The Cartographic Journal *Vol. 0 No. 0 pp. 1–11 Month 2014*

REFEREED PAPER

An Empirical User Study for Measuring the Influence of Colour Distance and Font Size in Map Reading Using Eye Tracking

Alzbeta Brychtova[1] *and Arzu Coltekin*[2]

[1]Palacky University in Olomouc, Olomouc, Czech Republic. [2]University of Zurich, Switzerland
Email: alzbeta.brychtova@gmail.com

The primary goal of this study is to empirically analyse the influence of colour distance *and* font size *on map readability. We utilized eye-tracking to complement the classical usability metrics; thus, we studied performance metrics such as* effectiveness *(i.e. success, accuracy),* efficiency *(i.e. time to answer, task completion time), and selected eye tracking metrics* fixation frequency, fixation duration *and* scanpath speed *as well as conducting an* area-of-interest (AOI) analysis *to understand the performance and strategy issues that may be influenced by colour distance and font size during map reading. The user experiment was carried out in a controlled laboratory where participants were asked to conduct a visual search task and mark the correct answer with a mouse click on a static map on a computer screen. Collected data was analysed through descriptive and inferential statistics. Task completion times for the five tested colour distances show that as the colour distances grow larger, the relative differences in task completion times become statistically significant; empirically confirming our intuition that larger colour distances are better for map readability. The comparison of the scanpath speeds for the tested font sizes suggests that the medium font size leads to a more efficient search.*

Keywords: usability, readability, eye-tracking, colour, map labelling

INTRODUCTION

Cartographers have long been interested in evaluating the usability of visual representations (e.g. effectiveness, efficiency and user satisfaction) to be able to distinguish a 'good map' from a 'bad map'. A common overarching goal in cartographic user research appears to be obtaining rules to ensure *quality*, and use these rules to create a 'good map' (note that we will use the word *map* to express the entire spectrum of geographic visualisations throughout this paper).

The quality of a map depends on various factors – e.g. design choices should be supported by empirical evidence where possible, the map must be a good fit for its intended purpose and ideally, it must be tailored for its audience. Besides these, another important factor determining the map quality is the background of the said audience, i.e. the map user. When people of different backgrounds and different skill levels use maps, it is likely (and to some degree, demonstrated) that each group will perform best with a different map, and subjectively rate also different maps as the 'best map' (Olson, 1979). To address such subjective differences, typically usability (or in a wider sense, user-experience) tests are conducted with human subjects, and most commonly *efficiency* (speed), *effectiveness* (accuracy) and *satisfaction* are measured (ISO, 1998). Performance metrics (efficiency and effectiveness) can influence the satisfaction, that is, the users are likely to prefer the map that facilitates faster/more accurate results, especially given that the preference questions are asked after the tasks are executed (Schnürer *et al.*, 2014). This is, however, not always the case – sometimes users prefer a particular design, but do not perform well with it (Hegarty *et al.*, 2009).

Besides the user-centric ('bottom-up') thinking, there are theoretical ('top-down') approaches in determining the quality of a map. Among these top-down quality criteria, a central and self-evident one is a map's readability. But what are the factors that limit the readability of a map? Harrie and Stigmar (2009) report two kinds of map readability limitations: The first is connected with *complexity* of the map (i.e. quantity and distribution of the information included on the map display), while the second is concerned with the *visualisation choices and design of cartographic symbols.* Some design principles are well established for cartography (e.g. Slocum *et al.*, 1999; Dent *et al.*, 2009) based on conventions that were applied in map-making for centuries (Garlandini and Fabrikant, 2009). In one of the most established and well-recognized cartographic theories,

DOI: 10.1179/1743277414Y.0000000103

Bertin (1983) classifies the basic elements of a visualisation design into seven visual variables; size, colour value, colour hue, texture, orientation and shape. Bertin's (1983) systematisation of the visual variables was later extended by MacEachren (2004), who proposed including dynamic variables relevant to digital products: display time, duration, order, rate of change, frequency and synchronisation. Modifying visual variables will change the meaning of the map symbol, therefore, visual variables are regularly and purposefully modified in visualisations to convey the desired meaning.

Another important and commonly applied top-down design consideration is to maintain 'sufficient' *visual distance* between map symbols to keep them readable and thus, their meaning distinguishable (e.g. Bjorke, 1996). While the visual distance is often conceptualized as the spatial distance between graphic objects, it is also relevant in determining other kinds of quantitative and qualitative distances between visual variables – e.g. on a choropleth map, a slightly different shade of the same colour represents a different value and we need to distinguish the two shades to be able to process the visual information. Determining a sufficient visual distance between the values of the same variable (so that we can tell them apart) has been an important subject in cartography and is relevant today as the visualisation hardware and software keeps changing and we have newer methods to obtain more information from user studies.

In this paper, we present a study on the readability of map labels as we modify the visual distance between two fundamental visual variables (size and colour value), using eye tracking in addition to the traditional usability metrics similar to others in recent literature (e.g. Çöltekin *et al.*, 2009). These two visual variables are critical to readability of map labels, and recent studies show that size and colour are among the strongest visual variables (e.g. Chesneau, 2007; Gartner and Hiller 2008; Garlandini and Fabrikant, 2009). Therefore, we contribute indirectly to the research on visual variables and directly to visualisation design research by empirically investigating the effect of varying levels of colour distance and font size on the users' ability to read the labels. More specifically, we conduct a user experiment in which users need to identify place names as we vary the size of the labels, and the colour of the background (label colour remains the same, but the distance between label's colour and the background changes). We study the eye movement data to observe users' visual strategies and how they divide their attention in task-relevant and irrelevant areas of the display.

RELATED WORK

Map labels

There are dozens of research papers in literature on label placement and finding algorithmic solutions for placing labels on a map. Many of these studies are concerned with the spatial position of the labels in which the goal is to avoid occlusion or overlaps with other map objects (e.g. Wagner and Wolf, 1995; Doddi *et al.*, 1997; van Dijk, 2001; Polishchuk and Vihavainen, 2010; Ooms *et al.*, 2012). One of the goals in our study (i.e. to find what kind of colour distance is appropriate between a label's colour and the background colour) is relevant to label placement problem, as the automatic label placement approaches should be concerned about the colour configuration of the algorithmically proposed position of the label.

One of the earliest experimental studies on readability of map labels in relation to colour is by Robinson (1952). He studied selected set of labels with varying colours on varying background colours. According to this study, the best combination is black letters on a white background, while the worst are black on purple, red on green or orange on white background (Robinson, 1952). His study sets the basis for the questions we study here at a conceptual level, however, the paper lacks precise specification of stimuli settings, i.e. investigated colours are not reported numerically in any colour space coordinates; and thus, we are unable to repeat the experiment, or re-use the colour combinations. Robinson's (1952) study does not include a font-size variation for labels. In relation to label size, an early study appears to be by Bartz (1970) in which she reports a study on visual search efficiency for map labels in various conditions. Bartz (1970) experimented with maps containing labels under what she called 'fixed conditions' (constant font size and type) and 'mixed conditions' (various font sizes and types on a map sheet). In her study, no differences were observed for the readability of serif and serif-sans fonts, nor did varying font size produce a significant effect on the search time.

