map, performance on the highest-resolution display D_4 is significantly lower than on the other two displays combined already for the largest size presented, and is consistently lower than on both other displays at all size levels. This strongly *contradicts* hypothesis HS3.3, which assumed that the lowest-resolution displays would not facilitate higher performance than a higher-resolution one, and also *contradicts* hypothesis HS3.9 which assumed that for this task, which involves to some extent peripheral vision, display resolution would have negligible influence. The difference in performance at the initial size level also means that a reliable baseline performance level could not be established across the three displays for this task. However, the 98% accomplished on D_5 and D_3 appear to be a plausible baseline performance level. For any authoritative conclusions, the experiment would have to be repeated in a setting that involves stimuli at larger sizes for this task, and, perhaps, better instructions for participants, in order to completely rule out inflated incorrect responses due to a misunderstanding of the instructions.

In Section 7.2.2 the attempt was made to derive hypothetical minimum dimensions from the simple sampling models of one-arcminute of visual acuity and one CSS-pixel of resolution. For Task 1 it was assumed that the width of the arrows embedded in the line needs to be 3 units wide, which resulted in a predicted minimum width of 0.45 mm for the one-arcminute model and 0.55 mm for the CSS-pixel model. Results for Task 1 align roughly with the prediction of the one-arcminute model in this case, with significant drops in performance detected for a line width below 0.5 mm on the two lower-resolution displays. For Task 4, both models predicted that for a base line width of 0.5 mm, successful discrimination should be possible for any of the three width factors, as the width difference for the smallest factor of 1.25 results in an absolute difference of line width that is larger than the sampling unit. The results show, however, consistently lower performance for width factors below 1.5. This demonstrates that, as has been well established in Section 3.1, the human visual system is not a "sampling and measuring" device that operates uniformly at a certain resolution, and that the judgement of the width of lines is not limited only by resolution. For base line width of 0.2 mm, the one-arcminute model correctly predicted diminished performance at smaller width factors, and for the base line width of 0.1 mm diminished performance (albeit at a non-significant difference) at even the largest width factor. For the internal hachure lines, again an assumption was made that each "lane" of hachures needs to be at minimum 3 units wide in order to be clearly legible. This predicted a minimum overall line width of 1.0 mm for the one-arcminute model, and 1.25 mm for the CSS-pixel model. In fact, performance was slightly reduced at 1.0 mm (unfortunately the stimulus was not presented at a width of 1.25 mm), dropping off significantly on all displays for smaller sizes. So again, the one-arcminute model, combined with the assumption of a minimum width of 3 units for graphical detail embedded within a line, made an approximate prediction.

8. Guidelines for producers of digital maps derived from the presented studies

The goal of this chapter is to summarize and generalize the findings of the presented studies in a form that makes them applicable in the practical context of designing and producing maps. Any generalisation of empirical findings into more universal, actionable guidelines carries the risk that the guidelines may turn out to not be valid outside of the narrow set of circumstances of the empirical investigation – the exact stimuli, participant population, tasks, and viewing conditions present for a study. Such generalisation is therefore outside of the domain of strictly empirical research, and subject to ethical considerations, represented by questions like: Are the derived guidelines plausibly related to the empirical findings? What is the risk associated with the incorrect or incomplete transmission of information? Does a map design resulting from revised guidelines empower the map user and enrich their perception of a spatial situation, or does it frustrate users or exclude certain demographic segments from access to important information?

Some ethical principles that may be related to the question of the legibility of map symbols have been proposed by various authors in the field of cartography and related fields (Dent et al., 2008, p. 19f; British Cartographic Society, 2020; American Geographical Society, 2021; K. Field, 2022; World Wide Web Consortium, 2022). Among those, the following principles for ethical conduct in cartography can be found:

- Show all relevant data whenever possible
- At a given scale, strive for an accurate portrayal of the data
- Realize opportunities
- Strive to know your audience (the map reader)
- Understand impacts
- Do no harm
- The web is for all people
- Protect the vulnerable

Some of these principles can be read as supporting the production of maps with smaller symbology and higher information density, even at the risk of excluding map users with lower visual capabilities, while others could be read as advocating for a more cautious approach that mandates to always use symbology that can be reliably read by a large proportion of the demographics and in potentially adverse circumstances. The author of this thesis proposes an additional ethical value for the purpose of the formulation of map design guidelines, which is not explicitly addressed by the principles above: to increase the space of possible map design solutions that can be considered by a map designer, while not relieving them of the ultimate responsibility (and agency) of making decisions that result in a map design that works for the intended specific circumstances of map use. Applying such a principle also makes the proposal of guidelines subject to the historical context in which they are made: as has been discussed in Section 3.3, presently the common recommendation that screen-based maps should use much coarser symbology than printed maps is still largely undisputed by the cartographic literature, and the coexistence of display media of various

resolutions is considered a "big mess" (Muehlenhaus, 2014). Proposing simplified guidelines for minimum dimensions for modern screens, rooted in empirical investigation, would therefore have the potential to significantly increase the number of design alternatives that can be considered by map designers, even if the full nuance of all aspects of the empirical findings cannot be transported in such communication, and even if there remains a risk that not all users may reliably read all of the information under all circumstances. The provision of tools that can be used by map designers to empirically verify the legibility of a particular map design would be a service that could help dispel any remaining doubts regarding the reliability of the cartographic communication for any concrete map use scenario.

Another argument that can be made in favour of guidelines that propose symbology sizes at the limit of what can be reliably read by map users with good visual acuity is that a digital map today is rarely a static product with fixed dimensions. If a symbol is too small to read without ambiguity, the information providing reliable identification can be provided with a single click or gesture; on mobile phones and tablet devices, the two-finger pinch gesture is widely established for the purpose of zooming and enlarging graphical content; map applications can offer user preferences in which the desired symbology size can be configured by the user to their needs; and, as has been discussed in Section 3.2.2, modern operating systems allow for the system-wide configuration of the relationship between measurements specified by software and physical pixels on the screen. All of these possibilities can be used to adapt the presentation of a map to the user's needs, and access additional details on demand. This relieves the map designer of the responsibility to produce a static map image that reliably conveys all layers of information to all possible users under all conceivable circumstances.

In the light of the aspects discussed above, this section will proceed with proposing guidelines for map designers, rooted in the empirical investigations described in chapters 5 to 7, at three levels of generality: Section 8.1 will present a number of simplified general guidelines for map and symbology design that do not require detailed assumptions of the map user's visual capabilities or display device; Section 8.2 will attempt to provide more detailed guidelines differentiated by display resolution, user capabilities and symbology design; Section 8.3 will provide guidelines for designers who intend to empirically test their designs or design alternatives with real users. By structuring the recommendations in this way, it is hoped that the findings of this research project are presented in a way that is found useful in various application contexts, without overloading practitioners with too many details and caveats. For each guideline derived from the empirical results presented earlier in this thesis, these results will be referred to in footnotes for documentation and further investigation by interested parties.

8.1. General guidelines and minimum dimensions for cartographic symbology for modern digital displays

The guidelines in this section are intended to be valid for a target demographics with good effective visual acuity (corrected to "20/20 vision" / logMAR 0.0 or better), for smartphone screens with a resolution above 260 ppi (which, as of 2022, encompasses all commercially available smartphones) and desktop screens with a resolution above 130 ppi, and for designs that utilize high contrast (close to black on white). The empirical investigations underpinning the guidelines do not allow to draw conclusions for desktop monitors of conventional resolution, generally assumed to be at around 90-100 ppi – for maps targeting these devices, map designers should refer to the established guidelines discussed in Section 3.3.3.

The following general guidelines can be formulated based on the empirical studies presented in this thesis and the considerations discussed above:

- Cartographers would generally be right to **advocate for the use of high-resolution displays**, as these have shown to be able to provide significant benefits[49] for the legibility of cartographic symbology [50]. However, it has to be noted that for some particular combinations of symbology designs and display pixel densities, legibility on a higher resolution device may actually be slightly *worse* for some users than on a lower resolution display[51] – this is hypothesized to be caused by the "enlargement effect" that larger pixels can have for some fine details when rendered with activated anti-aliasing, which is enabled by default for all modern graphics systems. Generally, the advantages of higher resolutions seem to be dominant over any potential disadvantages, and of general practical relevance. Even an ultra-high resolution (~800 ppi for a phone screen, corresponding to ~400 ppi for a desktop screen) has been shown to provide an advantage over high-resolution displays (~500 ppi mobile / 250 ppi desktop) for some symbology designs[52]. Highly detailed maps utilizing multiple levels of a visual hierarchy of map symbology may therefore even be a "killer app" for justifying the purchase of ultra-high-resolution displays, to be potentially used in settings in which map users can be expected to reliably have exceptional visual acuity (such as, for example, emergency services, pilots, or the military). Although desktop monitors have not been directly tested in the studies presented in this thesis, a translation of the results of the lowest-resolution display used in the first study to desktop viewing distances and -resolutions suggest that replacing monitors of conventional pixel density with higher-resolution ones should have a noticeable effect on the legibility of cartographic symbology.

- However, the **advantage to be gained by high- and ultra-high resolution devices is not a dramatic one** as compared to the current low-end of available devices at ~260 ppi

---

[49] As measured by an initial drop of recognition rate at a smaller size and/or a significant difference in recognition rates for mobile phone displays of varying pixel densities.

[50] Tasks 1, 2, 3, 4, 5 of the first study; tasks 1, 3, 9, 10 of the second study; tasks 1, 5, 6 of the third study.

[51] Tasks 1, 5, 6 of the first study; task 10 of the second study; task 8 of the third study. Detailed analysis of the results of task 3 of the third study (see Figure 7.13) also shows that the detrimental effect of a high-resolution device may be possibly overcome by users with very good visual acuity.

[52] Tasks 1, 6 of the first study; tasks 1, 3 of the second study; tasks 1, 5 of the third study.

(corresponding to ~130 ppi for desktop screens) – for many types of symbology, an improvement in recognition rate of a few percentage points, or a slightly smaller size of symbology that can be reliably identified (reduced by a factor of ~1.2–1.5) can be expected for map users with good or excellent visual acuity[53].

- Even the smartphone at the low end of the spectrum of resolutions which are commercially available today (the LG K50S, at a resolution of 265 ppi) has shown to be able to reproduce cartographic symbology that can be reliably discriminated from similar graphical stimuli at **much smaller sizes than has been conventionally recommended for screen-based symbology**. For the sets of graphically similar icons identified for the second study[54], the minimum size at which an icon could be correctly identified among similar icons with a success rate of 98% or above was between 0.7 mm and 1.5 mm[55]. **A minimum size of 0.7 mm may therefore be realistically assumed for a smartphone viewing situation for monochrome icons that are well differentiated** graphically as well as semantically, and are reproduced near maximum contrast – a size that is consistent with some recommendations published for printed maps (e.g. it approximates the 0.6 mm recommended for filled geometric figures by Imhof, 1972; however, Imhof recommended a significantly larger size of 1.5 mm for graphical icons). This corresponds with **a minimum size of 1.4 mm for point icons on desktop screens with a resolution of 130 ppi** and above.

- The simple **model of an apparent size of one minute of arc as the minimum size of graphical detail has been shown to seriously underestimate the capabilities of map users** with good visual acuity with respect to the discrimination of monochrome map icons at maximum contrast[56]. However, **for internal detail of lines (triangular arrows, one "track" of internal hachures), a model of three minutes of arc (~0.3 mm at 30 cm viewing distance) as minimum width** for which the orientation of such elements can reliably be detected (corresponding to a minimum width of ~0.55 mm at 60 cm viewing distance) has been shown to plausibly match the empirical results[57].

- **The model of one (CSS-) pixel as minimum size of graphical detail has also been shown to seriously underestimate the capabilities of modern smartphone displays**, even for the lowest resolutions commercially available today[58]. But, since the concepts of the CSS pixel or "density independent pixel" (see Section 3.2.2) or similar units are widely available in graphics frameworks today and provide a practical and reliable way to specify apparent size independent of display resolution or viewing distance, these units of measurement – corresponding to a metric size of ~0.125 mm at smartphone viewing distance or ~0.22 mm for a desktop situation[59] – should be of high relevance for the specification of dimensions

---

[53] Based on the size levels for which a recognition rate of 98% or higher was accomplished by participants with good visual acuity for tasks 1, 2, 9, 10 of the second study and 1, 5, 6 of the third study.

[54] See Figure 6.5 on page 127.

