HELSINKI UNIVERSITY OF TECHNOLOGY

Department of Computer Science and Engineering

Tomas Lindberg

Effects of Information Density and Size on the Perception of Graphics in User Interfaces

Master's thesis

Submitted 10.3.2003

Supervisor: Professor Marko Nieminen

Instructor: Risto Näsänen, Ph.D., Docent

HELSINKI UNIVERSITY OF TECHNOLOGY	ABSTRACT OF MASTER'S THESIS
Author:	Tomas Lindberg
Title of thesis:	Effects of Information Density and Size on the Perception of Graphics in User Interfaces
Date: 10.3.2003	Number of pages: 69
Department:	Computer Science and Engineering
Professorship:	Tik-121 User Interfaces and Usability
Supervisor:	Professor Marko Nieminen
Instructor	Risto Näsänen, Ph.D., Docent

Alphanumeric and graphical information needs to be presented in such a way that its perception is accurate, fast and as effortless as possible. This study investigated the effects of spacing and size of individual interface elements on their perception. Experiment 1 investigated the effect of spacing on the speed of perception for elements in groups of varying size and defined the number of icons that can be seen with one glance (that is, with one eye fixation), experiment 2 studied the effect of size and experiment 3 the subjective preferences of people.

In experiment 1 three subjects searched for a target icon from among a group of distractors. In the experiment the array size (set size) and inter-element distances were varied. The array sizes were 2x2, 3x3, 5x5, 7x7 and 10x10 icons and the inter-element distances were zero icons, ¼ icon, ½ icon, one icon and two icons. Eye movements were measured during the experiment. In experiment 2 subjects also searched for an icon (the same as in experiment 1) from a group of distractors. This time the array and inter-element distance were kept constant (10x10 and one icon respectively) with varying viewing distance. Viewing distances used were 28 cm, 57 cm, 114 cm and 228 cm corresponding to icon sizes in degrees of visual angle of 1.48, 0.74, 0.37, and 0.18, respectively. No eye movements were recorded. In experiment 3, fourteen subjects arranged paper printouts of different levels of inter-element spacing of one array size into their preferred order. That is, the one they liked the best was placed first and the one liked the least was placed last.

The results of experiment 1 showed that spacing does not have an effect on threshold search times. The perceptual span for icons was found to be 25 in total arranged in a 5x5 array. The size of the interface elements on the other hand was found to have a great effect. Icons smaller than 0.7 degrees of visual angle resulted in significantly raised threshold search times. For both experiments, however, there were differences between subjects. Experiment 3 revealed that an inter-element spacing of one icon is to be preferred and a spacing of zero icons is to be avoided.

Keywords: Visual search, icons, visual acuity, eye movements, perceptual span, stimulus density, crowding, stimulus size, design guidelines

TEKNISKA HÖGSKOLAN	SAMMANDRAG AV DIPLOMARBETE

Författare:	Tomas Lindberg
Titel:	Inverkan av täthet och storlek på hur grafik i användargränssnittet uppfattas
Datum: 10.3.2003	Antal sidor: 69
Avdelning:	Avdelningen för datateknik
Professur:	Tik-121 Användargränssnitt och användbarhet
Övervakare:	Professor Marko Nieminen
Handledare	FD, Docent Risto Näsänen

Alfanumerisk och grafisk information bör presenteras så, att den kan tolkas korrekt, snabbt och så enkelt som möjligt. Denna rapport undersökte hur tätheten och storleken av individuella element i grafiska användargränssnitt påverkar uppfattningen av dem. Experiment 1 undersökte hur tätheten påverkar söktiden för olika grupper av varierande antal element och definierade antalet ikoner som kan ses med en fixering. Experiment 2 undersökte inverkan av storleken av element och experiment 3 mänskors subjektiva åsikter.

I experiment 1 sökte tre försökspersoner efter en specifik ikon från en grupp distraktorer. I försöket varierades gruppstorleken och tätheten mellan elementen. Gruppstorlekarna var 2x2, 3x3, 5x5, 7x7 och 10x10 ikoner och deras inbördes avstånd var noll ikoner, ¼ ikon, ½ ikon, en ikon och två ikoner. Under försöket mättes försökspersonernas ögonrörelser. I experiment 2 sökte försökspersonerna också efter en specifik ikon (samma som i experiment 1) från en grupp distraktorer. I detta försök hölls gruppstorleken och tätheten konstanta (10x10 respektive en ikon) medan försökspersonens avstånd från monitorn varierades. Avstånden var 28 cm, 57 cm, 114 cm och 228 cm vilka motsvarar ikonstorlekar av 1,48 0,74 0,37 och 0,18 grader av visuell vinkel. Inga ögonrörelser mättes. I det tredje experimentet placerade 14 försökspersoner papperversioner av ikongrupper av en storlek men olika grader av täthet i den ordning de föredrog. Försökspersonerna placerade alltså den täthetsgrad de föredrog mest först och den de föredrog minst sist.