According to Deeb *et al.* (2011, 2012) determining the most legible typographic variable depends on two factors, map *aesthesis* and map *efficiency*. They reported a study which focused on identifying preferred visual variables for map labels taking the influence of expertise level of map users (i.e. cartography education and practice) into account. They found that preferences of the examined groups vary unsystematically, thus, based on this study, determination of most legible typographic variable appears to remain inconclusive. Deeb *et al.*'s approach (2011, 2012) has some methodological similarities to our work; however, the study asks a different question and does not include colour variations.

Colour distance

Colour perception is a subjective process (Gegenfurtner and Sharpe, 2001), however, to enable reproduction, it has been measured and quantified into reference systems – termed as *colour spaces*. Colour spaces are divided into two groups in relation to human colour perception: perceptually non-uniform and uniform (CIE, 2012). The typical colour systems integrated in cartography and GIS software packages (such as RGB, HSV, HSB and HSL[1]) use perceptual terminology to label dimensions (such as hue, value, lightness, brightness, saturation), but they are not necessarily perceptually uniform (Brewer *et al.*, 2003). In order to ensure conformity between the measured colour distances and human perception, it is necessary to work with perceptually uniform colour spaces (Slocum *et al.*, 1999). Currently, most commonly used perceptually uniform colour spaces appear to be the CIE 1976 (L^*, a^*, b^*)

– *CIELAB* and the CIE 1976 ($L^* u^* v^*$) – *CIELUV* (Landa and Fairchild, 2005). These two models were developed to describe (ideally) all colours perceived by human eye (Dent *et al.*, 2009).

Colour, as a dominant visual variable, operates in a preponderant way in readability problems (Stigmar, 2010). Therefore, it is evident that *colour distance* should be sufficiently large to allow identification of symbols, preferably with maximum ease. To address this concern, International Commission on Illumination (CIE) has introduced ΔE as their colour distance metric. The most commonly used calculations of colour distance are based on determining the linear distance in the CIELAB colour space, such as the CIE76 (ΔE^*ab) and the CIEDE2000 (ΔE_{00}), where the latter is a refinement of the former (Werman, 2012). According to Carter and Huertas (2009), CIEDE2000 is more reliable than CIELAB when applied to very small colour differences (as intended) as well as very large colour differences (as later demonstrated). For this study, it is important that the formulation works well for larger colour distances, thus we chose to work with the CIEDE2000. The specifications of the formulas are published, for example by Sharma *et al.* (2004).

There is a substantial amount of research with the goal to optimize colours for map users with various colour vision deficiencies (e.g. Culp, 2012; Jenny and Kelso, 2007; Olson and Brewer, 1997). Majority of existing studies are focused on investigating the efficiency of sequential colour schemes (Kimerling, 1985; Gilmartin and Shelton, 1989; Chesneau, 2007; Buard and Ruas, 2009), qualitative data visualisation (Vondráková *et al.*, 012; Kröger *et al.*, 2013) or investigating colour scales for special mapping purposes - for example noise maps (Schiewe and Weninger, 2013). Related to our work, Steinrücken and Plümer (2013) have introduced an approach for optimizing colour configuration for on-demand maps based on the idea of keeping the minimum colour distance as high as possible. Arguably, the most significant contribution in the visualisation domain regarding colours is the ColorBrewer[2] tool developed by Harrower and Brewer (2003). This online software provides specifications of colour scales of different shades and numbers of categories taking their distinguishability on liquid-crystal displays (LCD) or printed materials into account and suggests optimized colour scales for people with colour vision impairments. The tool was designed based on colour theory and empirical studies with colour vision impaired map users (Olson and Brewer, 1997).

In summary, from the perspective of the two visual variables we study, font size and colour use appear to have been studied to some degree. However, how the font size interacts with colour distance, i.e. the numerical difference that determines how far two colours or shades are apart from each other appears to have never been studied in the context of map readability.

EXPERIMENTAL DESIGN

The aim of the presented study is to detect the influence of *colour distance between map labels and the background*, *the font size*, and *the combinations of changing size and colour* on the readability of the labels. The general hypothesis of the study is that increasing colour distance between map labels and background and increasing font size will have a positive impact on the readability, i.e. map users will have less trouble while searching for and reading the labels. To test this hypothesis, we studied two **independent variables**: *colour distance* and *font size*.

We used eye-tracking to determine the efficiency and strategy of stimulus reading (Salvucci and Goldberg, 2000). Raw eye-tracking data are typically classified based on spatial and temporal thresholds into fixations and saccades. It is assumed that humans can perceive a particular object during a fixation (i.e. only when looking at this point a certain amount of time), while during fast saccadic movements they do not actually register what they see (Holmqvist *et al.*, 2011). Eye-tracking metrics can be derived based on fixations and saccades (e.g. number, duration and dispersion of fixations, length and direction of the recorded trajectories of the view).

Our **dependent variables**, therefore, were *accuracy*, *speed*, and a selected set of eye movement metrics, namely *fixation frequency*, *fixation duration* and *scanpath speed*, to interpret users' strategies, as well as an area of interest (AOI) analysis.

Materials

Experimental stimuli were significantly simplified maps presented as static images of the size 1920 × 1080 pixels (see Figure 1 for an illustration). The simplification of the thematic and graphical content was necessary for experimental control. Stimuli include names of administrative units of five different states of the USA (Arizona, Nevada, Maine, New Jersey and Wyoming). The shapes of the states as well as number of labels (administrative units) on stimuli varied (Arizona 15, Nevada 17, Maine 18, New Jersey 21 and Wyoming 24). However, the within-subject design[3] of the experiment ensures that this variation does not influence the results for the independent variable *font size* (all participants work with all conditions). For the independent variable *colour distance*, participants worked with a subset for each condition, but the variation here was subtle enough that we did not expect any effect on the results after the pilot study (this was later confirmed in a dedicated analysis as well).

A total of 15 maps were prepared (Figure 1). Each stimulus is characterized by the concrete value of colour distance between labels and the background colour. Five levels of colour distance (which are of equal distance in RGB space and correspond to CIEDE2000 formula ΔE_{00}=30, 50, 70, 85 and 100) and three levels of font size (8, 11 and 14 pt) were examined.

Selected font sizes express precisely what was shown to the participants and were selected because they are commonly used in regular displays as well as traditional atlases (Robinson, 1952). Regarding the colour choice, we decided to begin with studying primary colours and among them, green; as the human vision is most sensitive to green spectrum (Dent *et al.*, 2009). The labels were in all cases in pure black (RGB (0,0,0)), because this is the most common choice for labels and sans-serif font type (Arial), because this