[55] Tasks 1–4 of the second study. The stated success rate was slightly missed (at 94%) for the OSM-castles set for the largest size of 1.5 mm on the lowest-resolution display at 265 ppi.

[56] Tasks 1–4 of the second study. Also see sections 6.3.2 and 6.5.

[57] Tasks 1, 5, 6 of the third study. Also see sections 7.2.2 and 7.4.

[58] See sections 6.3.2 and 6.5.

[59] One CSS pixel is defined as having a size of 0.26 mm at a distance of 28 inches (≈71 cm), which corresponds with a size of ~0.125 mm at a distance of 34 cm and a size of 0.22 mm at 60 cm.

for screen-based cartographic symbology. **For graphical icons, a minimum size of 0.5–0.75 CSS pixels for the extent of the smallest details differentiating the icons may be considered by map designers**[60]**, corresponding with an effective minimum size of 6 CSS pixels for well-differentiated point icons. For internal details of line symbology, like triangular arrows or hachures, a minimum width of 2.5 CSS pixels** corresponds to the guideline of 3 minutes of arc stated above[61].

- For map icons and finely detailed shapes, **horizontal or vertical graphical elements separated by a gap width close to one physical pixel can potentially result in aliasing artefacts detrimental to legibility**, depending on the precise alignment with the pixel grid[62]. When producing a design for a particular known display resolution where the reproduction of such fine details in alignment with the pixel grid is crucial, map designers could consider shifting the geometry by fractions of a pixel for a more favourable alignment of graphical details and a reduction of aliasing artefacts. However, this will be hard to implement in practice, as each particular physical resolution the map may be presented on needs to be taken into consideration.

- Otherwise **attempting to engage with physical pixels**, either as a minimum manipulable graphical unit, or as a unit of measurement for graphical elements, **has shown to provide no practical benefits** over the specification of graphical elements using dimensions with an arbitrary relationship to the physical pixel grid[63]. **The practice of "grid-fitting" icons manually for ideal alignment with physical pixels and optimal contrast has been shown to largely have no significant effect on legibility.** While there may be a few very specific combinations of icon size and display resolution for which a manually grid-fitted icon provides a slight benefit in legibility over vector-based geometry at a corresponding scaling factor[64], such edge cases would have to be identified on a case-by-case basis, and recognition rates even for the optimized icon would probably be below acceptable levels due to the small size of pixels of modern high-resolution displays. Given the considerate manual work involved in adapting an icon design to a series of representations at discrete pixel sizes, it is not clear whether a real-world use case exists that would justify the investment of time into such optimization. The physical pixel as a model of smallest graphical detail has also been shown to not yield useful estimates of graphical fidelity, as it overestimates the required size of graphical detail for lower-, and underestimates the size for higher-resolution displays.

- **Modifying the geometry of shapes to amplify the appearance of small details or the difference from other shapes has been shown to have the potential** to overcome the

---

[60] See sections 6.3.2 and 6.5.

[61] Tasks 1, 5, 6 of the third study. Also see sections 7.2.2 and 7.4.

[62] As discussed in section 5.3.2.2 for the "Tumbling E's" task (task 1) of the first study, and also hypothesized to potentially cause the worse performance of the icons of the "Maki-rectangular" set as compared to the "Maki-triangular" set in the second study.

[63] As observed for the interventions of manual grid-fitting of icons (tasks 6 & 7 of study 2), thresholding for maximum contrast (task 5 of study 2) or highlighting of green subpixels at the corners of arrows (task 3, study 3).

[64] For the Maki-rectangular set, one single such case has been found for a size of 13 physical pixels / 0.6 mm on a display resolution of 522 ppi for task 6 of the second study. Recognition rates were improved to 93% for the grid-fitted icon as compared to 81% for the scaled vector graphics with randomized placement on the pixel grid. For the Maki-triangular set, no significant advantage of grid-fitted icons has been found for any pixel size.

disadvantages of lower-resolution displays and improve legibility at very small sizes. Although further research is needed to fully explore the experimental methods of "shape difference amplification" and highlighting of arrows embedded in lines that were tested[65], in order to potentially develop automated methods for such interventions, map designers are encouraged to experiment with similar techniques in situations where the legibility of very small symbology needs to be improved.

8.2. Minimum dimensions differentiated by symbol type, display resolution and visual acuity of map users

This section presents a table of minimum dimensions derived from the empirical investigations. Strictly speaking, the stated dimensions are valid only for the particular stimuli and circumstances used in the studies. However, it is hoped that the chosen graphical elements may also represent map symbology of a similar or better differentiated design with respect to discriminability of individual symbol types. Table 8.1 compiles the minimum dimensions derived from the studies for a smartphone viewing scenario, three different resolution bands and two classes of visual acuity.

While a desktop scenario has not been directly tested in the studies presented in this thesis, the minimum dimensions presented in Table 8.1 can be adapted to such a viewing situation by multiplying the metric dimensions with a factor of 2 (corresponding to the typical assumption of a viewing distance increased from 30 cm to 60 cm), while leaving the CSS pixel values unchanged (since the definition of the CSS pixel already takes viewing distance into account – see Section 3.2.2). Also, the resolution thresholds need to be adjusted by dividing the values in the top table row by 2, such that the "medium resolution" column applies to desktop monitors with a pixel density above 130 ppi, and the "high resolution" column applies to a pixel density above 250 ppi.

---

[65] Task 8 of the first study, tasks 2 and 3 of the second study.
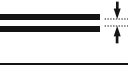
	Medium Resolution (>260 ppi)		High Resolution (> 500 ppi)		Ultra-High Resolution (~800 ppi)	
	Visual Acuity		**Visual Acuity**		**Visual Acuity**	
	Good[66]	Excellent[67]	Good	Excellent	Good	Excellent
Icons "Maki-triangular" (or better differentiated)[68]	**0.7 mm**	0.7 mm	**0.7 mm**	0.6 mm	0.7 mm	0.5 mm
	6 px	6 px	**6 px**	5 px	6 px	4 px
Icons "Maki-rectangular" (or better differentiated)[69]	**1.0 mm**	0.7 mm	**0.85 mm**	0.7 mm	0.7 mm	0.5 mm
	8 px	6 px	**7 px**	6 px	6 px	4 px
Icons "NPS-vertical" (or better differentiated)[70]	**1.5 mm**	1.25 mm	**1.5 mm**	1.25 mm	1.25 mm	1.25 mm
	12 px	10 px	**12 px**	10 px	10 px	10 px
Line width (to discriminate dash patterns)[71]	**0.15 mm**	0.12 mm	**0.12 mm**	0.11 mm	0.12 mm	0.11 mm
	1.2 px	1.0 px	**1.0 px**	0.9 px	1.0 px	0.9 px
Line width (to discriminate line widths)[72]	**0.2 mm**	0.2 mm	**0.2 mm**	0.2 mm	0.2 mm	0.2 mm
	1.6 px	1.6 px	**1.6 px**	1.6 px	1.6 px	1.6 px
Line separation (for parallel lines)[73]	**0.15 mm**	0.15 mm	**0.15 mm**	0.12 mm	0.15 mm	0.12 mm
	1.2 px	1.2 px	**1.2 px**	1.0 px	1.2 px	1.0 px
Line with internal arrows, width (internal arrows 2/3rd of total width)[74]	**0.5 mm**	0.5 mm	**0.5 mm**	0.4 mm	0.4 mm	0.4 mm
	4.0 px	4.0 px	**4.0 px**	3.2 px	3.2 px	3.2 px
Text labels, capital letter height (font "Roboto")[75]	**1.1 mm**	1.0 mm	**1.0 mm**	1.0 mm	1.0 mm	1.0 mm
	8.8 px	8.0 px	**8.0 px**	8.0 px	8.0 px	8.0 px

Table 8.1 Minimum dimensions established by empirical verification in studies 1-3, differentiated by display resolution and visual acuity of map users, for a smartphone-based viewing scenario. Pixel values are given in CSS pixels, which are independent of physical display resolution. Results are valid for scenarios in which map users examine symbols individually, near maximum contrast. Values in bold indicate guidelines of most practical relevance. To adapt the dimensions in the table for a desktop viewing scenario, metric dimensions need to be doubled, ppi values for screen resolution halved, and CSS pixel dimensions will remain unchanged. Pixel dimensions for rows 1-3 have been rounded up to full pixel values, with only values with a fractional part < 0.1 rounded down. Pixel dimensions for rows 4-7 have been rounded to multiples of 0.1 pixels.

8.3. Guidelines for empirical examination of minimum dimensions

The static guidelines proposed in the previous sections may not always be applicable to new map designs, which may incorporate symbology from any number of symbol libraries or employ truly novel design ideas. Instead of attempting to relate the symbology of a novel

---

[66] Map users with a logMAR score of -0.05 or lower, roughly corresponding to "20/20" vision.

[67] Map users with a logMAR score of -0.25 or lower.

[68] Based on task 2 of the second study.

[69] Based on task 1 of the second study.

[70] Based on task 3 of the second study.

[71] Based on task 3 of the first study.

[72] The stimuli used for task 4 of the third study included base line widths of 0.1 mm and 0.2 mm, with worse performance across displays for the narrower base line width. More research would be needed for a more fine-grained assessment of the perceptual threshold in between those line widths (e.g., at a line width of exactly 1 CSS pixel ≈ 0.125 mm).

[73] Based on the results of task 8 of the third study and task 2 of the first study.

[74] Based on task 1 of the third study.

[75] Based on task 6 of the first study.

design to static guidelines, map designers should be encouraged to run their own empirical investigations to verify the legibility of a set of symbols under the specific intended circumstances. While the development of a ready-to-use software solution for the simple testing of symbology in professional contexts is beyond the scope of the work undertaken for this thesis, nevertheless a few guidelines for map designers wishing to verify their designs can be distilled from the experience with designing the studies presented in this thesis, in order to encourage and facilitate such investigation by map design practitioners:

- Any empirical investigation must be clear about the **intended demographics of map users, and employ a recruiting strategy** to ensure sufficient participation of the target demographics in the empirical investigation. The studies presented in this thesis have shown that for a target demographics of young, healthy adults with good visual acuity, recruitment of participants among students at a university may be a viable strategy. For a different target demographics, other recruitment strategies may be needed in order to ensure participation of a representative sample of the population.

- While mobile phone displays are commercially available today in a wide variety of resolutions, for an initial investigation it could be sufficient to test at the lower end of the pixel density spectrum (~260 ppi) and the higher end of mainstream phones (~520 ppi). Phones with ultra-high resolutions are not manufactured any more as of 2022, and currently have a negligible market share. This means that **a practical study may need to incorporate only two phones**, one of which will be an affordable entry-level phone. If symbology should be tested for desktop use, the two phones should be **complemented by a conventional desktop LCD monitor** with a pixel density of 90-100 ppi. Conclusions for higher-resolution desktop monitors can potentially be derived by adjusting results from mobile phones in proportion to the increased viewing distance. Using this strategy, a wide range of conditions can be tested with only three commonly available and affordable devices.

- While the studies presented in this thesis sought to standardize viewing conditions for the purpose of making results comparable with potential future studies, informal investigations in the context of practical map design may not need to create such standardized conditions. The phones can be handed to participants in an indoor environment for a "casual" yet reasonably controlled viewing situation, resembling more closely an actual map use setting. Responses can be collected verbally or directly on the device. Graphics can be rendered as bitmap images and displayed using the phones "image gallery" app. Thus, many **aspects of the experiment setup and apparatus described in Chapter 4 may be simplified** in the context of practical experimentation outside of the more rigorous requirements of academic research.

- For establishing minimum dimensions, the designer of the experiment needs to make sure that **symbology is presented at large enough sizes to reliably establish a baseline performance,** if this is not known or assumed. In any progression towards smaller sizes, the largest two levels should not show a drop in performance, else it remains unclear if a valid baseline performance has indeed been recorded for the largest size.

- **Symbology should be presented at size levels that can be meaningfully converted to dimensions of practical relevance, in appropriate units**. Examples for such levels would be multiples of 0.1mm or 0.1 CSS pixels. Size levels should progress through a sequence

of such predefined sizes, and not be adjusted purely mathematically (e.g., by reducing by a constant factor at each step). This will ensure that the resulting minimum size is meaningful and easy to remember, and not an arbitrary floating-point number.

- **Subsequent size levels to be tested should differ by about 10-25%.** A larger difference may fail to identify the threshold at which performance drops with sufficient precision.

