Efficiency, Task Dependency and User Preferences of Three Declutter Methods for Mobile Maps

Rosemarijn Looije
TNO Defence and Security
P.O. Box 23, 3769 ZG Soesterberg, the Netherlands
+31 (0)346 356 370
Rosemarijn.looije@tno.nl

Guido M. te Brake
TNO Defence and Security
P.O. Box 23, 3769 ZG Soesterberg, the Netherlands
+31 (0)346 356 253
Guido.tebrake@tno.nl

Mark A. Neerincx
TNO Defence and Security
P.O. Box 23, 3769 ZG Soesterberg, the Netherlands
+31 (0)346 356 298
Mark.neerincx@tno.nl

ABSTRACT
Map-based applications such as Google maps and navigation software are often used on mobile devices. These applications help users find directions and provide different types of location-based information. The usability of map-based applications on mobile devices is a critical factor for adoption of these applications by the general public. One of the main problems with displaying information on the small screen of a mobile device is that the display can become cluttered with icons and text. This paper presents three declutter methods and a user study on their efficiency, effectiveness, task dependency, and user satisfaction. Results show that for a search and a navigate task, decluttering by spreading the icons to minimize overlap at the cost of exact positioning was the preferred approach. An aggregate method, with compound icons, was preferred for a locate and an identify task. Interestingly, the method preferred by the users was not always the most effective option.

Categories and Subject Descriptors
H.1.2 [**Information Systems**]: User/Machine Systems – *Human Factors.*

H.5.2 [**Information Systems**]: User Interfaces –*Screen Design.*

J.7 [**Computer Applications**]: Computers in Other Systems – *Consumer Products.*

General Terms
Measurement, Performance, Design, Experimentation, Human Factors

Keywords
User study, usability, geospatial systems, mobile devices, display clutter.

1. INTRODUCTION
Map-based applications such as Google Earth and Yahoo Maps are frequently used on mobile devices. Using these applications, consumers find directions and look for information about hotels, restaurants, shops, monuments, events, etcetera. Usability is a key prerequisite for adoption and regular usage of such services. However, the increase of information causes clutterd displays in which it can be difficult to find the relevant information. So, we need to find effective ways to prevent information clutter. This is complex, because mobile devices have many restrictions in comparison with desktop computers, such as small screen sizes, clumsy input methods, slow CPU speed, and slow wireless data connections (Chittaro, 2006).

This study focuses on dealing with many closely positioned similar objects, or objects overlapping each other (sometimes called 'icon collisions'). Various approaches are in use, such as spreading out the icons over a larger area to minimize occlusion by other icons (Fuchs & Schumann, 2004) or combining several icons in one group icon (Edwardes, Burghardt, & Weibel, 2005). However, little research has been done to compare the different approaches. Therefore, it is unclear which of the existing methods will present the user the most usable visualization. Which method is best may also depend on the task the user is conducting. To find answers for these questions, we compared three declutter approaches on several different tasks. The main research questions were which method works best, is declutter effectiveness task dependent, and which approach is preferred by the users.

Section 2 presents a review of recent literature on clutter and overlapping icons. Section 3, describes a study in which three different declutter methods were evaluated. Section 4 and 5 contain respectively the results of the user study and the discussion/conclusion.

2. Background

2.1 Defining and measuring clutter
According to the definition of Rosenholtz, Li & Nakano (2007), clutter is the state in which excess items, or their representation or organization, lead to a degradation of performance at some task. From this definition, it follows that whether a display is cluttered depends on the goals and tasks of the user. Excess and/or disorganized display items can cause crowding, masking, decreased recognition performance due to occlusion, greater

difficulty at both segmenting a scene and performing visual search, and so on.