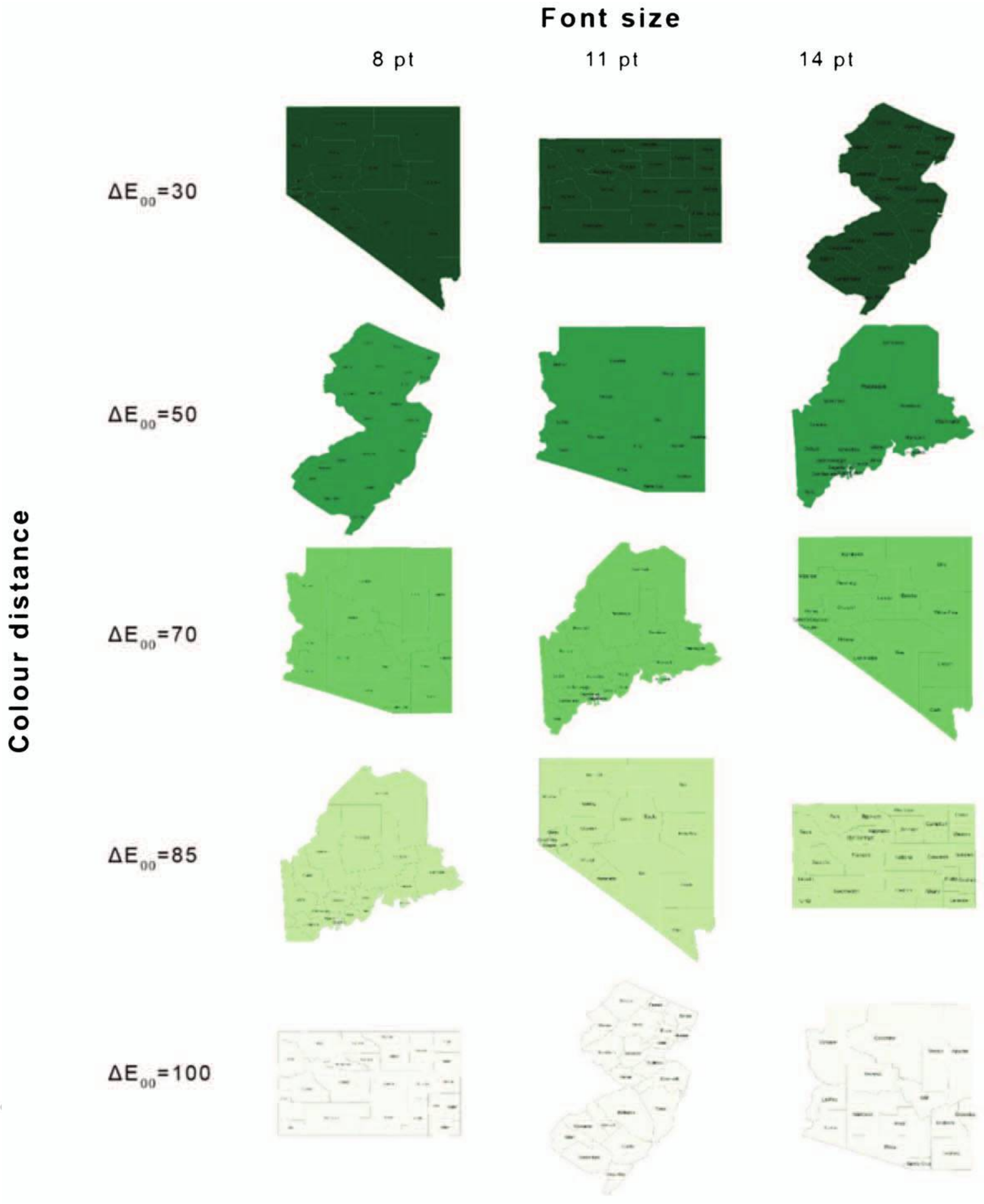

Figure 1. Experimental stimuli. All examined combinations of colour distance and font size are presented. For comparability, geographical scales of the states were distorted so their extent is equally large. Size of labels given in typographic points (pt; 1 pt=0.37 mm) corresponds to sizes that participants have seen during the experiment

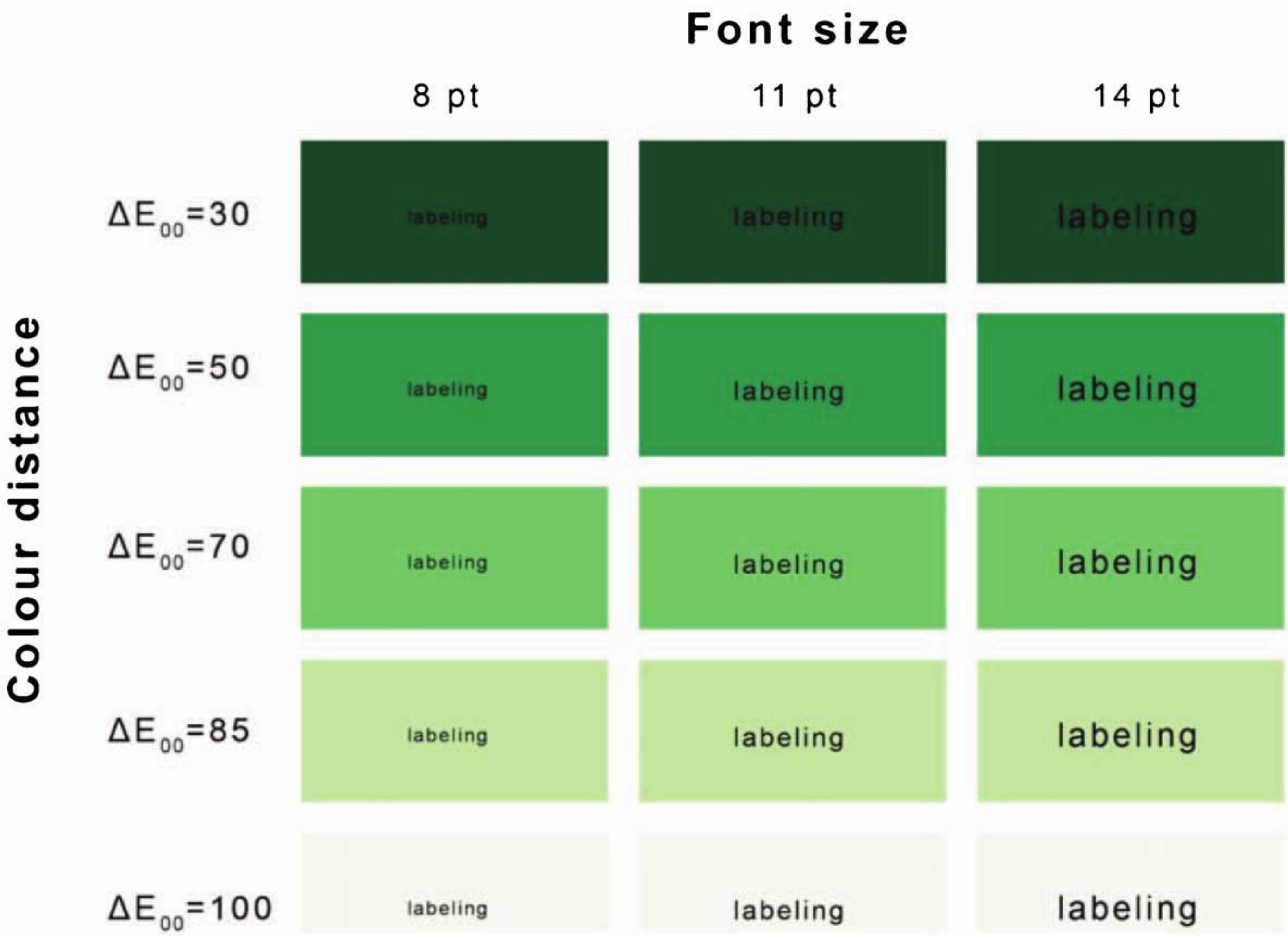

Figure 2. Examined colour combinations and font sizes (independent variables). Note that shown colours are tentative; i.e. they might look different than how it was displayed during the experiment

is considered a legible font and assumed to facilitate visual search for place names on maps well (Slocum *et al.*, 1999).