9. Conclusions and future work

"Important questions remain."
(Westheimer, 1981, p. 3)

The work presented in this thesis has started out with a simple question: How small can we make symbols on a map so that they can still be reliably read? The investigation of this question has resulted in three empirical studies, some theoretical considerations, and has arguably even allowed room for creativity in the design of the experimental apparatus and software, the selection of symbology for use in the experiments, the interventions devised for potentially improving legibility of cartographic elements and, finally, the translation of empirical results into guidelines for map makers. In the opinion of this author, the whole project represents the appeal of most cartographic projects: the task may seem simple at first, and needs to be done according to the state of the art and science; but it will always involve subjective and creative decisions that make the work a personal one, and make a contribution to the overall "cartographic conversation", in which multiple approaches, styles, and schools of thought are always present.

Chapter 2 presented the overall research objectives (RO) and research questions (RQ) stated at the outset of the research project. RO 1, a compilation of existing guidelines from the cartographic literature, has been attempted in the tables presented in sections 3.3.2.1 and 3.3.3.3. RO 2, an empirical investigation of minimum dimensions for cartographic symbology, has been undertaken in three studies, presented in chapters 5, 6 and 7, based on the software, apparatus and procedures proposed in Chapter 4, which have been refined during the course of the studies (RO 2.1). Section 6.1.2 presented an approach for the identification of sets of similar map icons to use in the empirical investigation (RQ 2.2.1), and for each of the tasks of the three studies, the geometry of the used stimuli and the parameters of presentation have been documented (RO 2.2). Studies 2 and 3 did not only investigate the discriminability of graphical stimuli deemed relevant for the design of cartographic symbology, but also attempted to employ methods for increasing the legibility of maps symbols, with the interventions made for tasks 5–8 of the second study, and tasks 2 and 3 of the third study (RO 2.4). Finally, a set of practically applicable guidelines have been derived from the findings of the empirical investigations, presented in Chapter 8 (RO 3).

Research Objective RO 2.3 called for investigating the impact of display pixel density, with the main question being the impact of increased resolution for cartographic applications (RQ 2.3.1). Four general hypotheses were formulated in the context of this research question, which were partially translated into hypotheses for the individual studies presented in chapters 5, 6, and 7. Synthesizing the results of the three studies, the hypotheses can be commented on as follows:

Hypothesis HG2.3.1.1: An increase in resolution, from devices at the low end of the spectrum of available resolutions to devices with higher pixel density, results in improved recognition accuracy at small sizes for some cartographic symbology.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

This hypothesis can be **retained**. Particularly for the lowest-resolution device D_1 (228 ppi), used only in the first study, performance of participants with good visual acuity was significantly increased for 5 out of 6 tasks on the display of higher pixel density. For display D_5 (265 ppi), which marks the low end of commercially available devices today, performance was reduced, as measured by a "significant drop" at a larger size or a significantly worse performance at a specific size level, compared to other displays, for 3 out of 6 tasks that can be compared across displays for the second study, and for 4 out of 6 such tasks for the third study. However, the differences are less pronounced, and in many cases manifested by only a small difference in recognition rate or a shift of the significant drop by one level.

Hypothesis HG2.3.1.2: *An increase in resolution is never detrimental to the legibility of cartographic symbology at small sizes.*

This hypothesis has to be **rejected**. As apparent in the results of task 1 of the first study, and task 8 of the third study, stimuli exist for which performance can be reduced for displays of a higher resolution. Further research into the exact circumstances under which such phenomena occur would certainly be justified. Both of the tasks mentioned involve fine parallel lines, which indicates that interference with the pixel grid at lower resolutions could have been exploited by participants to successfully deduce the type of stimulus, despite the lower fidelity of reproduction. However, overall this effect can be considered less relevant than the advantage of a higher resolution, so a general recommendation to use devices of high pixel densities can be maintained.

Hypothesis HG2.3.1.3: *Participants with high visual acuity will benefit more from an increase in resolution for some cartographic symbology, showing in better recognition accuracy at small sizes, than participants of lower visual acuity.*

This hypothesis can be **retained**. One example has been shown for the results of Task 3 of the third study, for which participants with the highest visual acuity could reliably discriminate the arrows embedded in a line at a level for which the performance of participants with lower acuity already approached chance level (see Figure 7.13 on page 176). However, performance of participants with highest visual acuity has been consistently better than that of other participants at all display resolutions.

Hypothesis HG2.3.1.4: *Mobile devices of the highest pixel density (>800 ppi) have surpassed the capabilities of humans to resolve small details; legibility of small cartographic symbology will not be improved, compared to presentation on a display with a pixel density around 520 ppi.*

This hypothesis has to be **rejected**. The results of Task 1 of the first study, tasks 1 and 2 of the second study, and Task 1 of the third study show that for some stimuli, the performance on display D_4 with highest pixel density (801 ppi) was better than the performance on display D_3 (522 ppi), as measured by a significant difference at a relevant size level or an initial drop in performance at a smaller level. However, for some tasks, performance on the highest resolution display was lower than on other displays, as has been mentioned above. Overall, a weak recommendation to use the highest possible resolution can be made, as such displays potentially avoid detrimental aliasing artefacts that were observed for some stimuli. However, such a recommendation is currently of little practical relevance because, as of 2022,

devices with ultra-high pixel densities, matching the ~800 ppi of the highest-resolution device used in the presented studies, do not seem to be commercially available any longer.

The remaining sections of this final chapter attempt to summarize the contributions of this thesis to the overall academic cartographic conversation, and point out future threads of investigation that may be picked up by researchers and practitioners going forward.

9.1. A definition of "minimum dimension" of cartographic symbols

Any conversation needs a shared understanding of key terms and concepts. As the review of the literature in Section 3.3 has shown, minimum dimensions for cartographic symbols were proposed by many authors, but it often remained unclear what precisely was referred to, and how those guidelines came to be established. Based on the empirical investigations undertaken in this thesis, the following **definition for a cartographic minimum dimension** is now proposed:

> **For a given** *viewing situation* **and** *target demographics***, the smallest size out of a downwards** *progression of sizes***, starting at a clearly legible size, for which a symbol design can be discriminated from other,** *similar symbols* **without a** *significant drop* **in performance as compared to larger sizes.**

Any proposal for cartographic minimum dimensions should specify the underlying choices for the elements of the definition set in cursive: Which viewing situation was investigated, what were the target demographics, and how were participants recruited, which sizes have been tested, which symbol choices were offered as alternatives, and by which means was a significant drop in performance detected.

For this study, the elements of the proposed definition were chosen as follows: The viewing situation was a controlled indoor environment free of disturbance, in which map symbols for most tasks were presented in isolation at optimal contrast on a range of smartphone displays. The target demographics were young, healthy adults with good visual acuity, which were recruited mainly among students of cartography courses at a European university, and whose visual acuity was assessed without medical examination, using the experiment apparatus and software. The progression of sizes has been chosen based on pilot experiments performed by the author and a small group of colleagues. Both the size levels tested and the choice of symbols can be found in the detailed description of each of the tasks for each study. For identifying sets of similar point symbols, a simple automated raster-based analysis of icon similarity was applied as the basis for identifying clusters of similar icons out of real-world collections of map icons. Finally, significant drops in performance have been detected by statistical analysis with Fisher's exact test, whose results have been complemented by reporting symbol sizes for which the recognition rate remained above a high percentage value (typically 98% or 95%).

The elements of the proposed definition may also serve as an index to discuss a particular investigation's shortcomings, or potential future improvements. These questions will be

picked up in sections 9.3 and 9.4 below. Before that, the following section will briefly discuss some potential practical applications of the findings of the studies presented in this thesis.

9.2. Further practical implications and application ideas

An attempt to extract guidelines of practical relevance from the studies presented in this thesis has already been undertaken in Chapter 8. This section attempts to contextualize the relevance of the findings within current applied cartographic practices and challenges. As stated in the introduction, minimum dimensions can be seen as fundamental parameters which establish the lowest levels of the visual hierarchy of information on a map, as well as influence generalisation operations performed on the map content. As such, smaller minimum dimensions can potentially contribute on multiple levels to a higher visual information density and "delay the loss of information" (Spiess, 1990b). Cartographic applications requiring such a higher information density may not necessarily be found in scenarios which are already served by digital cartography to our full satisfaction, as could be argued is the case for simple monomodal route planning applications or single-layer thematic maps. Instead, one may look towards application areas for which fully satisfactory cartographic solutions have not yet been established.

One such area where cartographic innovation is ongoing is in the field of multimodal-, bicycle- and walking-based urban navigation. Potentially reflecting the need of more detailed information for urban mobility behaviours other than point-to-point movement, Google has recently added visualizations of "areas of interest" (Bliss, 2016) and detailed geometry of sidewalks (see Figure 9.1, left) in urban areas to their Google Maps product, complemented by optional layers for public transportation and cycling infrastructure. Other authors have acknowledged the need of bicycle navigation maps to show more or different information than what is needed for car-based navigation, such as slope, number of lanes, traffic volume and speed limits (Wessel & Widener, 2015), to help cyclists with diverse individual preferences make informed decisions. Recently, detailed visualizations of urban road layout and features, such as the precise arrangement and interaction of multiple lanes of traffic (Carlino, 2022; see Figure 9.1, right) or the arrangement of curbside parking (Seidel, 2021; see Figure 9.1, centre) have become feasible, based on the detailed modelling of such features in OpenStreetMap.

Figure 9.1 Left: Screenshot of Google Maps showing sidewalk and building geometry (original size). Middle: "Straßenraumkarte Neukölln" (Seidel, 2021)[76], visualizing parking arrangement along streets in Berlin (scaled down to 50%). Left: "A/B Street" (Carlino, 2022)[77], visualizing multiple lanes of traffic and complex intersection geometry (scaled down to 75%).

These current and upcoming applications indicate that there may be demand for more accurate representation of urban transport infrastructure and conditions, and that the data and algorithms to generate such visualizations are being currently developed. Clear guidelines on minimum dimensions may help designers of cartographic visualizations make confident decisions about the density of information to depict, and the map scales that are appropriate to fulfil a given information demand in detail while also providing as much overview of the overall spatial situation as possible.

Besides the potential that a renewed confidence in the fidelity of the cartographic medium may have for such novel applications in urban environments, it may also allow map designers to introduce additional elements to conventional digital maps. These may be additional topographic layers for a richer base map for thematic maps, traditional topographic elements such as contour lines, or artificial graphical elements such as grids or graticules. These elements have often been omitted from digital maps in the past, due to the assumption of a coarse pixel grid, in order to not overload the overall map image. Using finer lines and symbols would allow map designers to introduce such elements without conflicting with the main information layers of the map, which could improve subjective user satisfaction or objective map reading performance. For example, Edler et al. (2014) have shown that the introduction of grid lines in a topographic map design can improve the recall of object locations on the map. Using the finest lines that are clearly distinct from the symbology used for the layers of primary importance of the map, designers could introduce such aids without having to fear interference with the topographic content.

[76] https://strassenraumkarte.osm-berlin.org/?map=micromap, accessed 2022-07-12
[77] https://github.com/a-b-street/abstreet/, accessed 2022-07-12

9.3. Limitations of the presented studies

The empirical investigation of minimum dimensions involved many aspects and decisions, of which not all have been without difficulty or error. As mentioned above, the definition given in Section 9.1 may serve as a kind of index of the aspects involved: viewing situation, target demographics, selection of symbols, progression of sizes and method to detect a significant drop in performance. For each aspect, there are certainly ways in which the presented work was limited and may be improved by subsequent research projects.

The viewing situation presented participants with individual graphics, presented in isolation at near-optimal yet highly artificial viewing conditions with controlled viewing distance, lighting, and a single simple task of identifying the stimulus among a set of similar graphics. The main purpose of the controlled viewing situation was to establish a well-documented experimental setup that could be replicated by subsequent research. It is unclear how performance would be affected by a more realistic indoor or outdoor viewing scenario. Certainly, map viewing scenarios in the real world can vary a lot, from stable viewing in favourable lighting conditions, to taking a quick peek on the run at diminished contrast. As long as a standardized set of replicable viewing conditions for experimental cartographic research is not established, resorting to a stable, lab-based configuration will be a good choice for ensuring results that can be replicated across studies. Furthermore, the idea of a "minimum dimension" has been interpreted as implying favourable conditions. Establishing models for how cartographic symbology has to be adjusted to be reliably read in adverse conditions is a task left to future studies.