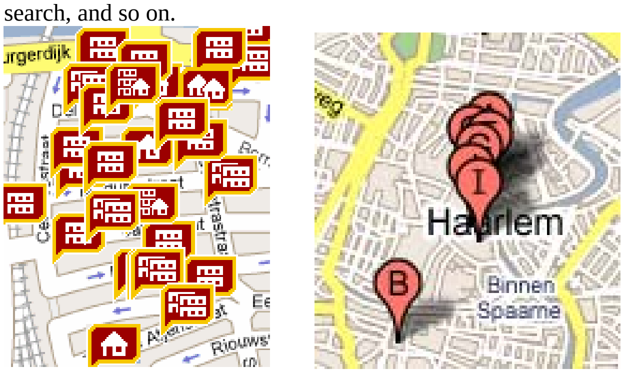

Figure 1: Cluttered maps of Funda (real estate program) on the left (Funda, 2008) and Google maps (Google maps, 2008). The funda picture shows next to house and apartment icons, aggregated icons of multiple apartments, multiple houses and of multiple apartments and houses at the same location.

Given a reliable measure of the visual clutter in a display, designers could be supported in choosing between different options of display design. A clutter measure can show how much clutter disappears when removing information from the screen and the designer can judge the necessity of the removed information. Of course, task and user differences are also of large influence on the experienced clutter and these are hard to measure. Rosenholtz et al (2007) present and compare several measures of visual clutter, which operate on arbitrary images as input. The first is a new version of the Feature Congestion measure of visual clutter (Rosenholtz & Mansfield, 2005). This measure is based on the analogy that the more cluttered a display or scene is, the more difficult it would be to add a new item that would reliably draw attention. A second measure of visual clutter, Subband Entropy (Rosenholtz et al, 2007), is based on the notion that clutter is related to the amount of visual information in the display. When similar objects are grouped the visual information is less than when there are many dissimilar objects. The third measure, Edge Density (Mack & Olivia, 2004) is used as a measure of subjective visual complexity. Edge density looks at the percentage of pixels that are edge pixels of objects, more edges means a more cluttered display. Rosenholtz has shown that all three measures of clutter are good predictors of response times for a visual search tasks. They could be used to replace set-size, the number of segmentable regions, a variable often found to have high correlation with response times for visual search tasks in artificial settings, but hard to define in many real-world (especially photographic) images.

The level of clutter is experienced very differently by individual persons, correlations of r=.61 (Mack & Olivia, 2004) and r=.70 (Rosenholtz & Mansfield, 2005) are found in experiments. They asked participants to rank 25 images of maps from the U.S. and San Francisco area according to how cluttered they found the image, they did not receive a definition of clutter before they started. The individual differences can depend on, amongst others, spatial abilities and expertise. A geologist will find contour lines on a map less cluttering then someone who is not familiar with them (Phillips & Noyes, 1982). In the same way some people prefer more information than others to build up a mental model of an area (Rayson, 1999).

2.2 Dealing with clutter

Although experienced clutter may depend on task requirements and user preferences (Rayson, 1999), there are some general guidelines to reduce clutter for visual search tasks. One should focus on limiting the use of symbols that resemble each other in, amongst others, shape, size, or color (lines clutter lines, points clutter points). It is also important to avoid symbols that resemble letters in front of a word, for example an open round before a city name can increase search time significantly (Phillips & Noyes, 1982). Lee, Forlizzi, and Hudson (Lee, Forlizzi, & Hudson, 2007) conducted a study to order different pop-out effects. Pop-out is a bottom-up way of drawing attention to an object through perception, in contrast to top-down where the cognition directs the attention. Prior studies have identified a range of visual features which can induce pop-out effects effect (Julesz, 1984; Treisman, 1986), such as color, brightness, animation, et cetera. Notably, size has been shown not to induce a pop-out effect (Baldassi & Burr, 2004). Other research has also shown that people are good in distinguishing colors. Phillips and Noyes (1982) coded 16 symbols either by color, texture, or a non-redundant combination of both. The colored symbols, with or without texture, were much easier to find than symbols coded by texture alone. A problem with the use of colors is that changes in brightness of a display can have a disastrous effect on the ability to distinguish different colors; furthermore it is dependent on the ability of people to distinguish colors (colorblindness). One has to keep in mind that pop-out effects have to be used sparsely, because otherwise the effect will be lost.