Based the previous research (as reported in the Section *Colour Distance*), we applied the perceptually uniform CIE 1976 $L^*a^*b^*$ colour space in this study. The distance between two colours was computed with the CIEDE2000 method. $L^*a^*b^*$ values were related to CIE standard illuminant D65. Experimental stimuli were shown in a digital environment and therefore only colours from CIE 1976 $L^*a^*b^*$ gamut, that can be converted into real values of RGB colour space, were used. The computation of colour distance and transformation between CIE 1976 $L^*a^*b^*$ and RGB colour space were done with use of the web calculator designed by Lindbloom (2012). A tentative view of all examined colour distances is shown in Figure 2. Related numerical colour specification (RGB and $L^*a^*b^*$ codes) of the background and colour distance levels to the black labels (calculated by three different methods) can be seen in Table 1.

Procedure

Experiment was carried out under controlled laboratory conditions at the Department of Geoinformatics, Palacký University, Olomouc, which is equipped with a low-frequency non-contact eye-tracker SMI RED 250 with a sampling

Table 1. Colour specifications of examined map background in RGB and Lab and corresponding colour distances between specified backgrounds and black labelling

Colour distance			RGB specifications			Lab specifications		
ΔE_{00}	ΔE^*_{ab}	ΔE_{RGB}	R	G	B	L	a	b
30	48	20	0	80	37	29.08	−32.7	19.02
50	80	40	36	154	80	56.15	−48.43	29.37
70	95	60	117	207	119	75.91	−44.52	34.88
85	94	80	199	232	175	88.42	−20.48	23.49
100	100	100	249	255	245	99.32	−3.3	3.12

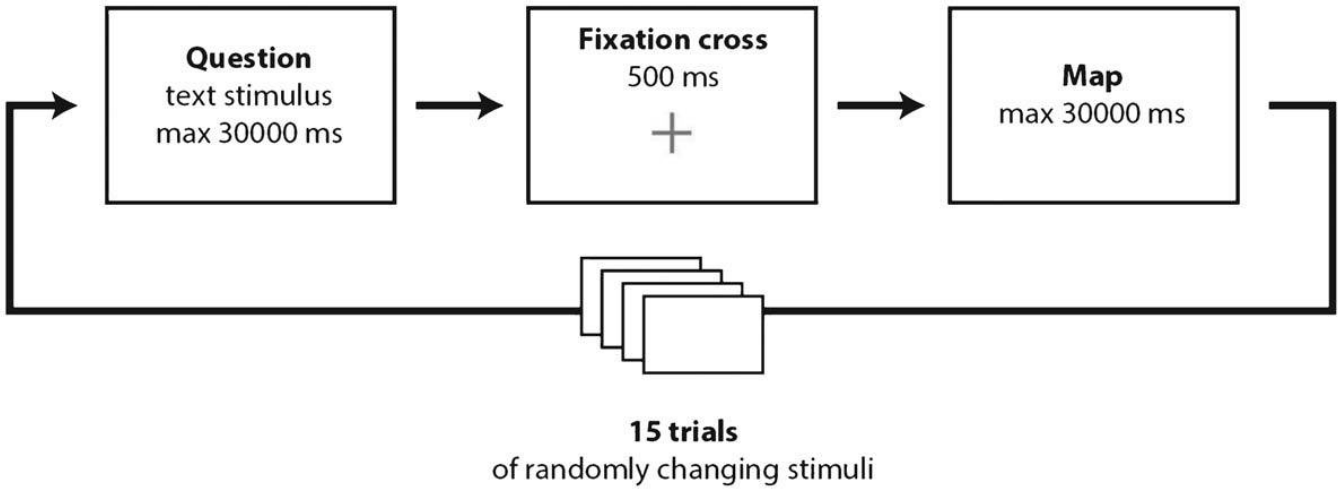

Figure 3. Basic experimental design

frequency of 120 Hz. Stimuli were presented on 23″ LG Flatron monitor IPS231P. Experiment was prepared using SMI Experiment Center™ (SensoMotoric Instruments, 2013a).

In a within-subject manner, participants were asked to *find an administrative unit by its name* and mark it with a mouse click within a given time limit. Stimuli were prepared so that we could avoid the effect of previous geographical knowledge of the areas, i.e., we asked participants if they were familiar with the studied areas and admitted only those who were not. We randomized the order of the stimuli to distribute the learning effect, which is a common practice in experimental studies (e.g. see Holmqvist *et al.*, 2011).

As can be seen in Figure 3, the experiment was composed of 15 randomized trials. Each trial contained a text stimulus with the *task description*, a *fixation cross* and a *map*. We presented the task description for 30 seconds, the fixation cross for 0.5 seconds and the map stimulus for 30 seconds. Time limits were introduced as a measure to control the experiment's length and have been obtained based on pilot studies. We optimized the time limits so that participants have enough time to solve the task without feeling time pressure. If a participant did not provide a response within this time frame, we considered them unsuccessful at that task.

Participants

53 volunteers took part in the experiment and they were offered no compensation. Gaze data from a total of 50 participants were used, i.e. data from three participants with tracking ratio less than 90% and calibration accuracy higher than 1° were removed. Participants were asked to state their age, sex, whether they had colour vision deficiencies, and their levels of expertise in cartography. The participant age ranged between 20–25 years. All of them were students of Palacký University, Olomouc. Thirty out of 50 participants (~60%) took at least one cartography course ('experts'); 20 out of 50 participants (~40%) have no previous cartography experience ('novices'). Twenty-eight of them were males (18 experts and 10 novices) and 22 were females (12 experts and 10 novices). None of participants reported any colour vision deficiency.

ANALYSIS METHODS

Collected data were analysed with regards to usability metrics *effectiveness* (accuracy of the answers), *efficiency* (time participants took to find the correct answers), and the eye-tracking metrics *fixation frequency*, *average fixation duration* and *scanpath speed*. First, overall performance differences between stimuli based on colour distances and font sizes were evaluated (regardless of gender or expertise). Following this, group differences between females/males and experts/novices were examined.

Fixation detection was performed through the SMI BeGaze software (SensoMotoric Instruments, 2013b) using ID-T (dispersion threshold algorithm). Dispersion threshold was set to 50 pixels and a minimum length of 80 ms-based pilot experiments[4]. Basic eye-tracking metrics has been calculated in the open source software OGAMA (Voßkühler, 2013) and statistical analysis of the measured data has been performed using the R software (R Core Team, 2013). According to Goldberg and Kotval (1999), more overall fixations and longer scanpaths (the length of gaze trajectory over the stimulus) could indicate a less efficient search process. Longer fixation duration could mean difficulty in extracting information, or the object is more engaging/relevant for the task in some way (Eastman, 1985; Poole and Ball, 2005). Time to answer (response time/task completion time) reflects the performance and success during information search. Absolute values of eye-tracking metrics (fixation count and scanpath length) positively correlates with the time to answer, thus we examined proportional fixation frequency (count per second) and scanpath speed (pixels per second), because these metrics are independent of time.