For most parts of the presented studies, the tasks asked of participants consisted of choosing, among some alternatives, the symbol that was shown on the stimulus display. While such task design may help participants to some extent by offering a limited set of choices among which only the best matching had to be selected, it was chosen partly due to its relationship with real-world cartographic applications, where either some sort of legend can be used to explain the symbology and define the available symbols, or some familiarity with the symbology can be assumed. Furthermore, in any real cartographic application, one would assume that cues can be extracted by the map reader from the spatial context of the symbol. In a real-world map usage scenario, there may be aspects making it harder to identify a given symbol (unclear set of choices, absence of a legend etc.), and other aspects making it easier (spatial context and patterns, local knowledge and experience, familiarity with the symbols etc.). Overall, the forced-choice design delivered initial results that certainly need to be reflected upon and adapted in the context of a specific real-world map use scenario. Other study designs, potentially involving free-form responses to a query of "what does this icon show?" would certainly be possible, and could deliver insights into scenarios with a lot of ambiguity or little knowledge of the map user about what information to expect at a certain location.

One of the main limitations of the presented studies is certainly the target demographics that was recruited for participation. For the purpose of generalising the results of participants with good visual acuity to the general population, one would hope that the medical sciences would provide a comprehensive model of the expected distribution and development of

habitual visual acuity in the overall population. However, as discussed in Section 3.1.2.1, a simple, ready-to-use and undisputed model that would allow the cartographer to calculate how a map design would need to be adapted to fit the needs of a certain target demographics does not exist. Within the constraints of this research project, a comprehensive recruiting effort to ensure participation of a representative cross-section of the overall demographics was found to be elusive. Instead, the focus was set on recruiting and identifying participants with "good" visual acuity, in order to establish plausible results for map users near the maximum of their visual capabilities. The group of participants with lower visual acuity was too small to draw valid conclusions from, and the test of visual acuity done during the experiment was basically used to exclude those participants with diminished visual capabilities from the main parts of the analysis. It would, however, be urgent to establish reliable models and guidelines for how cartographic symbology needs to be adjusted for the expected habitual visual acuity of different segments of the population, structured by age, occupation or place of residence. Whether such work can be undertaken within the domain of cartography or in collaboration of cartographers with other scientific fields, or whether such work would be considered wholly outside of the domain of cartography and only eventual outcomes would be adopted, is presently unclear.

Regarding the choice of symbols to present for icon discrimination tasks, the second study presented an approach to identify groups of similar icons from real-world icon sets by raster-based similarity analysis. This can be seen as an improvement over earlier studies, in which only generic geometric figures were used as stimuli, with questionable transferability of the resulting guidelines to other, more complex shapes or larger sets of icons. Although more sophisticated methods of shape similarity analysis exist, the raster-based method was the only one found to fulfil the specific requirements encountered for analysing the available sets of map icons taken from real-world cartographic applications, and that could be realistically implemented within the available resources. A comparison of the results of the chosen method with other automated methods and/or subjective judgement of perceptual similarity would certainly be desirable. As for the symbology of lines, there were no previous studies or collections of standardized symbology that served as blueprints for the designs used in the study. The used stimuli, with now well-documented geometry, can hopefully serve as a starting point for future investigations and potential standardization of the types of line symbology to be used for similar studies.

The method used to determine the progression of sizes at which stimuli are presented changed from the first study to the later studies. The "staircase" procedure, as adopted in the first study, used in psychophysics as a simple method to quickly advance towards a critical stimulus level near the individual perceptual threshold for each participant, was found to have two main disadvantages that make it less attractive for cartographic investigations: the irregular sampling of stimulus levels, with individually different distribution of stimuli at various size levels for each participant, and the resulting threshold values at size levels of arbitrary decimal precision and at success rates below the high values expected for cartographic applications. For this reason, the method for adapting the sizes has been changed in subsequent studies to a simpler method of a constant number of repetitions at predefined size levels. While this allowed for more meaningful analysis of results in many

cases, it also showed to have some limitations: for some tasks, the chosen stimulus levels failed to include sufficient trials at large enough sizes, so that the baseline performance could not be reliably established, or included levels way below an initial drop in performance, which wasted participants' time on stimulus levels that were clearly too small to be used in real-world applications anyway. Researchers attempting similar studies in the future would be advised to run more elaborate pilot studies to determine the levels at which stimuli are presented, or to err on the side of caution and include a wider range of size levels initially, of which the smallest levels can eventually be removed once an initial trend becomes apparent in the results. For most tasks of the presented studies, the stimuli were presented at monotonically decreasing size levels, which removed the need to include explicit instructional "trial tasks" at large enough size levels, as it ensured participants were always first confronted with a task at a clearly legible size. However, this means that trials at smaller sizes could benefit from a short-term learning effect of having just completed the tasks at larger sizes. For insights about a fully naïve map user, who encounters a certain symbol for the first time at a very small size, a different study design would be required that presents the symbols at small sizes without prior runs at larger sizes. In order to keep the participant in a "naïve" state, such study would probably have to use a much larger and randomized set of candidate icon sets, or ask participants for free-form responses.

The final aspect covered by the proposed definition for minimum dimensions is the method used for detecting a significant drop in performance. Driven by the goal of providing clear guidelines to map designers, studies two and three adopted a binary framework of analysis: a response is either correct or incorrect, and the grouped responses for a particular size level and display are either significantly different from another group or not. As the discussion of the results of both later studies has shown, for many tasks this framework of analysis has delivered useful and intuitively understandable results. However, in some cases, results were not as clear-cut as would have been desirable. Using an arbitrary static threshold of $p<0.05$ to detect "significance" has been widely criticized as potentially delivering misleading results (e.g. see A. Field, 2017, Chapter 3). This author believes that the p-value, as an assessment of the probability of two different success rates originating from an actual change in performance, is defensible as a means of quantifying a drop in performance (which, undoubtedly, can be assumed to be present at *some* point in a progression of increasingly smaller symbol sizes). However, the binary classification at above or below the arbitrary threshold of 0.05 could be questioned, and future analyses may benefit from an approach that visualizes the p-value of each drop on a continuous scale with appropriate graphical means, in order to better portray the results. But even with such a "continuous" approach, a threshold would have to be set above which the difference is not considered noteworthy at all (e.g. at $p<0.1$ or *0.2*?), and it would come at the cost of presenting a more complex picture instead of clear guidelines. Future proposals or different formalizations by other researchers of what a "significant drop" in performance means in the context of legibility of cartographic symbology would certainly be welcome. In the light of above considerations, the author of this thesis believes the choices made are defensible and delivered meaningful results that could be translated to the actionable guidelines presented in Chapter 8.

For some tasks, however, the combination of the chosen size levels of the stimulus and the analysis framework failed to deliver unambiguous results. The detection of the "initial drop" depends on reliably identifying baseline performance at large enough levels, which has been assumed to be achieved if a drop was neither detected between the initial two size levels, nor between performance levels on different displays for those initial levels. This was not the case for the results of Task 9 of the second study (counting icons from the set "Maki-triangular" on a map), and Task 8 of the third study (counting lines on a pseudo-map). Generally, the results of the counting tasks suffered from poor choices of size levels, and delivered somewhat surprising but inconclusive results (showing, contrary to the hypotheses, significantly worse results for the lowest-resolution display for counting icons, and significantly worse results for highest-resolution display for counting lines). For the purpose of minimum dimensions, only the results for isolated presentation were taken into account, and it can be argued that in the age of interactive maps, correct visual identification of information may be more important than more complex manual operations like counting or visual search, for which the presentation should be suitably adapted or the computer should provide interactive tools to assist with such tasks. Nevertheless, clear guidance on the impact of more complex map use tasks or tasks involving peripheral vision on recommended minimum dimensions would be desirable, and could not be unambiguously established in the course of the research presented in this thesis.

The framework adopted for the presented studies was a strictly behaviouralist one: success or failure have been determined solely based on the response of the participant to a given stimulus. Subjective aspects, such as the participants' preferences or comfort for performing the task at a given size level, have not been assessed. This decision was made in order to include as many task variants as possible in a single experiment session, which at an overall duration of 45-60 minutes per participant were felt to already be at the limit of participants' ability and willingness to participate in a concentrated manner. Subsequent studies would be necessary to assess additional aspects such as subjective preference or confidence, or speed of response in addition to the rate of correct responses this study focused on.

9.4. Future research opportunities

The limitations documented in the previous section can also be read as a plan for subsequent research that would be needed to overcome these limitations and to generate further insight expanding upon the findings presented in this thesis. For some of the limitations, potential next steps to work on overcoming them have already been mentioned above. Besides the compensation of shortcomings of the presented studies, further avenues of research can be conceived that expand on the work presented in this thesis.

It has already been mentioned that collaboration with researchers in ophthalmology and psychophysics could be pursued in order to refine and generalize the findings presented here. However, it is also important to note that the methods proposed in this thesis for the purpose of *cartographic* research deviate from the methods established in other disciplines.

While the stimuli were presented in isolation in the context of simple experimental tasks, the usage of real-world map icons and graphics resembling cartographic symbology marks a difference from highly artificial stimuli typically used in many psychophysical experiments. Also, the specific interest in stimulus levels allowing for recognition rates at or near 100% differs from the thresholds usually assessed for psychophysical experiments. Thirdly, initial attempts to increase the ecological validity of the results by including more complex tasks involving map-like stimuli have been presented and should be pursued further. These aspects make the investigation of minimum dimensions a genuinely cartographic endeavour, which will likely not be pursued further by any other discipline. From the design of stimuli and tasks, the requirements for software running the experiments, down to the statistical methods used to analyse the results, the requirements of and methods employed by cartography are distinct from other fields. A broader discussion and further development of genuinely cartographic empirical methods, connected to the long history of cartographic investigations of issues related to perception and cognition, seems to be called for.

While the presented studies attempted to cover a range of mobile phone displays, the results have not been directly compared to other display media. Both an empirical comparison of screen-based presentation to printed material, for which the established guidelines for that media type could be re-evaluated and compared to performance on modern digital displays, as well as to other digital media, such as e-ink displays or virtual- and augmented reality viewing devices, would have the potential to contextualize the findings in the current heterogeneous landscape of media technology available for the reproduction of cartographic content. In addition to investigating different media technologies, also the impact of the map viewing situation could be investigated, by taking into account situations in outdoor environments, while on the move, or under stress. Such investigations may contribute to a multi-dimensional model of minimum dimensions for cartography, which map designers could consult for guidance on symbology design based not only on map medium and demographics, but also map use task, viewing situation and other factors.

The presented simple method of raster-based shape similarity analysis provided a robust method for identifying clusters of similar icons, which were manually curated into sets of similar icons. Future improvement and verification of this method could lead to the creation of reliable models of shape similarity, which may form the basis of dynamic and automatically computed models for minimum symbol dimensions. This could facilitate the creation of automated tools which would allow the map designer to submit the collection of map symbols to be used, receiving advice on minimum dimensions and an analysis of problematic symbol pairs which are most likely to be confused, allowing for optimization of the design. The ColorBrewer project (Harrower & Brewer, 2003) is an example of the potential of academic cartography to produce practical tools which are widely adopted by practitioners in cartography, and even beyond the boundaries of the discipline. The raster-based map icon analysis presented in Section 6.1.2 could form the basis of such a browser-based analysis tool for map icons. The intervention of automated shape difference amplification, presented in Section 6.1.4, has not delivered unambiguous improvement of icon legibility, but should be examined further in order to potentially also suggest algorithmically determined interventions to improve map icon legibility at small sizes to map designers.

Even if legibility of small map symbology can be ensured by proposing suitable minimum dimensions, we do not know much about how very small symbols affect the user's behaviour. It may be the case that small symbols are perceived as an affordance (Meng, 2008), implicitly suggesting the user to zoom in to reveal more detail, or that small symbols may be well suited to guide the attention of the user (Swienty, 2008) to particular areas of the map for further exploration. Maybe a recognition rate close to 100% is, after all, not a goal that needs to be accomplished under all circumstances, if information represented too small to be reliably read has the potential to prompt the user to take further action to investigate complex spatial situations. After all, in the age of interactive maps, tools can often be provided that allow the user to resolve any uncertainty stemming from the graphical representation, by providing additional information upon mouse movement, clicking, tapping, or zooming in. This would allow map designers to use symbology even smaller than recommended by the guidelines presented in Chapter 8, if it could be shown that such representations have the potential to trigger desirable behaviour – but more research is needed to understand how this would affect user performance and satisfaction.