A more advanced approach is generalization, defined as the process of reducing detail on a map as a consequence of reducing the map scale (Blinn, Queen, & Maki, 2008). Map generalization work focuses on changing representations of information at different map scales to reduce clutter and improve usability.

If it is known which information is not essential to task execution, this information could be removed from the display (Wickens, 2000). This turning on and off of specific layers may be done automatically if a reliable task model is available, but generally it is the user's responsibility to turn-off unnecessary information layers. However, if for a task information is required from various information domains, turning layers off may not be an option. Also, interaction with the system to select or deselect layers comes at a penalty in effort and time required (Yeh & Wickens, 2000).

Maps with high information density can suffer from overlapping icons, sometimes called 'icon stacking' or 'icon collisions'. Several solutions can be found in modern software applications, of which applying semi-transparency is the most common (e.g. {Kamba}. With semi-transparency overlapping icons have different levels of transparency which makes that even though an icon might be overlapped by another icon it is still partially visible. Alternatives are the grouping of several icons of the same type in a single aggregate icon, or spreading the overlapping icons such that overlap is prevented. However, little research has been done on the usability of these alternatives. Fuchs and Schumann (Fuchs & Schumann, 2004) present an approach for intelligent icon placement and conflict resolution that can be used in real-time interactive systems, but little other research has been done. In this study we will focus on reducing the overlap between icons (aggregate and spread) and will not consider reducing overlap by transparency, mainly because transparency can not be implemented with the .Net Compact framework.

In existing applications, several overlap reducing methods are in use. Three examples are:

1. Aggregate, combining several overlapping objects of the same type in a single large object (Figure 2a). When the object is selected, properties or names of the various included objects can be shown. When zoomed in, the aggregate object may fall apart in multiple separated objects.
2. Spread, shifting (partly) occluded objects until no longer serious overlap is present. Objects are not depicted anymore on their exact location, but for some tasks this may not be a problem. This method is only suitable if several objects are very close together (such as shops in a mall or a string of bars) in a sparse context (Figure 2b).
3. Aggregate-spread, combines both spread and aggregate. A large icon is placed nearby the overlapping objects, using the same algorithm as for aggregate. The original objects are represented with small icons presenting the exact location (Figure 2c).

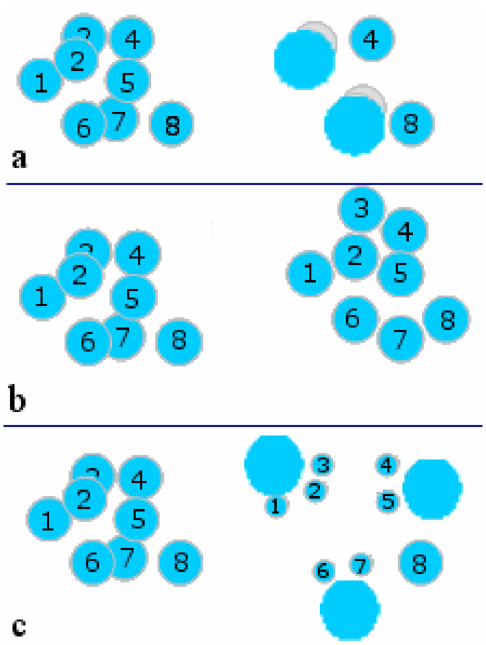

Figure 2: Three different declutter methods: aggregate (a), spread (b) and aggregate spread (icon 8 is not aggregated) (c)

3. Method

3.1 Design

Based on the literature review two hypotheses can be formulated:

1. Which declutter approach is best is task dependent.
2. Which declutter approach is best is user dependent.

In our experiment users had to conduct tasks in four different conditions: using the original cluttered display (NM), using the spread-aggregate method (AGS), using the aggregate method (AG) and using the spread method (S).