We conducted statistical analysis on the measured dependent variables. Prior to inferential statistics, we tested the data for normality with Shapiro-Wilk test (Shapiro and Wilk, 1965). Results showed that on the significance level $\alpha=0.05$ none of measurements were normally distributed: fixation frequency $W=0.96$, $p=6.94\times10^{-12}$; average fixation duration $W=0.93$, $p<2.2\times10^{-16}$; scanpath speed $W=0.99$, $p=0.03$; time to answer $W=0.82$, $p<2.2\times10^{-16}$. Because none of the data are normally distributed, data does not meet

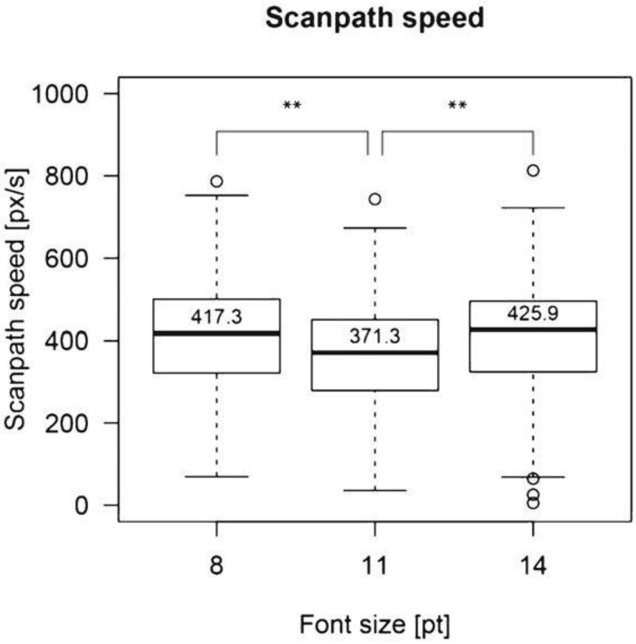

Figure 4. Boxplot showing the scanpath speed values observed for the three different font sizes. Median scanpath speed for font size 11 pt is significantly lower than for the others. Exact median values are given inside boxplots[5]

the parametric assumption, and therefore, only non-parametric statistical tests were used in the analysis.

RESULTS AND DISCUSSION

All participants were able to find the correct answer within the given time limit of 30 seconds; thus, the accuracy of answers was not evaluated further.

Font size and colour distance

Font size examination was conducted as an overall statistical analysis between the tested font sizes, regardless of the colour distance and user backgrounds. For this purpose, the visual search efficiency (time to answer) and visual search strategies (eye-tracking metrics) for the three groups of stimuli with three types of font size were compared by applying the Kruskal–Wallis test (Kruskal and Wallis, 1952). Only the scanpath speed metric (H= 17.56, p=0.00) yields statistically significant results. No statistically significant results were observed for other metrics (Table 2).

Using the *post-hoc* Kruskal–Wallis we observed that the participants had significantly lower scanpath speeds (less pixels viewed per second) while searching for the answer on stimuli with font size 11 pt (*Mdn*=371.3 px/s, Figure 4). No significant differences were observed between stimuli with font sizes 8 pt (*Mdn*=417.3 px/s) and 14 pt (*Mdn*=425.9 px/s). While our overall findings suggested that the tested font sizes perform somewhat similarly in terms of visual search performance and confirm the results of Bartz (1970), at this point, we found that the font size 11 facilitated a more decisive search compared to the others. A lower scanpath speed means participants can find the information they look for without having to examine as much information as in other conditions.

In the next step, we analysed the main effects for the colour distance (regardless the font size and user groups) using Kruskal–Wallis test. A significant result was observed for the performance metric *time to answer* (H=17.05, DF=4, P=0.00, Table 2). *Post-hoc* Kruskal–Wallis tests reveal that the observed difference refers to stimuli with the lowest colour distance (ΔE_{00}=30, *Mdn*=4.5 seconds) and two highest colour distances (ΔE_{00}=85 with *Mdn*=3.5 seconds and ΔE_{00}=100 with *Mdn*=3.6 seconds). No other significant differences were observed in the middle steps (Figure 5).

To further explore the influence of colour distance, we analysed its interactions with the different font sizes. Kruskal–Wallis test was performed separately for groups of stimuli with the same font size. Significant results were observed in all three cases (8 pt, 11 pt and 14 pt) for time to answer metric (Table 2). In the case of the medium and largest font size the scanpath speed metric appears to be significant as well. *Post-hoc* Kruskal–Wallis tests revealed further differences among various combinations of stimuli with different font sizes and colour distances (Figure 6).

At this point, to study the attention that users paid to the task-relevant areas versus task-irrelevant areas on the display for our tested conditions, we performed an analysis of gaze data in specified AOIs. To counter the possible registration issues with the eye tracking, we constructed regions around each label to include all gaze points that are closer to that label than to any other, similarly to the idea of Voronoi diagrams.

Based on the AOI analysis, we observe a vast difference between dwell times on the task-relevant AOIs (where the correct answer is) and irrelevant AOIs (other labels on stimuli). In general participants need to spend only a short time to identify that a label is not the correct one (M=134.1 ms), while to find and mark the correct AOI they spend considerably more time (M=1286.2 ms).

Table 2. Results of Kruskal–Wallis test to compare maps varying in font size, colour distance and both variables

		Fixation frequency		Average fixation duration		Scanpath speed		Time to answer	
source	df	H	p	H	p	H	p	H	p
font size	2	4.81	0.09	2.38	0.30	17.56	**0.00**	5.05	0.07
colour distance	4	0.00	0.94	2.21	0.69	4.55	0.33	17.05	**0.00**
colour distance ~ font size 8	4	4.02	0.40	3.17	0.52	8.25	0.08	17.70	**0.00**
colour distance ~ font size 11	4	2.17	0.70	3.96	0.41	12.51	**0.01**	24.64	**0.00**
colour distance ~ font size 14	4	4.80	0.30	2.74	0.60	23.25	**0.00**	12.11	**0.01**

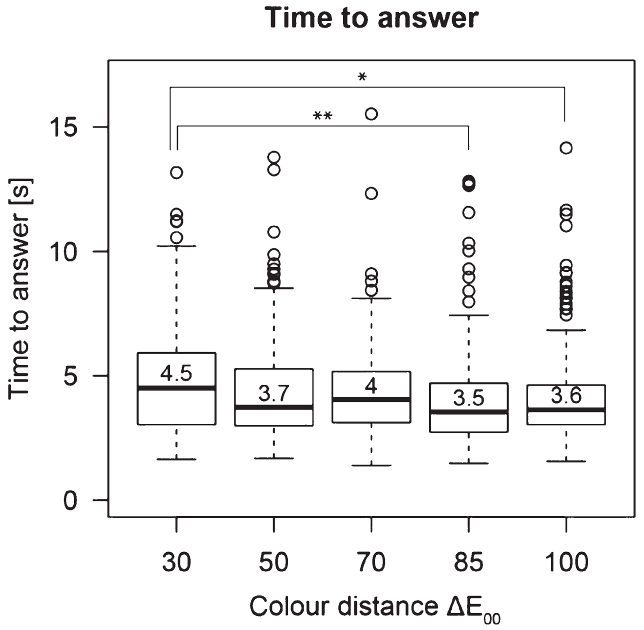

Figure 5. Box plots showing time to answer values observed for stimuli with different colour distances. Exact median values are given inside boxplots. A general trend can be observed that the larger the colour distance the less time people use to identify the labels

Analysing the first fixations on the ‘correct AOI’; our observations confirmed that when there is a larger colour distance between the label and the background, people need less time to find the correct answer. This decreasing trend is illustrated in Figure 7. Kruskal–Wallist test proved significant differences (H=25.84, DF=4, p=0.00) between ΔE_{00}=30 and ΔE_{00}=85, ΔE_{00}=30 and ΔE_{00}=100 and also ΔE_{00}=70 and ΔE_{00}=100. This corresponds squarely with our previous finding about time to answer and further validates our findings about the colour distance.