Judy Olson (1976) argues that there are two fundamental possibilities of improving map reading performance: changing the design, or training the map user. So far, all of the discussion has addressed issues of symbology design, and the performance of the designs has been tested without giving individuals the opportunity of training. However, it became apparent in the pilot studies that the author of this thesis was able to perform better than most participants in most tasks, despite not having better visual acuity, which can be hypothesized to be caused by the training effect that may have been achieved during the course of designing, implementing and repeatedly testing the experiments. Studies have shown that for some hyperacuity tasks, training can have an effect of a potential reduction of stimulus size needed for successful recognition of up to 40%, even if no feedback on the correctness of each trial is provided (Edelman & Weiss, 1995; McKee & Westheimer, 1978). While users of online maps can usually not be expected to spend time on training prior to reading a map, in certain professional settings (e.g. pilots, emergency services or the military), the effect of training on map reading performance for small symbols could be considered relevant and should certainly be studied.

As an alternative to static guidelines, which will be subject to simplification and generalisation even when differentiated by various aspects of map medium and viewing situation, Section 8.3 proposed guidelines for researchers and practitioners of cartography who wanted to conduct their own empirical examinations. The further development of tools and methods to empirically verify the suitability of cartographic designs for a stated purpose should certainly be on the agenda of academic cartography. The development of methods for automated adaption of stimulus sizes during experiments, to facilitate a process that minimizes the time spent by participants and experiment supervisors while maximizing generated insight, would certainly be an important aspect to investigate in the context of such a project. Again, providing web-based tools, ready to be used by map designers to test their designs, would be of benefit both to academic cartography and real-world cartographic projects.

9.5. Concluding remarks

On the outset, one of the questions this research project wanted to investigate was whether established guidelines mandating screen-based cartography to use much coarser symbology as compared to printed maps were still valid, or whether they could be updated, given the state of currently available display technology. The goal of this research was to fill the perceived gap in contemporary cartographic advice by empirical investigation, and to provide updated guidelines of practical use to map making practitioners. The presented studies adopted a psychophysical-behaviouralist approach, in the sense that the responses of participants to stimuli presented in a controlled lab situation were used to determine the success of a particular simple act of cartographic communication (in most cases, the classification of an isolated cartographic symbol as one of a given number of alternative choices). Such approach was considered best suited to answer the particular research question, and to close the identified gap in our knowledge by producing empirical evidence. The particular choice of method should not be read as advocating for a universal adoption of "neo-behaviouralist" positions and methods in cartography. It is a great privilege to be able to work in a discipline which produces so many exciting questions, many of which are too complex to answer with simple behaviouralist approaches. Thus, the many paradigms of cartography that were proposed and explored in the 20th and early 21st centuries (Azócar Fernández & Buchroithner, 2014) have every right to co-exist, and will hopefully continue to inform and inspire researchers of various backgrounds and diverse personal styles and preferences. This thesis has hopefully shown that a strictly empirical research design still has the potential to contribute to our increasingly detailed understanding of the mediated processes of cartographic communication and representation, and to deliver not only useful results but also raise some interesting further questions on the way that may inspire researchers and practitioners in cartography to explore further.

Besides the academic contribution, it is hoped that the work undertaken for this thesis has contributed something of practical value. Not only the guidelines presented in Chapter 8, but also the symbol geometries created for the studies, the icon sets identified for the second study, the framework for establishing and conducting the lab-based studies presented in Chapter 4, and the software created for the work presented in this thesis, which are all released under open-source licenses, are hoped to be of practical use for cartographers, academics and map-makers in one way or another. Also, the results data collected during the three studies is made available to allow other researchers to conduct further analyses on it, if so desired. While Chlupac in 1982 expressed gratitude that entering the results data of his experiments on map symbol legibility could be completed in a time span of merely six weeks, thanks to the advances of computer technology at the time, today 40 years later the code and results data for experiments can be shared globally, to be analysed, replicated and built upon by others, immediately. It is hoped that by sharing the code and results of the experiments described in this thesis, a small contribution towards such a globally shared collaborative approach to research has been made, and that the further verification, extension and development of the findings and ideas presented here will not have to wait for another 40 years.

Supplementary Materials

Code and documentation of stimsrv, the software framework developed for implementing the studies presented in this thesis, is available at

https://github.com/floledermann/stimsrv

Materials for Study 1

The source code and materials of the first study are available at
https://github.com/floledermann/experiment-pixel-density

The raw results data of Study 1 is available at
https://doi.org/10.5281/zenodo.5520608

A computational notebook for analysing the results data is available at
https://observablehq.com/@floledermann/pixel-density

Materials for Study 2

The source code and materials of the second study are available at
https://github.com/floledermann/experiment-point-symbols

A computational notebook for visualizing icon similarity metrics is available at
https://observablehq.com/@floledermann/icon-similarities

A computational notebook for analysis of results, including the raw data, is available at
https://observablehq.com/@floledermann/data-analysis-point-symbols

Materials for Study 3

The source code and materials of the third study are available at
https://github.com/floledermann/experiment-line-symbols

A computational notebook for analysis of results, including the raw data, is available at
https://observablehq.com/@floledermann/data-analysis-point-symbols

List of Figures

List of Tables

Bibliography

American Geographical Society. (2021). *The Locus Charter*. https://ethicalgeo.org/locus-charter/

Android Open Source Project. (2021). *Support different pixel densities*. Android Developers Documentation. https://developer.android.com/training/multiscreen/screendensities

Apple Inc. (2012a). *IOS Drawing Concepts*. Drawing and Printing Guide for IOS. https://developer.apple.com/library/archive/documentation/2DDrawing/Conceptual/DrawingPrintingiOS/GraphicsDrawingOverview/GraphicsDrawingOverview.html

Apple Inc. (2012b). *Supporting High-Resolution Screens in Views*. Drawing and Printing Guide for IOS. https://developer.apple.com/library/archive/documentation/2DDrawing/Conceptual/DrawingPrintingiOS/SupportingHiResScreensInViews/SupportingHiResScreensInViews.html

Apple Inc. (2022). *Apple Human Interface Guidelines: Images*. Apple Developer. https://developer.apple.com/design/human-interface-guidelines/foundations/images

Arnberger, E., & Kretschmer, I. (1975). *Wesen und Aufgaben der Kartographie – Topographische Karten* (Vol. 1). Franz Deuticke.

Azócar Fernández, P. I., & Buchroithner, M. F. (2014). *Paradigms in Cartography*. Springer.

Bach, M. (1996). The Freiburg Visual Acuity test – Automatic Measurement of Visual Acuity. *Optometry and Vision Science*, *73*(1), 49–53. https://doi.org/10.1097/00006324-199601000-00008

Bach, M. (1997). Anti-aliasing and Dithering in the "Freiburg Visual Acuity Test." *Spatial Vision*, *11*(1), 85–89. https://doi.org/10.1163/156856897x00087

Bach, M. (2007). The Freiburg Visual Acuity Test – Variability unchanged by post-hoc re-analysis. *Graefe's Archive for Clinical and Experimental Ophthalmology*, *245*(7), 965–971. https://doi.org/10.1007/s00417-006-0474-4

Bach, M. (2020). *Freiburg Vision Test (FrACT)* (Version 10) [JavaScript; Browser Application]. https://michaelbach.de/fract/

Balabash, M., & Oskin, A. (2020). The Inside of Headless Chrome. *Materials of the 12th Junior Researchers' Conference*. https://elib.psu.by/bitstream/123456789/31117/1/154-155.pdf

Beier, S. (2009). *Typeface Legibility: Towards Defining Familiarity* [PhD Thesis, Royal College of Art]. https://researchonline.rca.ac.uk/957/

Bertin, J. (1983). *Semiology of Graphics: Diagrams, Networks, Maps*. University of Wisconsin Press.

Birsak, L. (2000). Konstanten und Unterschiede im Kartenentwurf für gedruckte und Bildschirmkarten. In F. Kelnhofer & M. Lechthaler (Eds.), *Interaktive Karten (Atlanten) und Multimedia-Applikationen* (pp. 14–25). Institut für Kartographie und Reproduktionstechnik, TU Wien. https://repositum.tuwien.at/handle/20.500.12708/368

Bliss, L. (2016). *The Real Problem With "Areas of Interest" on Google Maps*. CityLab. http://www.citylab.com/design/2016/08/google-maps-areas-of-interest/493670/

Bollmann, J. (1981). *Aspekte kartographischer Zeichenwahrnehmung: Eine empirische Untersuchung*. Kirschbaum-Verlag.

British Cartographic Society. (2020, July 20). *Code of Ethics*. https://www.cartography.org.uk/code-of-ethics

Brodie, S. E. (2019). Light. In M. Yanoff & J. S. Duker (Eds.), *Ophthalmology* (5th ed., pp. 38–47). Elsevier.

Brown, A. (1993). Map design for screen displays. *The Cartographic Journal*, *30*(2), 129–135. https://doi.org/10.1179/000870493787860030

Bruce, V., Green, P. R., & Georgeson, M. A. (2003). *Visual Perception: Physiology, Psychology, and Ecology* (4th ed.). Routledge.

Brunner, K. (2000). Neue Gestaltungs- und Modellierungsaufgaben für den Kartographen – Ein Plädoyer für eine attraktive Kartengraphik zur Bildschirmvisualisierung. In F. Kelnhofer & M. Lechthaler (Eds.), *Interaktive Karten (Atlanten) und Multimedia-Applikationen* (pp. 53–62). Institut für Kartographie und Reproduktionstechnik, TU Wien. https://repositum.tuwien.at/handle/20.500.12708/368

Brunner, K. (2001). Kartengraphik am Bildschirm – Einschränkungen und Probleme. *KN – Journal of Cartography and Geographic Information*, *51*(5), 233–239. https://doi.org/10.1007/BF03544835

Brunner, K. (2002). Topogramme für kleinformatige Displays mobiler Endgeräte. *Kartographische Nachrichten*, *52*(3), 103–106. https://doi.org/10.1007/BF03544889

Brunner-Friedrich, B., & Nothegger, C. (2002). *Concepts for user-oriented cartographic presentation on mobile devices – A pedestrian guidance service for the TU Vienna*. 189–196.

Buchroithner, M. F. (2016). Will paper maps remain smart apps? *Kartographische Nachrichten*, *66*(1), 3–6. https://doi.org/10.1007/BF03545179

Buchwald, J. Z., & Feingold, M. (2013). *Newton and the Origin of Civilization*. Princeton University Press.

Burghardt, D., Edwardes, A., Purves, R., & Weibel, R. (2005). Automatische Generalisierung dynamisch generierter Karten für mobile Endgeräte. *KN – Journal of Cartography and Geographic Information*, *55*(3), 119–125. https://doi.org/10.1007/BF03544008

Byford, S. (2021, October 8). *The OLED Nintendo Switch doesn't have a Pentile screen*. The Verge. https://www.theverge.com/2021/10/8/22715977/oled-nintendo-switch-no-pentile-screen-macro-photo

Campbell, F. W., & Maffei, L. (1974). Contrast and Spatial Frequency. *Scientific American*, *231*(5), 106–115.

Carlino, D. (2022). *A/B Street* (0.3.24). https://doi.org/10.5281/zenodo.6636265

Cartwright, W. (2003). Maps on the Web. In M. P. Peterson (Ed.), *Maps and the internet* (pp. 35–56). Elsevier.

Cauvin, C., Escobar, F., & Serradj, A. (2010). *Thematic Cartography. Vol. 1, Thematic Cartography and Transformations*. Wiley.

Chang, D. F. (2017). Ophthalmologic Examination. In P. Riordan-Eva & J. Augsburger (Eds.), *Vaughan & Asbury's General Ophthalmology* (19th ed., pp. 17–65). McGraw Hill.

Chen, H.-W., Lee, J.-H., Lin, B.-Y., Chen, S., & Wu, S.-T. (2018). Liquid crystal display and organic light-emitting diode display: Present status and future perspectives. *Light: Science & Applications*, *7*(3), Article 3. https://doi.org/10.1038/lsa.2017.168

Chen, H.-W., Tan, G., & Wu, S.-T. (2017). Ambient contrast ratio of LCDs and OLED displays. *Optics Express*, *25*(26), 33643–33656. https://doi.org/10.1364/OE.25.033643

Cheung, Y. K., Li, Z., & Chen, W. (2009). Integration of Cognition-based Content Zooming and Progressive Visualization for Mobile-based Navigation. *The Cartographic Journal*, *46*(3), 268–272. https://doi.org/10.1179/000870409X12472347560542

Chlupac, J. (1982). *Die Erkennbarkeits- und Unterschiedsschwelle verschiedener geometrischer Signaturenformen* [PhD Thesis]. Universität Wien.