Experiment 1 visade att täthet inte har en inverkan på söktiden. Antalet ikoner som ses med en fixering visade sig vara 25 arrangerade i en 5x5 matris. Storleken visade sig å andra sidan ha en stor inverkan på söktiden. Ikoner mindre än 0,7 grader av visuell vinkel hade en avsevärt förhöjd söktid. Det fanns dock skillnader mellan försökspersonerna. Experiment 3 avslöjade att ett avstånd mellan elementen på en ikon är att föredra medan ett avstånd på noll ikoner bör undvikas.

Nyckelord:	Visuell sökning, ikoner, skarpsynthet, ögonrörelser, täthet, storlek, design riktlinjer

Preface

This thesis was written at the Brainwork Laboratory at the Finnish Institute of Occupational Health as part of a National Technology Agency (Tekes) sponsored project. Thank you to Kristian Lukander for alerting me to the possibility of writing the thesis at the laboratory and to Kiti Müller M.D. for giving me the opportunity.

I am sincerely grateful to my instructor Risto Näsänen Ph.D. for all the invaluable help and guidance he gave me. Had it not been for him, writing the thesis would have been a much longer and more painful process. I would also like to express my gratitude to Helena Ojanpää for answering numerous questions and helping me whenever I needed it. Huge thanks also to all the subjects who participated in the experiments and to my supervisor professor Marko Nieminen.

Finally, a warmest of thanks to my lovely girlfriend Hanna Sandvik for listening to me talking about my thesis on countless occasions and for understanding and helping me in my worries.

Helsinki, 10.3.2003

Tomas Lindberg

Table of contents

1 Glossary

df	degrees of freedom
GUI	Graphical User Interface
LCD	Liquid Crystal Display
LED	Light Emitting Diode
MS-DOS	Microsoft Disk Operating System
ns	not significant
p	probability
PC	Personal Computer
PDA	Personal Digital Assistant
RGB	Red, Green, Blue
χ_F	Friedman chi-square

2 Introduction

The mobile phones of the late 1980's and early 1990's were designed mainly for speech. Modern mobile phones and PDAs on the other hand are crammed with various features like calendars, calculators, games etc. regardless of whether people need them or not. At the same time the displays of the devices have developed a great deal. From having been simple monochrome LCD displays unable to display graphics, the displays when this thesis is being written (2002) feature colour and graphics enabling the design and use of complex interface elements. See Figure 1 and Figure 2 for a comparison of mobile phone and PDA interfaces respectively. Many of the contemporary devices have icon-based interfaces with which the user interacts either through the keypad of the device or through a pen and touch sensitive display.

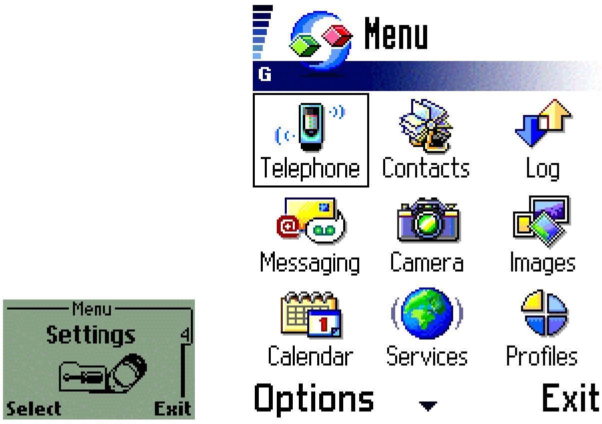

Figure 1. On the left a monochrome display of the Nokia 6210 mobile phone with a resolution of 96x60 pixels. On the right the display of the Nokia 7650, featuring 4096 colours and a resolution of 176x208 pixels.

Working with these mobile devices and also other devices with complex interfaces people receive large amounts of alphanumeric and graphical information that they need to interpret. This information should be presented in such a way that its perception is accurate, fast and as effortless as possible. Consequently, the user interface designers face the problem of having to make difficult decisions. Do they for example design an interface with large elements spaced sparsely or one in which the elements are small and close to each other?[1] People using the devices also vary greatly, adding to the problem. People may have poor sight and motor skills and no prior experience or they may be experts with years of experience and training. Also the situations in which devices are used vary, from a quiet evening at home in good lighting conditions to a noisy, dusty workplace where there might be rain or bright sunshine which may reduce visibility and contrast.