CONCLUSIONS AND OUTLOOK

In this manuscript, our main contribution is an empirical user study to analyse the influence of colour distance (between the colour of the map labels and their background) and font size on map readability.

All participants were successful in locating the searched items in all conditions within the time limits (i.e. accuracy was 100%). This level of accuracy suggests that conditions we provided are essentially usable – our smallest ΔE is much higher than the so-called ‘just noticeable difference’ (Linhares *et al.*, 2008), and the smallest font size we used (8 pt) is used in certain map scales. However, we investigated finer differences based on other metrics. Analysing the *time to answer* for the five colour distances empirically validates what is common sense; the relative differences in time to answer become statistically significant as the colour distances grow larger (30–85 and 30–100). More precisely, observed median value of time to answer on the stimuli with ΔE_{00}=30 was Mdn=4.5 seconds. Median time on ΔE_{00}=85 was Mdn=3.5 seconds and on ΔE_{00}=100 was Mdn=3.6 seconds. These results were further validated in the following AOI analysis. These findings mean that even if a search task is successfully completed, a map user will lose time with a combination of a dark background with a dark (in this case black) label. In certain map use cases, this can be critically important (e.g. emergency and rescue, driving) in time pressure situations or simply frustrating.

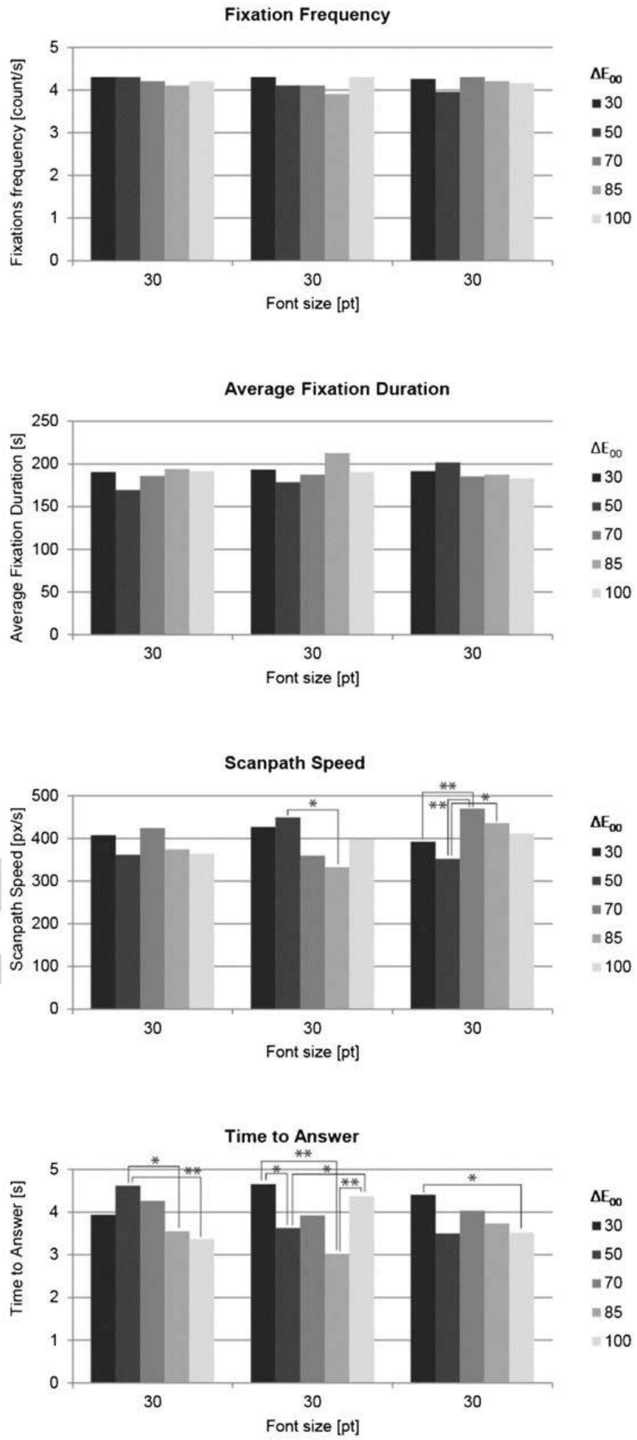

Figure 6. Charts show median values of observed metrics separated for groups of stimuli with different label size (8, 11 or 14 pt) and colour distance between labels and background (ΔE_{00}=30, 50, 70, 85 or 100)

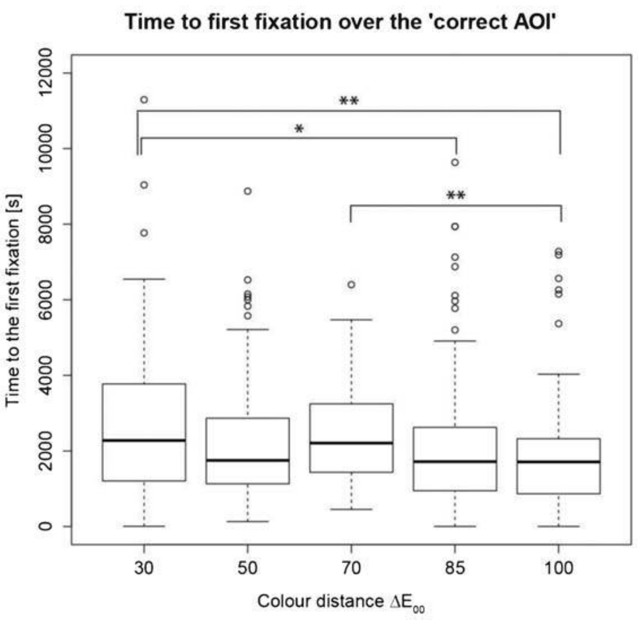

Figure 7. Boxplot shows time to first recorded fixation over the AOI covering the correct answer observed for five groups of stimuli with different colour distance

The comparison of the font size based on time metrics suggests that the medium font size (11 pt) is the most efficient. This result is mainly based on the eye movement metric scanpath speed showing us the number of pixels covered per second as the participants move their eyes on the display. In this case, arguably, we interpret that covering less pixels per second indicates some level of certainty. Having a better result for a medium size conforms to the previous knowledge – it has been demonstrated that too little or too much information can influence decision performance (e.g. Çöltekin and Rechenbacher, 2011). Furthermore, we studied to see if we could observe an interaction between font size and colour distance, and our current results suggest that the performance of people based on size and colour do not strongly interact. Nonetheless, these initial results about font sizes warrant further tests to obtain a rule for font sizes especially in relation to colour distance.