Clairvoyante. (2010). *Measuring Display Resolution with Contrast Modulation Methodology* [Application Note]. Clairvoyante.

Colenbrander, A. (2016). *Measuring Vision and Vision Loss*. Ento Key – Fastest Otolaryngology & Ophthalmology Insight Engine. https://entokey.com/measuring-vision-and-vision-loss/

Çöltekin, A., Bleisch, S., Andrienko, G., & Dykes, J. (2017). Persistent challenges in geovisualization – a community perspective. *International Journal of Cartography*, *3*(sup1), 115–139. https://doi.org/10.1080/23729333.2017.1302910

Cornsweet, T. N. (1962). The Staircase-Method in Psychophysics. *American Journal of Psychology*, *75*(3), 485–491. https://doi.org/10.2307/1419876

Credelle, T. L., & Brown Elliott, C. H. (2005). High-Pixel-Density PenTile Matrix RGBW Displays for Mobile Applications. *Proceeding of the 5th International Meeting on Information Display (IMID 2005)*, 867–872.

Cunningham, D. W., & Wallraven, C. (2011). *Experimental Design: From User Studies to Psychophysics*. CRC Press.

Dash, P., & Hu, Y. C. (2021). How much battery does dark mode save? An accurate OLED display power profiler for modern smartphones. *Proceedings of the 19th Annual International Conference on Mobile Systems, Applications, and Services*, 323–335. https://doi.org/10.1145/3458864.3467682

Dent, B. D. (1972). Visual Organization and Thematic Map Communication. *Annals of the Association of American Geographers*, *62*(1), 79–93. https://doi.org/10.1111/j.1467-8306.1972.tb00845.x

Dent, B. D., Torguson, J., & Hodler, T. (2008). *Cartography: Thematic Map Design* (6th edition). McGraw-Hill Education.

Diaz, J. (2010, June 7). *iPhone 4: The Definitive Guide*. Gizmodo. https://gizmodo.com/iphone-4-the-definitive-guide-5557101

Diniz, D., Irochima, F., & Schor, P. (2019). Optics of the Human Eye. In M. Yanoff & J. S. Duker (Eds.), *Ophthalmology* (5th ed., pp. 26–37). Elsevier.

Dobson, M. W. (1985). The future of perceptual cartography. *Cartographica: The International Journal for Geographic Information and Geovisualization*, *22*(2), 27–43. https://doi.org/10.3138/E482-2512-732Q-2T11

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, *96*, 433–458. https://doi.org/10.1037/0033-295X.96.3.433

Durrie, D., Stulting, R. D., Potvin, R., & Petznick, A. (2019). More eyes with 20/10 distance visual acuity at 12 months versus 3 months in a topography-guided excimer laser trial: Possible contributing factors. *Journal of Cataract and Refractive Surgery*, *45*(5), 595–600. https://doi.org/10.1016/j.jcrs.2018.12.008

Eaton, R. M. (1993). Designing the electronic chart display. *The Cartographic Journal*, *30*(2), 184–187. https://doi.org/10.1179/000870493787860094

Edelman, S., & Weiss, Y. (1995). Visual Hyperacuity. In M. A. Arbib (Ed.), *The handbook of brain theory and neural networks* (pp. 1009–1012). MIT Press.

Edler, D., Bestgen, A.-K., Kuchinke, L., & Dickmann, F. (2014). Grids in Topographic Maps Reduce Distortions in the Recall of Learned Object Locations. *PLOS ONE*, *9*(5), e98148. https://doi.org/10.1371/journal.pone.0098148

Elliott, D. B., Yang, K. C., & Whitaker, D. (1995). Visual acuity changes throughout adulthood in normal, healthy eyes: Seeing beyond 6/6. *Optometry and Vision Science*, *72*(3), 186–191. https://doi.org/10.1097/00006324-199503000-00006

Ellsiepen, I., & Morgenstern, D. (2007). Der Einsatz des Bildschirms erweitert die kartographischen Gestaltungsmittel. *KN – Journal of Cartography and Geographic Information*, *57*(6), 303–309. https://doi.org/10.1007/BF03544040

Eysenck, M. W., & Keane, M. T. (2020). *Cognitive Psychology: A Student's Handbook* (8th ed.). Psychology Press.

Fairbairn, D. (2005). Geovisualization Issues in Public Transport Applications. In J. Dykes, A. M. MacEachren, & M.-J. Kraak (Eds.), *Exploring Geovisualization* (pp. 513–528). Elsevier. https://doi.org/10.1016/B978-008044531-1/50444-9

Fang, L., Au, O. C., Tang, K., & Wen, X. (2012). Increasing image resolution on portable displays by subpixel rendering – a systematic overview. *APSIPA Transactions on Signal and Information Processing*, *1*(1), e1. https://doi.org/10.1017/ATSIP.2012.3

Field, A. (2017). *Discovering Statistics Using IBM SPSS Statistics* (5th ed.). SAGE.

Field, K. (2022, January 11). Ethics in mapping. *ArcGIS Blog*. https://www.esri.com/arcgis-blog/products/arcgis-pro/mapping/ethics-in-mapping/

Fish, C., & Brewer, C. A. (2015). GIS as a Tool for Map Production. In M. S. Monmonier (Ed.), *The History of Cartography, Volume 6 – Cartography in the Twentieth Century* (pp. 504–510). University Of Chicago Press.

Flannery, J. J. (1971). The relative effectiveness of some common graduated point symbols in the presentation of quantitative data. *Cartographica: The International Journal for Geographic Information and Geovisualization*, *8*(2), 96–109.

Frisén, L., & Frisén, M. (1981). How good is normal visual acuity? *Graefe's Archive for Clinical and Experimental Ophthalmology*, *215*(3), 149–157. https://doi.org/10.1007/BF00413146

Fujii, T., Kon, C., Motoyama, Y., Shimizu, K., Shimayama, T., Yamazaki, T., Kato, T., Sakai, S., Hashikaki, K., Tanaka, K., & Nakano, Y. (2018). 4032 ppi High-resolution OLED Microdisplay. *Journal of the Society for Information Display*, *26*(3), 178–186. https://doi.org/10.1002/jsid.656

Gardiner, R. A. (1981). Cartography and the Mind of Man. *The Cartographic Journal*, *18*(2), 112–115. https://doi.org/10.1179/caj.1981.18.2.112

Gartner, G. (2009). Web Mapping 2.0. *Mitteilungen Der Österreichischen Geographischen Gesellschaft*, *151*, 277–290. https://doi.org/10.1553/moegg151s277

Ghose, T. (2015, May 14). *Brain Training Exercise Gives Athletes "Super Vision."* LiveScience. https://www.livescience.com/50835-brain-training-improves-vision.html

Giraud, T., & Lambert, N. (2017). Reproducible Cartography. In M. P. Peterson (Ed.), *Advances in Cartography and GIScience* (pp. 173–183). Springer International Publishing. https://doi.org/10.1007/978-3-319-57336-6_13

Glasser, A. (2011). Accomodation. In L. A. Levin, S. F. E. Nilsson, J. V. Hoeve, & S. Wu (Eds.), *Adler's Physiology of the Eye* (11th ed., pp. 40–70). Saunders.

Goldstein, E. B., & Brockmole, J. (2016). *Sensation and Perception* (10th ed.). Cengage Learning.

Gooding, K., & Forrest, D. (1990). An examination of the difference between the interpretation of screen based and printed maps. *The Cartographic Journal*, *27*(1), 15–19. https://doi.org/10.1179/000870490786961852

Google Inc. (2021). *Change text & display settings*. Android Accessibility Help. https://support.google.com/accessibility/android/answer/11183305

Griffin, A. L., Robinson, A. C., & Roth, R. E. (2017). Envisioning the future of cartographic research. *International Journal of Cartography*, *3*(sup1), 1–8. https://doi.org/10.1080/23729333.2017.1316466

Griffin, A. L., White, T., Fish, C., Tomio, B., Huang, H., Sluter, C. R., Bravo, J. V. M., Fabrikant, S. I., Bleisch, S., Yamada, M., & Picanço, P. (2017). Designing across map use contexts: A research agenda. *International Journal of Cartography*, *3*(3), 90–114. https://doi.org/10.1080/23729333.2017.1315988

Grünreich, D. (2008). Perspektiven und Aufgaben der Kartographie. *KN – Journal of Cartography and Geographic Information*, *58*(6), 301–307. https://doi.org/10.1007/BF03543997

GSM Arena. (2015). *Sony explains why the Z5 Premium only uses its native 4K resolution when needed*. https://www.gsmarena.com/sony_explains_why_the_z5_premium_only_uses_its_native_4k_resolution_when_needed-news-14078.php

Hake, G., Grünreich, D., & Meng, L. (2002). *Kartographie: Visualisierung raum-zeitlicher Informationen* (8th ed.). de Gruyter.

Hamm, L. M., Yeoman, J. P., Anstice, N., & Dakin, S. C. (2018). The Auckland Optotypes: An open-access pictogram set for measuring recognition acuity. *Journal of Vision*, *18*(3), 13–13. https://doi.org/10.1167/18.3.13

Harrower, M., & Brewer, C. A. (2003). ColorBrewer.org: An Online Tool for Selecting Colour Schemes for Maps. *The Cartographic Journal*, *40*(1), 27–37.

Healey, C., & Enns, J. (2012). Attention and Visual Memory in Visualization and Computer Graphics. *IEEE Transactions on Visualization and Computer Graphics*, *18*(7), 1170–1188. https://doi.org/10.1109/TVCG.2011.127

Hearn, D., & Baker, M. P. (1996). *Computer Graphics, C Version* (2nd Edition). Prentice Hall.

Heupel, A. (1977). Über die „wissenschaftliche Tätigkeit" bei kartographischen Arbeiten. *Kartographische Nachrichten*, *77*(3), 90–93.

Hitchcock, G. (2005, August 11). Where does 96 DPI come from in Windows? *Windows Font Blog*. https://docs.microsoft.com/en-us/archive/blogs/fontblog/where-does-96-dpi-come-from-in-windows

Hooke, R. (1674). An attempt to prove the motion of the earth by observations. In *Lectiones Cutlerianae, or a collection of lectures: Physical, mechanical, geographical & astronomical*. John Martyn, Printer to the Royal Society. https://library.si.edu/digital-library/book/lectionescutler00hook

Imhof, E. (1972). *Thematische Kartographie*. Walter de Gruyter.

International Council of Ophthalmology. (1984). *Visual acuity measurement standard*. International Council of Ophthalmology. http://www.icoph.org/dynamic/attachments/resources/icovisualacuity1984.pdf

International Council of Ophthalmology. (2002). *Visual Standards – Aspects and Ranges of Vision Loss* [Report prepared for the International Council of Ophthalmology at the 29th International Congress of Ophthalmology]. International Council of Ophthalmology. http://www.icoph.org/downloads/visualstandardsreport.pdf

Jenks, G. F., & Knos, D. S. (1961). The Use of Shading Patterns in Graded Series. *Annals of the Association of American Geographers*, *51*(3), 316–334. https://doi.org/10.1111/j.1467-8306.1961.tb00381.x

Jenny, B., Jenny, H., & Räber, S. (2008). Map design for the Internet. In M. P. Peterson (Ed.), *International perspectives on maps and the internet* (pp. 31–48). Springer.

Kalat, J. W. (2018). *Biological Psychology* (13th ed.). Cengage Learning.

Kalmykova, Y., Patrício, J., Rosado, L., & Berg, P. E. (2015). Out with the old, out with the new – The effect of transitions in TVs and monitors technology on consumption and WEEE generation

in Sweden 1996–2014. *Waste Management*, *46*, 511–522.
https://doi.org/10.1016/j.wasman.2015.08.034

Kelley, E. F., Lindfors, M., & Penczek, J. (2006). Display daylight ambient contrast measurement methods and daylight readability. *Journal of the Society for Information Display*, *14*(11), 1019–1030. https://doi.org/10.1889/1.2393026

Kleiner, M. (2013). *Introduction to Psychtoolbox-3*. European Conference on Visual Perception, 25-29 August 2013, Bremen, Germany.

Knorr, H. (1967). Über Gegenwartsprobleme der Kartographie. *Kartographische Nachrichten*, *5*(67), 149–161.

Koenker, R., & Zeileis, A. (2009). On reproducible econometric research. *Journal of Applied Econometrics*, *24*(5), 833–847.