To test the hypothesis that a dependency exists between task and optimal declutter method, the participants had to perform four different tasks for each condition. These tasks are chosen, because they address different aspects of using maps (Reichenbacher, 2003) potentially benefiting different declutter methods.

1. A search task: Find the nearest X (shop, parking or restaurant) and find out on which street it is. For the search task a spread display is expected to be most usable, because all icons are visible and the "you are here" icon is easy to locate. Zooming is not necessary to see exact (cluttered) or near exact (spread) locations as it is with the aggregate methods.
2. A locate task: In which street is X (shop, parking or restaurant) with number Q. For the locate task the aggregate methods is likely to be the most usable, because by clicking on an aggregated icon, information about multiple icons can be viewed at once.
3. An identification task: Find an X (shop, parking or restaurant) with label Y and find out which number it has. For the identify task aggregate methods will usable, because by clicking on an aggregated icon information about multiple icons can be viewed.
4. A navigation task: You are in parking with number Q, and must plan the shortest circuit of three shops and a restaurant. For the navigate task the spread method is the favorite, because all icons are visible at the same time and, in most cases, there is no need for exact locations.

All participants conducted the four tasks with the four different conditions, making a total of 16 task assignments. The order of the conditions was balanced between the participants (see Table 1**)**. In each condition, the order of the tasks was the same for every participant, because it was built up according to increasing difficulty (search, locate, identify and navigate).

Table 1 Experimental design

Pp 1/5	NM	AGS	AG	S
Pp 2/6	AGS	S	NM	AG
Pp 3/7	AG	NM	S	AGS
Pp 4/8	S	AG	AGS	NM

3.2 Materials

A map-based software application has been developed in the .Net Compact framework using Visual Studio .Net 2005 to experiment with various declutter methods. In this framework applications for Windows Mobile 5 systems can be developed. The environment provides an emulator for debugging the program. The GUI is written in C#.

The user is provided with a PDA (Dell x51) presenting a map of a city as is shown in Figure 3. At the bottom right, three information layers can be selected: shops, parking, and restaurants. All the items on the layers are numbered, as is common in many map-based applications (see Google in Figure 1). Furthermore, they have an information label associated with them dependent on the icon type: type of shop, the parking price per hour, and type of cuisine. When an object is selected, the number and information label are shown in a text box in the left bottom corner of the screen. Next to this text box, three zoom

levels (using three screenshots from Google maps) can be selected using a small slider. The green 'End' option was added for experimental purposes. Participants interacted with a stylus on the PDA touch screen; use of the PDA's buttons was not required. Each condition had 3 zoom-levels and 3 information levels.
Three different declutter methods were implemented as follows:

1. The aggregate method uses an algorithm that checks if two icons of the same type are less than an icon size apart from each other, thus overlap each other. When this is the case the icons are aggregated and the rest of the icons are checked and compared with the first icon. Once an aggregated object is clicked the numbers of the objects that are contained in the aggregated object are listed in the textbox instead of the labels of an object. The labels can be accessed by clicking one of the object numbers in the list. Thus, when a Swedish (nr. 1) and a Chinese (nr. 2) restaurant are aggregated the textbox shows only 1, 2 instead of 1. Swedish, 2. Chinese.

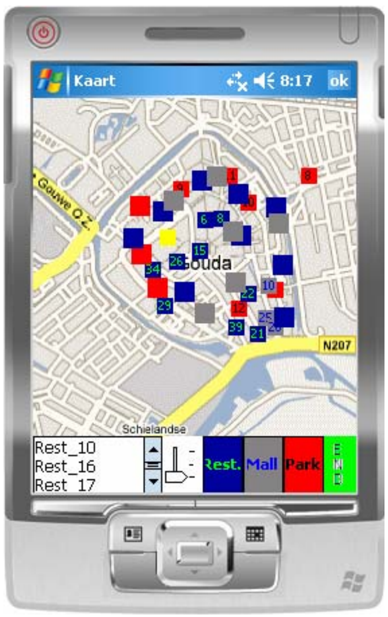

Figure 3: PDA provided to the users showing a cluttered display with overlapping objects

2. Because spread is a rather complicated algorithm to program, we positioned the icons by hand at appropriate locations. These positions were reloaded when the spread method was used.
3. Aggregate-spread uses basically the same algorithm as aggregate method. Labels can be directly accessed by clicking a small icon or indirectly by first clicking an aggregated icon and then accessing the labels by clicking on the objects number in the list.