To **summarize** our results; the tested colour distances essentially should not impair basic success for readability of map labels, but it will slow people down, i.e., larger colour distances yield consistently better results in terms of 'time to answer'. This paper documents empirically (based on a user study and in a quantified manner) that a ΔE_{00} of 70 (difference between $\Delta E_{00}=30$ and $\Delta E_{00}=100$) will change the speed of label identification consistently. Among the three font sizes, we tested the middle size (font size 11) allows a better experience in visual search with this combination of colours according to our interpretation of the scanpath speed metric.

These results, insights and observations allow us to build new hypotheses and the next steps involve in further experiments to test the effective limits of colour distance until we understand thresholds for colour discrimination, and visual behaviour during the colour discrimination process in cartographic tasks. Future experiments will take more factors into account that influence the colour perception including the effect of spatial distance, and eventually surrounding colours. The combined results from this experiment and follow-up experiments should allow us to establish thresholds that can be used as guidelines in map design and help us all decide which colour distances can be applied in maps with no perceptual concerns for discernibility. Additionally, we intend to explore group differences for visual search strategies, possibly contributing how we can guide the users to create personalized maps for themselves or designers to create maps that work better for the target groups. In the long-term, we also aspire to contribute to developing educational strategies that work for all groups.

BIOGRAPHICAL NOTES

Alžběta Brychtová is currently a PhD student at Department of Geoinformatics, Faculty of Science, Palacký University in Olomouc, Czech Republic. She has a degree in Geography and Geoinformatics from the same university. Her professional skills and interests in cartography are cognitive aspects, usability studies and eye-tracking which are fully reflected in her dissertation thesis on the evaluation of colour distance influence on the map legibility.

ACKNOWLEDGEMENTS

Authors would like to thank Veronika Obadálková, student of Geoinformatics at Palacký University, Olomouc, who helped in the process of collecting eye-tracking data and all our participants. This article has been completed within the project IGA_PrF_2014007 of Palacký University, Olomouc.

NOTES

1. Listed names of colour models are based on abbreviations of their components: RGB=red, green, blue; HSV=hue, saturation, value; HSB=hue, saturation, brightness; HSL=hue, saturation, lightness.
2. http://colorbrewer2.org/
3. 'Within-subject', in experimental design, means that all participants were exposed to all tested conditions (Rubin and Chisnell, 2008).
4. Pilot experiments were conducted by Popelka (2014). There are no golden rules about these thresholds so far, the values we selected are within the window of values that are used in previous eye movement studies, e.g. see Popelka and Brychtová (2013) or Russo *et al.* (2014).
5. We differentiate between varying confidence levels of 0.01 and 0.05 with a notation of a two asterisks (**) and single asterisk (*), respectively. This notation will be used throughout the paper.

REFERENCES

Bartz, B. (1970). 'An analysis of the typographic legibility literature: assessment of its applicability to cartography', **The Cartographic Journal**, 7, p. 7.

Bertin, J. (2010). **Semiology of Graphics: Diagrams, Networks, Maps**, 1st ed., ESRI Press, Redlands, CA.

Bjorke, J. T. (1996). 'Framework for entropy-based map evaluation', **Cartography and Geographic Information Systems**, 23, pp. 78–95.

Brewer, C. A., Hatchard, G. W. and Harrower, M. A. (2003). 'ColorBrewer in print: a catalog of color schemes for maps', **Cartography and Geographic Information Society**, 30, pp. 5–32.

Buard, E. and Ruas, A. (2009). 'Processes for Improving the Colours of Topographic Maps in the Context of Map-on-demand', in **24th International Cartographic Conference**, p. 11, Santiago, Nov 15–21.

Carter, R. and Huertas, R. (2009). 'Ultra-large color difference and small subtense', **Color Research & Application**, 35, pp. 4–17.

CIE – International Commission on Illumination (2012). Termlist of International Commission on Illumination. Available at: http://eilv.cie.co.at/ [Accessed 20 February 2014].

Çöltekin, A. and Reichenbacher, T. (2011) 'High quality geographic services and bandwidth limitations', **Future Internet**, 3, pp. 379–396.

Çöltekin, A., Heil, B., Simone Garlandini, S. and Fabrikant, S. I. (2009). 'Evaluating the effectiveness of interactive map interface designs: a case study integrating usability metrics with eye-movement analysis', **International Journal of Geographical Information Science**, 36, pp. 5–17.

Chesneau, E. (2007). 'Improvement of Colour Contrasts in Maps: Application to Risk Maps', in **10th AGILE International Conference on Geographic Information Science 2007**, pp. 1–14, Aalborg, May 8–11.

Culp, G. M. (2012). 'Increasing accessibility for map readers with acquired and inherited colour vision deficiencies: a re-colouring algorithm for maps', **The Cartographic Journal**, 49, pp. 302–311.

Deeb, R., de Maeyer, P. and Ooms, K. (2011). 'The Influence of Jaques Bertin', in **25th International Cartographic Conference**, p. 8, Paris, Jul 3–8.

Deeb, R., Ooms, K. and de Maeyer, P. (2012). 'Typography in the eyes of Bertin, gender and expertise variation', **The Cartographic Journal**, 49, pp. 176–185.

Dent, B. D., Torguson, J. S. and Hodler, T. W. (2009). **Cartography: Thematic Map Design**, 6th ed., Thomas Timp, New York.

Doddi, S., Marathe, M. V., Mirzaian, A, Moret, B. M. E. and Zhu, B. (1997). 'Map Labeling and Its Generalizations', in **8th Annual ACM-SIAM Symposium on Discrete Algorithms**. pp. 148–157, New Orleans, LA, Jan 7–9.

Eastman, J. R. (1985). 'Cognitive models and cartographic design research', **The Cartographic Journal**, 22, pp. 95–101.

Garlandini, S. and Fabrikant, S. I. (2009). 'Evaluating the effectiveness and efficiency of visual variables for geographic information visualization', in **COSIT 2009**, ed. by S. K. Hornsby, pp. 195–211, Springer-Verlag, Berlin/Heidelberg.

Gartner, G. and Hiller, W. (2008). 'Size Matters: Correlation of Limited Presentation Media Size and the Degradation of Cognitive Representations of Space', in **3rd International Joint Workshop on Ubiquitous, Pervasive and Internet Mapping (UPIMap) 2008**, Shepherdstown, WV, Sep 10–11.

Gegenfurtner, K. R. and Sharpe, L. T. (Eds.). (2001). **Color Vision: From Genes to Perception**, Cambridge University Press, Cambridge.

Gilmartin, P. and Shelton, E. (1989). 'Choropleth maps on high resolution crts: the effest of number of classes and hue on communication', **Cartographica: The International Journal for Geographic Information and Geovisualization**, 26, pp. 40–52.

Goldberg, J. H. and Kotval, X. P. (1999). 'Computer interface evaluation using eye movements: methods and constructs', **International Journal of Industrial Ergonomics**, 24, pp. 631–645.