Koláčný, A. (1969). Cartographic Information – a Fundamental Concept and Term in Modern Cartography. *The Cartographic Journal*, *6*(1), 47–49. https://doi.org/10.1179/caj.1969.6.1.47

Lechthaler, M., & Stadler, A. (2006, February 13). *Cross-Media gerechte Kartengraphik in einem AIS*. CORP 2006, Vienna, Austria.

Ledermann, F., & Gartner, G. (2021). Towards Conducting Reproducible Distributed Experiments in the Geosciences. *AGILE: GIScience Series*, *2*(33). https://doi.org/10.5194/agile-giss-2-33-2021

Lieberman, H. R., & Pentland, A. P. (1982). Microcomputer-based estimation of psychophysical thresholds: The Best PEST. *Behavior Research Methods & Instrumentation*, *14*(1), 21–25. https://doi.org/10.3758/BF03202110

Lobben, A. K., & Patton, D. K. (2003). Design Guidelines for Digital Atlases. *Cartographic Perspectives*, *44*, 51–62. https://doi.org/10.14714/CP44.515

Lu, G., & Sajjanhar, A. (1999). Region-based shape representation and similarity measure suitable for content-based image retrieval. *Multimedia Systems*, *7*(2), 165–174. https://doi.org/10.1007/s005300050119

Luckiesh, M., & Moss, F. K. (1937). The visibility of various type faces. *Journal of the Franklin Institute*, *223*(1), 77–82. https://doi.org/10.1016/S0016-0032(37)90585-4

MacEachren, A. M. (1995). *How Maps Work: Representation, Visualization, and Design*. Guilford Press.

MacEachren, A. M. (2019). (Re)Considering Bertin in the age of big data and visual analytics. *Cartography and Geographic Information Science*, *46*(2), 101–118. https://doi.org/10.1080/15230406.2018.1507758

Madsen, M., Lhoták, O., & Tip, F. (2020). A Semantics for the Essence of React. In R. Hirschfeld & T. Pape (Eds.), *Proceedings of the 34th European Conference on Object-Oriented Programming (ECOOP 2020)* (p. 12:1-12:26). Schloss Dagstuhl – Leibniz-Zentrum für Informatik. https://doi.org/10.4230/LIPIcs.ECOOP.2020.12

Malić, B. (1998). *Physiologische und technische Aspekte kartographischer Bildschirmvisualisierung* [PhD Thesis]. Rheinische Friedrich-Wilhelms-Universität.

Mańk, A. K. (2019). *Cartographic Symbolization for High-Resolution Displays* [Master's Thesis, TU Wien]. https://doi.org/20.500.12708/2696

Marble, D. F. (2015). Computational Geography as a New Modality. In M. S. Monmonier (Ed.), *The History of Cartography, Volume 6 – Cartography in the Twentieth Century* (pp. 488–492). University Of Chicago Press.

McKee, S. P., & Westheimer, G. (1978). Improvement in Vernier acuity with practice. *Perception & Psychophysics*, *24*(3), 258–262. https://doi.org/10.3758/BF03206097

McMaster, R. B., & Shea, K. S. (1992). *Generalization in digital cartography*. Association of American Geographers.

Meng, L. (2008). *To see and see through graphics –Toward affordance-driven geovisualization*. Virtual Geographic Environments – An International Conference on Development in Visualization and Virtual Environments in Geographic Information Science.

Meng, L., Bandrova, T., Midtbø, T., & Voženílek, V. (2021). Toward a New ICA Research Agenda. *Abstracts of the ICA*, *3*, 1–2. https://doi.org/10.5194/ica-abs-3-205-2021

Metha, A. B., Vingrys, A. J., & Badcock, D. R. (1993). Calibration of a color monitor for visual psychophysics. *Behavior Research Methods, Instruments, & Computers*, *25*(3), 371–383. https://doi.org/10.3758/BF03204528

Microsoft Inc. (2021). *High DPI Desktop Application Development on Windows*. Windows App Development Documentation. https://docs.microsoft.com/en-us/windows/win32/hidpi/high-dpi-desktop-application-development-on-windows

Mileff, P., & Dudra, J. (2012). Efficient 2D software rendering. *Production Systems and Information Engineering*, *6*(2), 99–110.

Montello, D. R. (2002). Cognitive Map-Design Research in the Twentieth Century: Theoretical and Empirical Approaches. *Cartography and Geographic Information Science*, *29*(3), 283–304. https://doi.org/10.1559/152304002782008503

Morgenstern, D. (1974). Zur optimalen Auswahl einer Tonwertskala. *Kartographische Nachrichten*, *74*(2), 45–53.

Morrison, C., & Forrest, D. (1995). A study of point symbol design for computer based large scale tourist mapping. *The Cartographic Journal*, *32*(2), 126–136. https://doi.org/10.1179/caj.1995.32.2.126

Muehlenhaus, I. (2014). *Web Cartography*. CRC Press.

Neudeck, S. (2001). *Zur Gestaltung topografischer Karten für die Bildschirmvisualisierung* [PhD Thesis]. Universität der Bundeswehr München.

Nguyen, T. C. (2021, January 30). *The Brief History of Smartphones*. ThoughtCo. https://www.thoughtco.com/history-of-smartphones-4096585

Nissen, F., Hvas, A., Münster-Swendsen, J., & Brodersen, L. (2003). *Small-display cartography* (Project Report IST-2000-30090 / D3.1.1). Aalborg University. https://vbn.aau.dk/ws/files/453264/GiMoDig.pdf

Nüst, D., & Pebesma, E. (2021). Practical Reproducibility in Geography and Geosciences. *Annals of the American Association of Geographers*, *111*(5), 1300–1310. https://doi.org/10.1080/24694452.2020.1806028

Ohlsson, J., & Villarreal, G. (2005). Normal visual acuity in 17–18 year olds. *Acta Ophthalmologica Scandinavica*, *83*(4), 487–491. https://doi.org/10.1111/j.1600-0420.2005.00516.x

Olson, J. M. (1976). A Coordinated Approach to Map Communication Improvement. *The American Cartographer*, *3*(2), 151–160. https://doi.org/10.1559/152304076784080177

Peirce, J. W. (2007). PsychoPy – Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1), 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017

Petchenik, B. B. (1983). A mapmaker's perspective on map design research 1950–1980. In D. R. F. Taylor (Ed.), *Graphic communication and design in contemporary cartography*. John Wiley & Sons.

Phillips, R. J. (1979). An Experiment with Contour Lines. *The Cartographic Journal*, *16*(2), 72–76. https://doi.org/10.1179/caj.1979.16.2.72

Phillips, R. J., Lucia, A., & Skelton, N. (1975). Some Objective Tests of the Legibility of Relief Maps. *The Cartographic Journal*, *12*(1), 39–46. https://doi.org/10.1179/caj.1975.12.1.39

Pointer, J. S. (2008). Recognition versus Resolution: A Comparison of Visual Acuity Results Using Two Alternative Test Chart Optotype. *Journal of Optometry*, *1*(2), 65–70. https://doi.org/10.3921/joptom.2008.65

Radner, W., & Benesch, T. (2019). Age-related course of visual acuity obtained with ETDRS 2000 charts in persons with healthy eyes. *Graefe's Archive for Clinical and Experimental Ophthalmology*, *257*(6), 1295–1301. https://doi.org/10.1007/s00417-019-04320-3

Reichenbacher, T. (2001). Adaptive concepts for a mobile cartography. *Journal of Geographical Sciences*, *11*(1), 43–53. https://doi.org/10.1007/BF02837443

Reichenbacher, T. (2004). *Mobile cartography: Adaptive visualisation of geographic information on mobile devices* [PhD Thesis]. Technische Universität München.

Reichenbacher, T., & Meng, L. (2003). Forschungsfragen der mobilen Kartographie. *KN – Journal of Cartography and Geographic Information*, *53*(2), 67–70. https://doi.org/10.1007/BF03544954

Riordan-Eva, P. (2017). Anatomy & Embryology of the Eye. In P. Riordan-Eva & J. Augsburger (Eds.), *Vaughan & Asbury's General Ophthalmology* (19th ed., pp. 17–65). McGraw Hill.

Robinson, A. C. (2011). Challenges and opportunities for web-based evaluation of the use of spatial technologies. *Proceedings of the International Cartographic Conference (ICC 2011)*. ICC 2011, Paris, FR.

Robinson, A. C. (2019). Elements of viral cartography. *Cartography and Geographic Information Science*, *46*(4), 293–310. https://doi.org/10.1080/15230406.2018.1484304

Robinson, A. H. (1952). *The look of maps: An examination of cartographic design*. University of Wisconsin Press.

Robinson, A. H., Sale, R. D., Morrison, J. L., & Muehrcke, P. C. (1984). *Elements of Cartography* (5th ed.). Wiley.

Roth, R. E. (2015). *Challenges for Human Subjects Research in Cartography*. ICA Workshop on Envisioning the Future of Cartographic Research, Curitiba, Brazil.

Roth, R. E. (2019). *What is mobile first cartographic design?* ICA Joint Workshop on User Experience Design for Mobile Cartography, Beijing, China.

Roth, R. E., Baldrica-Franklina, G., Kellyb, M., Underwooda, N., & Sack, C. (2021, May 6). *Distributing map studies*. Workshop on Adaptable Research Methods For Empirical Research with Map Users. https://cogvis.icaci.org/21_adaptable/Roth.pdf

Roth, R. E., Çöltekin, A., Delazari, L., Filho, H. F., Griffin, A., Hall, A., Korpi, J., Lokka, I., Mendonça, A., & Ooms, K. (2017). User studies in cartography: Opportunities for empirical research on interactive maps and visualizations. *International Journal of Cartography*, *3*(sup1), 61–89.

Rozario, H. (2022, April 6). *How to Change or Remove Font Smoothing on MacOS Monterey & Big Sur*. OS X Daily. https://osxdaily.com/2022/04/06/change-remove-font-smoothing-macos/

Ruda, H. (2013). *Estimation of the parameters of a boundary contour system using psychophysical hyperacuity experiments* [PhD Thesis, Boston University]. https://doi.org/10.13140/RG.2.1.1875.0808/1

Rueckner, W., & Papaliolios, C. (2002). How to beat the Rayleigh resolution limit: A lecture demonstration. *American Journal of Physics*, *70*(6), 587–594. https://doi.org/10.1119/1.1463736

Salmon, J. F. (2019). *Kanski's Clinical Ophthalmology: A Systematic Approach* (9th ed.). Elsevier.

Schor, P., & Miller, D. (2011). Optics. In L. A. Levin, S. F. E. Nilsson, J. V. Hoeve, & S. Wu (Eds.), *Adler's Physiology of the Eye* (11th ed.). Saunders.

Schuurman, N. C. (2000). Trouble in the heartland: GIS and its critics in the 1990s. *Progress in Human Geography*, *24*(4), 569–590. https://doi.org/10.1191/030913200100189111

Schweizerische Gesellschaft für Kartografie (Ed.). (1980). *Kartographische Generalisierung: Topographische Karten* (2nd ed.). Schweizerische Gesellschaft für Kartografie.

Schweizerische Gesellschaft für Kartografie (Ed.). (2002). *Topografische Karten: Kartengrafik und Generalisierung*. Schweizerische Gesellschaft für Kartografie.

Seidel, A. (2021). *Parkraumanalyse für den Berliner Ortsteil Neukölln*. https://strassenraumkarte.osm-berlin.org/parkraumkarte/methodenbericht-2021-03.pdf

Semmo, A., Trapp, M., Jobst, M., & Döllner, J. (2015). Cartography-Oriented Design of 3D Geospatial Information Visualization – Overview and Techniques. *The Cartographic Journal*, *52*(2), 95–106. https://doi.org/10.1080/00087041.2015.1119462

Sheedy, J., Tai, Y.-C., Subbaram, M., Gowrisankaran, S., & Hayes, J. (2008). ClearType sub-pixel text rendering: Preference, legibility and reading performance. *Displays*, *29*(2), 138–151. https://doi.org/10.1016/j.displa.2007.09.016

Sloan, L. L. (1951). Measurement of visual acuity: A critical review. *AMA Archives of Ophthalmology*, *45*(6), 704–725.

Sloan, L. L. (1959). New test charts for the measurement of visual acuity at far and near distances. *American Journal of Ophthalmology*, *48*(6), 807–813.

Slocum, T. A. (1999). *Thematic Cartography and Visualization* (1st ed.). Prentice Hall.

Slocum, T. A., McMaster, R. B., Kessler, F. C., & Howard, H. H. (2009). *Thematic cartography and geovisualization* (3rd Edition). Prentice hall.