The experimental conditions are shown in Figure 4, in which a small part of the display is shown. The display always contained an extra icon that depicted the location of the user ("you are here icon"), which was at a different position in each task.

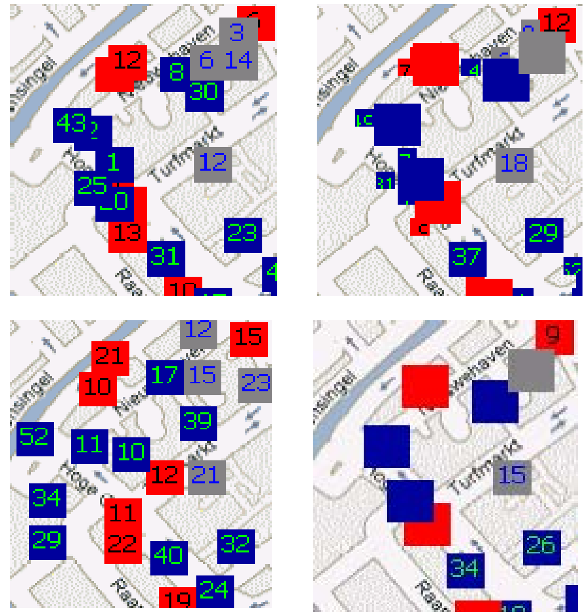

Figure 4: The four different conditions (clockwise, starting in the left upper corner): 1) Original image, 2) aggregate-spread, 3) aggregate, 4) spread. The numbers vary in each picture because for experimental reasons they were assigned differently in each condition.

3.3 Participants

Eight participants took part in the experiment, three male and five female (Mean age = 26.75, SD = 2.3). All participants had computer and mobile phone experience. Four of them had reasonable or much experience with touch-screens, the other four had little experience with touch-screens.

3.4 Measures

Objective measures gathered in this study were:

- Time: The total time participants needed to complete a task (efficiency).
- Number of clicks: How often does the participant; change the zoom-level, pan, click icons, turn information layers off and on (efficiency and strategy).
- Errors: Did the participant find a correct solution? (effectiveness).

Subjective measures regarding user preferences and satisfaction were gathered using questionnaires.

3.5 Procedure

Before the experiment, participants were asked about age, gender, education, computer experience, mobile experience, and touch-screen experience. They then received a short introduction and were asked to perform some simple tasks such as zooming and panning to get acquainted with the system.
Based on the design shown in Table 1 the participant is presented the first condition. The visualization is explained, and once understood, the participant conducts the four tasks. After a task is completed, the answer is given to the experiment leader, who writes the answer and the time required down. After completing the four tasks, a questionnaire was filled-in with ten questions to

acquire subjective measures on user preferences and satisfaction. This procedure was repeated for the other three conditions as well.
At the end of the experiment when all conditions were used by the participants, they were asked which condition they liked best and how much they liked it. The experiment took approximately 45 minutes to complete.

4. Results

Eight participants conducted 16 task assignments each, making a total of 128. One task assignment was considered an outlier and was excluded from analysis (a navigation task conducted with the aggregate spread condition), because the participant got totally confused. Excluding this task assignment meant that the mean of the task assignments for that condition changed from 78.3 to 62.7, the standard deviation from 100.9 to 49.9, and the longest time from 561 seconds to 172 seconds.
Given enough time, it was possible to conduct all tasks in all conditions. Because there were no time limits, no wrong answers were given in the experiments (sometimes, several correct answers were possible). Therefore our analysis looked into efficiency and user strategy, effectiveness was left out.