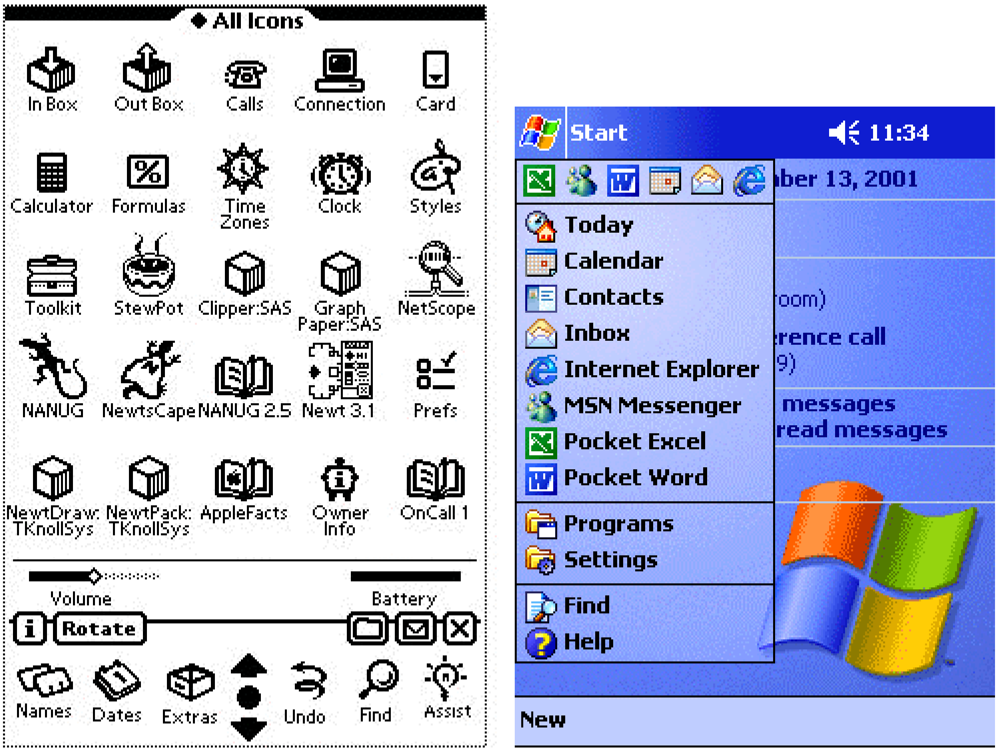

Figure 2. The Apple Newton 2.0 operating system on the left and Windows Pocket PC on the right.

As with alphanumeric information the efficiency with which graphical information is processed should depend on several different factors. For alphanumeric information these factors have been shown to be for example size

[1] In practice interfaces are often designed with a specific target group in mind (a process industry control application versus a children's educational program), or at least should be. Professionals have demands of their own, as do ordinary consumers and manufacturers do wisely in acknowledging these.

and contrast (Legge, Pelli, Rubin & Schleske, 1985; Legge, Rubin & Luebker, 1987; Näsänen, Karlsson & Ojanpää, 2001; Näsänen, Ojanpää & Kojo, 2001).

The processing of graphical information involves locating the correct piece of information, recognising the physical object and understanding what it means in the current context of use. Locating a specific piece of information, that is, visually searching for it is most difficult within a new environment and becomes easier as the environment becomes more familiar. Nevertheless, even a familiar environment requires some amount of visual search to pin down the exact location of the sought after information. Visual search, using the visual search paradigm (Treisman & Gelade, 1980), is an efficient way of measuring the effects of different presentation factors on the speed of information processing, the locating of a target. Consequently several studies have applied the paradigm to the search for visual information (Benbasat & Todd, 1993; Bewley, Roberts, Schroit & Verplank, 1983; Byrne, 1993; Moyes, 1994; Niemelä & Saarinen, 2000; Repokari, Saarela & Kurkela, 2002; Tanskanen, 1999; Wiedenbeck, 1999).

This study investigates the effects of the density and size of user interface elements, more specifically computer icons, on the speed with which users find what they are looking for using the visual search paradigm. It is hoped that as a result of this study, guidelines can be given to designers as to how densely elements in user interface can be packed and what the minimum size of interface elements should be.