Snellen, H. (1873). *Probebuchstaben zur Bestimmung der Sehschärfe* (4th ed.). H. Peters.

Snowden, R., Thompson, P., & Troscianko, T. (2012). *Basic Vision: An Introduction to Visual Perception* (Revised edition). Oxford University Press.

Spiess, E. (1990a). Generalisierung in thematischen Karten. In Schweizerische Gesellschaft für Kartographie (Ed.), *Kartographisches Generalisieren* (pp. 63–70).

Spiess, E. (1990b). Siedlungsgeneralisierung. In Schweizerische Gesellschaft für Kartographie (Ed.), *Kartographisches Generalisieren* (pp. 49–55).

Spolsky, J. (Ed.). (2008). Font Smoothing, Anti-Aliasing, and Subpixel Rendering. In *More Joel on Software* (pp. 85–87). Apress. https://doi.org/10.1007/978-1-4302-0988-1_11

Stamm, B. (1998). *The raster tragedy at low resolution or: Why correct math looks wrong on screen and how to fix it*. Microsoft Typography. https://web.archive.org/web/20050507121506/https://www.microsoft.com/typography/tools/trtalr.aspx

Stamm, B. (2009). *The Raster Tragedy at Low-Resolution Revisited: Opportunities and Challenges beyond "Delta-Hinting."* http://rastertragedy.com/

Swienty, O. (2008). *Attention-Guiding Geovisualisation* [PhD Thesis, Technische Universität München]. https://mediatum.ub.tum.de/node?id=652641

Tainz, P., & Weber, W. (1996). Kartographische Test-Software zur empirischen Untersuchung von Arbeitsprozessen mit Bildschirmkarten. *Kartographische Nachrichten*, *44*(3), 98–102.

The Chromium Project. (2009). *How Chromium Displays Web Pages*. Chromium Design Documents. https://www.chromium.org/developers/design-documents/displaying-a-web-page-in-chrome/

Tjukanov, T. (2021, April 8). *From traditional cartography to spatial data visualization* [Presentation]. International Spring School on Visualization, Olomouc, Czechia.

Tobler, W. R. (1979). A Transformational View of Cartography. *The American Cartographer*, *6*(2), 101–106.

Tobler, W. R. (1988). Resolution, Resampling, and All That. In H. Mounsey (Ed.), *Building Databases for Global Science* (pp. 129–137). Taylor & Francis.

Töpfer, F. (1974). *Kartographische Generalisierung*. Haack.

Tsujimura, T. (2017). *OLED Display Fundamentals and Applications* (2nd edition). Wiley.

US Department of Defense. (2014). *Department Of Defense Interface Standard – Joint Military Symbology* (MIL-STD-2525D).

van den Worm, J. (2001). Web map design in practice. In M.-J. Kraak & A. Brown (Eds.), *Web cartography: Developments and Prospects* (pp. 91–111). Taylor & Francis.

van Elzakker, C. P. J. M. (2001). Users of maps on the Web. In M.-J. Kraak & A. Brown (Eds.), *Web cartography: Developments and Prospects* (pp. 38–53). Taylor & Francis.

van Elzakker, C. P. J. M., Ormeling, F., Köbben, B. J., & Cusi, D. (2003). Dissemination of Census and other Statistical Data through Web Maps. In M. P. Peterson (Ed.), *Maps and the internet* (pp. 35–56). Elsevier.

Vanecek, E. (1980). *Experimentelle Beiträge zur Wahrnehmbarkeit kartographischer Signaturen* (Veröffentlichungen des Instituts für Kartographie der Österreichischen Akademie der Wissenschaften, Vol. 6). Verlag der Österreichischen Akademie der Wissenschaften.

Velasco e Cruz, A. A. (1990). Historical Roots of 20/20 as a (Wrong) Standard Value of Normal Visual Acuity. *Optometry and Vision Science*, *67*(8), 661. https://doi.org/10.1097/00006324-199008000-00022

Veltkamp, R. C., & Latecki, L. J. (2006). Properties and Performance of Shape Similarity Measures. In V. Batagelj, H.-H. Bock, A. Ferligoj, & A. Žiberna (Eds.), *Data Science and Classification* (pp. 47–56). Springer. https://doi.org/10.1007/3-540-34416-0_6

VESA Display Metrology Committee. (2001). *Video Electronics Standards Association Flat Panel Display Measurements Standard Version 2.0*. VESA Display Metrology Committee.

Virrantaus, K., Fairbairn, D., & Kraak, M.-J. (2009). ICA Research Agenda on Cartography and GI Science. *The Cartographic Journal*, *46*(2), 63–75. https://doi.org/10.1179/000870409X459824

Ware, C. (2020). *Information Visualization – Perception for Design* (4th ed.). Morgan Kaufmann. https://doi.org/10.1016/C2016-0-02395-1

Watson, A. B., & Pelli, D. G. (1983). Quest: A Bayesian adaptive psychometric method. *Perception & Psychophysics*, *33*(2), 113–120. https://doi.org/10.3758/BF03202828

Werner, F. (1970). Bemerkungen zur thematischen Kartographie. *Kartographische Nachrichten*, *3*(70), 103–109.

Wessel, N., & Widener, M. (2015). Rethinking the Urban Bike Map for the 21st Century. *Cartographic Perspectives*, *0*(81), 6–22. https://doi.org/10.14714/CP81.1243

Westheimer, G. (1981). Visual Hyperacuity. In H. Autrum, E. R. Perl, R. F. Schmidt, & D. Ottoson (Eds.), *Progress in Sensory Physiology* (pp. 1–30). Springer. https://doi.org/10.1007/978-3-642-66744-2_1

Westheimer, G. (2009). Visual acuity: Information theory, retinal image structure and resolution thresholds. *Progress in Retinal and Eye Research*, *28*(3), 178–186. https://doi.org/10.1016/j.preteyeres.2009.04.001

Wood, M. (1968). Visual Perception and Map Design. *The Cartographic Journal*, *5*(1), 54–64. https://doi.org/10.1179/caj.1968.5.1.54

Wood, M. (1993). The map-users' response to map design. *The Cartographic Journal*, *30*(2), 149–153.

World Wide Web Consortium. (2011). *Cascading Style Sheets Level 2 Revision 1 (CSS 2.1) Specification* (B. Bos, T. Çelik, I. Hickson, & H. W. Lie, Eds.). https://www.w3.org/TR/2011/REC-CSS2-20110607/

World Wide Web Consortium. (2016). *CSS Object Model: View Module* (W3C Working Draft TR/cssom-view-1). World Wide Web Consortium. https://www.w3.org/TR/cssom-view-1/

World Wide Web Consortium. (2019). *CSS Values and Units Module Level 3* [W3C Candidate Recommendation]. https://www.w3.org/TR/css-values-3/

World Wide Web Consortium. (2020a). *CSS Media Queries Level 4* (F. Rivoal & T. Atkins, Eds.). https://www.w3.org/TR/mediaqueries-4/

World Wide Web Consortium. (2020b). *Resize Observer* (W3C First Public Working Draft TR/resize-observer). World Wide Web Consortium. https://www.w3.org/TR/resize-observer/

World Wide Web Consortium. (2022, May 12). *W3C Ethical Web Principles* [W3C Group Draft Note]. https://www.w3.org/TR/ethical-web-principles/

Worth, C. (1989). Problems with experimental research in map design—A case study. *The Cartographic Journal*, *26*(2), 148–153. https://doi.org/10.1179/caj.1989.26.2.148

Xia, V. (2017, December 21). *What is Mobile First Design? Why It's Important & How To Make It?* https://medium.com/@Vincentxia77/what-is-mobile-first-design-why-its-important-how-to-make-it-7d3cf2e29d00

Zhang, H. (2020a). *Pixel-snapping in icon design*. https://uxdesign.cc/pixel-snapping-in-icon-design-a-rendering-test-6ecd5b516522

Zhang, H. (2020b, March 16). *Icon Grids & Keylines Demystified*. https://minoraxis.medium.com/icon-grids-keylines-demystified-5a228fe08cfd

Zipf, A. (2003). Forschungsfragen zur Benutzer- und kontextangepassten Kartengenerierung für mobile Systeme. *KN – Journal of Cartography and Geographic Information*, *53*(1), 6–11. https://doi.org/10.1007/BF03544939

Curriculum Vitae

Name:	Florian Ledermann
Date of birth:	7. April 1976
Place of birth:	Salzburg, Austria

Education

2016 – 2022	PhD studies, Cartography and Geoinformation, TU Wien
2004	MSc degree, "Computer Science and Design", TU Wien
1994 – 2004	Studies of Architecture & Computer Science, TU Wien
1994	"Matura" (school-leaving exam)
1986 – 1994	"Gymnasium" BRG Akademiestraße, informatics branch, Salzburg

Academic professional experience

2016 – 2022	University Assistant (PreDoc), TU Wien, Research Division Cartography
2014 – 2016	Researcher, TU Wien, Research Group Cartography
2005 – 2014	Lecturer, TU Wien, Institute for Software Technology & Interactive Media
2002 – 2007	Researcher, TU Wien, Institute for Software Technology & Interactive Media
1998 – 1999	Researcher, TU Wien, Institute for Design and Assessment of Technology

Awards & honours

2018	Best Paper Award, GI_Forum 2018 (best student paper), for paper *Analysing Digital Maps Online: A Reverse Engineering Approach*
2017	Best Teaching Award, TU Wien (best course of the faculty), for course *Web Mapping*
2017	1st prize, International Cartographic Exhibition, category educational products, for project *genderATlas*
2008	Best of Google Maps 2008 (Maps Mania Blog), for online novel *Senghor on the Rocks*

Selected publications

F. Ledermann (forthcoming): *Minimum Dimensions for Cartographic Symbology – History, Rationale and Relevance in the Digital Age*. (Accepted for publication in the International Journal of Cartography, forthcoming special focus issue on the State of the Art & Science in Cartography)

F. Ledermann (2022): *The Effect of Display Pixel Density on the Minimum Legible Size of Fundamental Cartographic Symbols*. The Cartographic Journal (in press).
https://doi.org/10.1080/00087041.2022.2055938

F. Ledermann (2022): *Minimum Dimensions for Cartographic Point Symbols on Mobile Phone Screens: Theoretical Considerations and Empirical Verification*. In: Abstracts of the ICA, Vol. 5 (EuroCarto 2022: Proceedings of the European Cartographic Conference, Vienna, Austria) p. 110. https://doi.org/10.5194/ica-abs-5-110-2022

F. Ledermann, G. Gartner (2021): *Towards Conducting Reproducible Distributed Experiments in the Geosciences*. In: Open-access proceedings of the Association of Geographic Information Laboratories in Europe (AGILE 2021). AGILE: GIScience Series, 2 (2021), p. 33–39.
https://doi.org/10.5194/agile-giss-2-33-2021

F. Ledermann (2021): *Small differences: Limits of Legibility of Cartographic Symbols on High- and Ultra-High-Resolution Mobile Displays*. In LBS 2021: Proceedings of the 16th International Conference on Location Based Services, p. 177-182. https://doi.org/10.34726/1785

F. Ledermann (2021): *Investigating the Effect of Display Pixel Density on the Minimal Legible Size of Cartographic Symbols*. In: Abstracts of the ICA, Vol. 3 (ICC 2021: Proceedings of the 30th International Cartographic Conference, Firenze, Italy), p. 174. https://doi.org/10.5194/ica-abs-3-174-2021

F. Ledermann, G. Gartner (2021): *Towards Reproducible User Studies and Behavioral Experiments for Cartography With the stimsrv Framework*. In: Abstracts of the ICA, Vol. 3 (ICC 2021: Proceedings of the 30th International Cartographic Conference, Firenze, Italy), p. 175. https://doi.org/10.5194/ica-abs-3-175-2021

F. Ledermann: *Analysing Digital Maps Online: A Reverse Engineering Approach*. GI_Forum 2018 (2). (Best Paper Award, GI Forum 2018)
https://doi.org/10.1553/giscience2018_02_s119

F. Ledermann, G. Gartner: *mapmap.js: A Data-Driven Web Mapping API for Thematic Cartography*. Brazilian Journal of Cartography, 67 (2015), 5 (Special Issue ICC2015).

M. Riegler, M. Wenk, E. Aufhauser, F. Ledermann, M. Schmidt, G. Gartner: *genderATlas Österreich - Entwicklung eines zielgruppenorientierten Online-Tools*. Mitteilungen der Österreichischen Geographischen Gesellschaft, 157 (2015).
https://doi.org/10.1553/moegg157s323