4.1 Task dependency

One of the hypotheses was that which condition is best depends on task characteristics. Figure 5 shows the performance for the four conditions on the four different tasks. Both time required to complete the task and the number of clicks are presented. Interesting to notice is that the number of clicks on the navigate task is close to zero in the spread condition. Three participants did not click at all to find the fastest route.

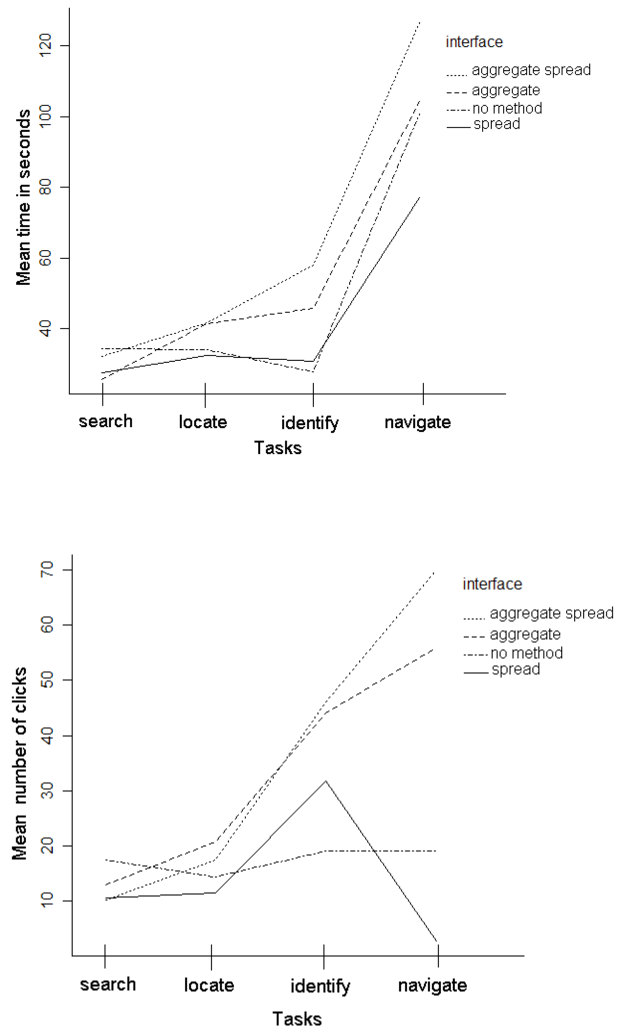

Figure 5: Mean time and mean number of clicks for the different conditions and tasks.

Table 2 shows the number of clicks per condition object per task and the mean number of clicks and time in seconds per task. There were no significant differences for time, but there were significant differences in number of clicks (Table 3).

An ANOVA showed a high interaction between tasks and speed ($F=60.2$, $p<0.001$) and number of clicks ($F=6.2$, $p<0.02$). However, no significant interaction was found between tasks and conditions ($F=0.5$, $p=0.46$).

Table 2 Mean number of clicks per condition object and mean number of clicks and mean time per task

Tasks	Menu clicks	Icon clicks	Pan-zoom clicks	Layer clicks	Clicks (sd)	Time in secs (sd)
Search	0.2	2.7	7.8	1.9	12.8 (11.5)	30.0 (19.3)
Locate	0.1	5.5	8.2	2.2	16.0	37.5

					(8.6)	(16.1)
Identify	4.8	21	3.3	2.0	35.3 (37.2)	40.7 (37.2)
Navi-gate	0.7	26.7	3.5	4.3	35.9 (31.9)	101.4 (45.3)

Table 3 Significant differences on clicks between the conditions on specific tasks

	NM vs. AGS	NM vs. AG	NM vs. S	AGS vs. S	AG vs. S
Locate				T=4.0 df=7 p<0.01	
Navigate	T=-2.2 df=7 p<0.05	T=-5.4 df=7 p<0.005	T=4.6 df=7 p<0.005	T=2.7 df=7 p<0.05	T=5.9 df=7 p<0.001

4.2 Comparison of condition efficiency

Table 4 shows the mean time and number of clicks per condition. Table 5 shows the significant differences between the conditions for time and number of clicks.