Harrie, L. and Stigmar, H. (2009). 'An evaluation of measures for quantifying map information', **ISPRS Journal of Photogrammetry and Remote Sensing**, 65, pp. 266–274.

Harrower, M. A. and Brewer, C. A. (2003). 'ColorBrewer.org: an online tool for selecting colour schemes for maps', **The Cartographic Journal**, 40, pp. 27–37.

Hegarty, M., Smallman, H. S., Stull, A. T. and Matt, S. (2009). 'Naïve cartography: how intuitions about display configuration can hurt performance', **Cartographica: The International Journal for Geographic Information and Geovisualization**, 44, pp. 171–186.

Holmqvist, K. et al. (2011). **Eye Tracking: A Comprehensive Guide to Methods and Measures**, 1st ed., Oxford University Press, Oxford.

ISO 9241-11. (1998). 'Ergonomic requirements for office work with visual display terminals (VDTs) – Part 11: Guidance on usability', ISO, Geneva, Switzerland, p. 22.

Jenny, B. and Kelso, N. V. (2007). 'Designing maps for the colour-vision impaired', **Bulletin of the Society of Cartographers SoC**, 41, pp. 9–12.

Kimerling, J. A. (1985). 'The comparison of equal-value gray scales', **The American Cartographer**, 12, pp. 132–142.

Kröger, J., Schiewe, J. and Weninger, B. (2013). 'Analysis and Improvement of the Open- StreetMap Street Color Scheme for Users with Color Vision Deficiencies', in **26th International Cartographic Conference**, p. 17, Dresden, Aug 25–30.

Kruskal, W. H. and Wallis, W. A. (1952). 'Use of ranks in one-criterion variance analysis', **Journal of the American Statistical Association**, 47, pp. 583–621.

Landa, E. R. and Fairchild, M. D. (2005). 'Charting color from the eye of the beholder', **American Scientist**, 93, pp. 436–443.

Lindbloom, J. M. M., Pinto, P. D. and Nascimento, S. M. C. (2008). The number of discernible colors in natural scenes. **Journal of the Optical Society of America**, 25(12), pp. 2918–24.

Linhares, B. J. (2012). Useful Color Calculators and Spreadsheets. http://www.brucelindbloom.com/ (accessed 12 February 2013).

MacEachren, A. M. (2004). **How Maps Work, Representation, Visualization, and Design**, 1st ed., The Guilford Press, New York.

Olson, J. M. and Brewer, C. A. (1997). 'An evaluation of color selections to accommodate map users with color-vision impairments', **Annals of the Association of American Geographers**, 87, pp. 103–134.

Ooms, K., de Maeyer, P., Fack, V., van Assche, E. and Witlox, F. (2012). 'Investigating the effectiveness of an efficient label placement method using eye movement data', **The Cartographic Journal**, 49, pp. 234–246.

Polishchuk, V. and Vihavainen, A. (2010). 'Periodic multi-labeling of public transit lines', **Lecture Notes in Computer Science**, Vol. 6292, pp. 175–188.

Poole, A. and Ball, L. J. (2005). 'Eye tracking in human–computer interaction and usability research: current status and future', in **Encyclopedia of Human–Computer Interaction**, p. 13, Idea Group Reference, Hershey, PA.

Popelka, S. and Brychtová, A. (2013). 'Eye-tracking study on different perception of 2D and 3D terrain visualisation', **The Cartographic Journal**, 50, pp. 240–246.

Popelka, S. (2014). 'Optimal eye fixation detection settings for cartographic purposes', in **SGEM 2014 Informatics, Geoinformatics and Remote Sensing Proceedings**, Vol. 1, I STEF92 Technology Ltd., Sofia, Bulgaria. pp. 705–712.

R Core Team. (2013). R: A Language and Environment for Statistical Computing, http://www.r-project.org/

Robinson, A. H. (1952). **The Look of Maps: An Examination of Cartographic Design**, ESRI Press, Redlands, CA.

Rubin, J. and Chisnell, D. (2008). **Handbook of Usability Testing: Howto Plan, Design, and Conduct Effective Tests**, 2nd ed., p. 384, Willey, New York.

Russo, P., Pettit, C., Coltekin, A., Imhof, M., Cox, M. and Bayliss, C. (2014). 'Understanding soil acidification process using animation and text: an empirical user evaluation with eye tracking', in **Cartography from Pole to Pole: Lecture Notes in Geoinformation and Cartography**, ed. by Buchroithner, M., Prechtel, N. and Burghardt, D., pp. 431–448, Springer, Berlin/Heidelberg.

Salvucci, D. D. and Goldberg, J. H. (2000). 'Identifying fixations and saccades in eye-tracking protocols', in **Proceedings of the symposium on Eye tracking research and applications ETRA '00**, pp. 71–78, ACM Press, New York.

Schnüer, R. Sieber, R. and Çöltekin, A. (2014). The Next Generation of Atlas User Interfaces: A User Study with Digital Natives. In A. Vondáková, J. Brus, & V. Voženílek (Eds.), *Modern Trends in Cartography: Selected Papers of CARTOCON 2014 (Lecture No., pp. 23–36)*. Springer.

SensoMotoric Instruments. (2013a). SMI BeGaze™, http://www.smivision.com/en/gaze-and-eye-tracking-systems/products/begaze-analysis-software.html.

SensoMotoric Instruments. (2013b). SMI Experiment Center™, http://www.smivision.com/en/gaze-and-eye-tracking-systems/products/experiment-center-software.html.

Sharma, G., Wu, W. and Dalal, E. E. (2005). The CIEDE2000 color-difference formula: implemen– tation notes, supplementary test data, and mathematical observations. **Color Research & Application**, 30(1), pp. 21–30

Shiewe, S. and Weninger, B. (2013). 'Visual encoding of acoustic parameters – framework and application to noise mapping', **The Cartographic Journal**, 50, pp. 332–344.

Shapiro, S. S. and Wilk, M. B. (1965). 'An analysis of variance test for normality (complete samples)', **Biometrika**, 52, pp. 591–611.

Slocum, T. A., McMaster, R. B., Kessler, F. C. and Howard, H. H. (1999). **Thematic Cartography and Geovisualization**, 1st ed., Prentice Hall, Upper Saddle River, NJ.

Steinrücken, J. and Plümer, L. (2013). 'Identification of optimal colours for maps from the web', **The Cartographic Journal**, 50, pp. 19–32.

Stigmar, H. (2010). **Making 21st Century Maps Legible – Methods for Measuring and Improving the Legibility and Usability of Real-Time Maps**, Lund University, Lund .

van Dijk, S. F. (2001). **Genetic algorithms for map labeling**, Utrecht University, Utrecht.

Vondráková, A., Vávra, A. and Voženílek, V. (2012). 'Climatic regions of the Czech Republic', **Journal of Maps**, 9, pp. 425–430.

Voßkühler, A. (2013). OGAMA (OpenGazeAndMouseAnalyzer) http://www.ogama.net/.

Wagner, F. and Wolf, A. (1995). 'Map Labeling Heuristics: Provably Good and Practically Useful', in **11th Annual Symposium on Computational Geometry**, pp. 109–118, Vancouver, BC, Jun 5–7.

Werman, M. (2012). 'Improving perceptual color difference using basic color terms', **Computer Research Repository**, pp. 1–14.