Table 4 Mean time and number of clicks per condition

Conditions	Time in seconds (sd)	Clicks (sd)
NM	49.4 (41.0)	17.5 (12.1)
AGS	62.7 (49.9)	34.8 (33.3)
AG	54.3 (46.9)	33.4 (35.4)
S	42.0 (27.8)	14.1 (14.8)

Table 5 Significant differences for time and clicks between the different conditions.

	NM vs. AGS	NM vs. AG	AGS vs. S	AG vs. S
clicks	T=-2.2 df=31 p<0.05	T=-2.6 df=31 p<0.01	T=2.5 df=31 p<0.01	T=3.1 df=31 p<0.01
time	T=-2.2 df=7 p<0.05	T=-5.4 df=7 p<0.005	T=2.3 df=31 p<0.05	T=2.0 df=31 p<0.05

4.3 User preferences

Because we only had eight participants, no statistical analysis of the questionnaires was done, except for the correlation between subjective and measured time (cor=0.93).

The participant were asked to rate the methods for different aspects. Table 6 shows the mean score for each condition. The spread method was favored in all aspects. Participants were also asked to rank the four different conditions for each task. Table 7 shows that the spread method is preferred for two of the four tasks, the aggregate method for the two other tasks.

Table 6 Mean scores per question from questionnaires green means best score in the column and red means worst score in the column.

	Clear	Easy to use	Good idea	Easy to learn	Enjoy able	Fast
NM	1.5	2.88	3.25	1.5	3.25	3.13
AGS	1.75	3.13	3	1.75	3.38	3.25
AG	1.75	2.75	2.63	1.38	2.88	2.88
S	1.38	2.25	2.13	1.25	2.5	2.38

Table 7 Rankings of the methods for the four different tasks.

	Search task	Locate task	Identify task	Navigate task
NM	1.75	2.5	3.375	3.13
AGS	1.75	2.125	3.625	3.75
AG	1.88	1.88	3.13	3.38
S	1.13	2.25	3.88	3.13

There is quite some variance between participants and their mean number of clicks and mean speed (Figure 6). Because of the small number of participants, we cannot draw conclusions on gender differences.

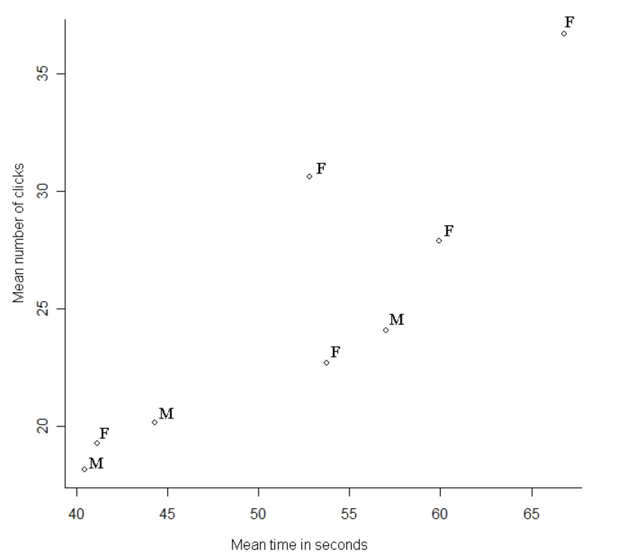

Figure 6: Number of clicks and mean speed per participant (F=Female, M=Male participant)

5. Discussion and Conclusion

We expected that the most efficient, effective and satisfying condition would depend on the task at hand. This is partly supported by the experimental results. The subjective results are in agreement with the expectations concerning the preferred method for the different tasks. The objective results on the other hand show that the preferred method for a certain task is not always the fastest or needs the least number of clicks (the aggregate condition is preferred for the identification task but was not the most efficient method). Overall, the spread method is the fastest and most click efficient. This method is perhaps not appreciated for the identification task, because for this task it is necessary to remember all the icons you clicked before. The spread method may create a higher workload than the aggregate method that has considerable less icons.

The implementation of both aggregate methods can be improved. In this study, the participants did not receive all the information about the aggregated icons when they clicked an aggregate icon, while this was the case when they pressed non-aggregated icons. The number of clicks can be significantly decreased when all relevant information is presented directly. The effect of directly available relevant information can be observed in the number of clicks needed for the navigate task with the spread condition. Three participants did not click on the screen at all in this task with this condition, because all information they needed was presented on the screen. To read street names it was necessary to zoom in completely. Therefore, for the two tasks for which street names were required, differences between the four conditions were minimal. In future experiments we will look further into the individual and gender differences.

6. ACKNOWLEDGMENTS

This publication is a product of the Dutch research program 'Space for Geo-information', project RGI-233 ‘Usable and well scaled mobile maps for consumers’.

7. REFERENCES

[1] Baldassi, S. & Burr, D. C. (2004). "Pop-out" of targets modulated in luminance or colour: the effect of intrinsic and extrinsic uncertainty. Vision Research, 44, 1227-1233.

[2] Blinn, C. R., Queen, L. P., & Maki, L. W. (2008). Geographic Information Systems: A glossary. http://www.extension.umn.edu/distribution/naturalresources/components/DD6097hr.html [On-line].

[3] Chittaro, L. (2006). Visualizing Information on Mobile Devices. COMPUTER, 40-45.

[4] Edwardes, A., Burghardt, D., & Weibel, R. (2005). Portrayal and Generalisation of Point Maps for Mobile Information Services. In L. Meng, A. Zipf, & T. Reichenbacher (Eds.), Map-based Mobile Services: Theories, Methods and Implementations New York: Springer.

[5] Fuchs, G. & Schumann, H. (2004). Intelligent Icon Positioning for Interactive Map-based Information Systems. In.

[6] Funda (2008). http://www.funda.nl/WoningAanbod/Koop/ [On-line].

[7] Google maps (2008). http://maps.google.nl [On-line].

[8] Julesz, B. (1984). A brief outline of the texton theory of human vision. Trends in Neuroscience, 7, 41-45.

[9] Lee, J., Forlizzi, J., & Hudson, S. E. Iterative Design of MOVE: A situationally Appropriate Vehicle Navigation System. International Journal of Human-Computer Studies, (in press).

[10] Mack, M. L. & Olivia, A. (2004). Computational estimation of visual complexity. In.

[11] Phillips, R. J. & Noyes, L. (1982). An Investigation of Visual Clutter in the Topographic Base of a Geological Map. Cartographic Journal, 19, 122-131.

[12] Rayson, J. K. (1999). Aggregate Towers: scale sensitive visualization and decluttering ofgeospatial data. In (pp. 92-99).

[13] Reichenbacher, T. (2003). Adaptive Methods for Mobile Cartography. In (pp. 1311-1321).

[14] Rosenholtz, R., Li, Y., & Nakano, L. (2007). Measuring visual clutter. Journal of Vision, 7, 17.

[15] Rosenholtz, R. & Mansfield, J. (2005). Feature congestion: a measure of display clutter. In (pp. 761-770).

[16] Treisman, A. (1986). Features and objects in visual processing. Scientific American, 255, 114-125.

[17] Wickens, C. D. (2000). Human Factors in Vector Map Design: The Importance of Task-Display Dependence. Journal of Navigation, 53, 54-67.

[18] Yeh, M. & Wickens, C. D. (2000). Attention filtering in the design of electronic map displays: A comparison of color-coding, intensity coding, and decluttering techniques (Rep. No. Technical Report ARL-00-4/FED-LAB-00-2). Savoy, IL: University of Illinois, Aviation Research Laboratory.