3 Related research

This section presents the theoretical background and the most relevant results from several different fields of research relating to this study.

3.1 Visual search

Using the visual search paradigm subjects are shown a set of objects, and the aim is to search for one specific object, the target, from among the other objects, called distractors. The independent variable used is set size and the dependent is often the reaction time. The reaction time is defined as the time from showing the objects to when the subject indicates his/her answer. If the time increases linearly with increasing set size the search is said to be serial and if the reaction time is independent of set size, it is said to be parallel. A serial search is also considered to be self-terminating, that is, the search is terminated when the target is found but if no target is present the entire set is searched. This results in a target absent / target present slope ratio of 2:1. For a review on visual search see Wolfe (1998b).

3.1.1 Feature integration theory

One of the important theories on visual search and attention is the feature integration theory of attention developed by Treisman and Gelade (1980). The importance of the theory can be assessed by the amount of research that has followed it, both extending it and creating new competing theories (for example Duncan & Humphreys, 1989; McLeod, Driver & Crisp, 1988; Nakayama & Silverman, 1986; Sireteanu & Rettenbach, 2000; Theeuwes & Kooi, 1994; Treisman, 1982; Treisman, 1986; Treisman, 1998; Treisman & Gormican, 1988; Treisman & Souther, 1985; Wolfe, 1994a; Wolfe, 1994b; Wolfe, 1998c; Wolfe, Cave & Franzel, 1989; Wolfe & Gancarz, 1996).

This theory by Treisman and Gelade (1980) states that the search for separable features (for example colour, size and orientation) is parallel, that is, independent of set size. The search time for a conjunction of features, on the other hand, increases as the set size grows and is said to be serial. The search for a conjunction of features could for example be searching for a green T among brown Ts and green Xs, that is, knowing the colour or the letter alone is not enough, instead they have to be combined in order to find the target. According to Treisman and Gelade conjunction search is serial because it requires that the subjects direct focal attention to relevant locations serially to combine the features. They do, however, suggest that attention does not necessarily need to scan items individually, but that it could be directed to groups of items given the condition that no features within the group could recombine to create illusory targets. Illusory targets are targets made up of the features of different objects and may appear when attention is not focused on a specific object to combine its features correctly. Features, on the other hand, are assumed to be 'registered early, automatically and in parallel across the visual field', with only visual acuity, discriminability of objects and their lateral interference setting the limits.

3.1.1.1 Development of the feature integration theory, by Treisman

The feature integration theory by Treisman and Gelade (1980) predicted that instead of directing attention only to single items it could be directed to groups. A study by Treisman (1982) confirmed this prediction. Treisman (1982) found that perceptual grouping of the items significantly speeded conjunction search. She explained this by a parallel search within the groups. In another article Treisman and Souther (1985) presented a model of visual processing. According to this model the basic features are encoded in feature maps from which focused attention using a 'master map of locations' integrates the features at a particular location. The targets and distractors are assumed to produce units of activity in the feature maps. These activations are then pooled together and the higher the ratio of target activity to distractor activity is, the faster the target is found. This could according to Treisman and Souther (1985) explain differences found in search slopes, that is, differences in search efficiencies. Whereas the original feature integration theory (Treisman & Gelade, 1980) assumed that conjunctions require attention either to single items or spread over groups to bind features together, a later modification to the theory (Treisman & Gormican, 1988) allows for it to vary along a continuum. Feature search would thus correspond to completely divided attention spread over the entire display. Treisman (1998) presents still another modification to the model of visual processing. According to this model different values of features are represented in only a few feature maps. All different orientations would for example only be represented in maps for vertical, diagonal and horizontal. An example Treisman (1998) gives is the search for a slightly tilted line among vertical ones. The tilted line would activate the map for vertical lines the most but also slightly the map for diagonal lines. The vertical line, on the other hand, would only activate the vertical map and the target would thus be easy to find. In the opposite case of searching for a vertical line among tilted ones would by the same principle be difficult with both target and distractors activating the vertical map. This has also been found to be the case in practice (Treisman, 1986; Treisman & Gormican, 1989). This modification would account for why search for features that according to the original theory should be detected in parallel can become more difficult as targets become more difficult to discriminate from the distractors.

In other words, the feature integration theory originally divided search into serial and parallel with the serial search requiring focal attention to individual items. After modifications it now allows the spread of attention along a continuum and the pooling of activity created in a few feature maps per dimension to account for a variation in search efficiencies.

3.1.2 Critique of the feature integration theory

The feature integration theory developed by Treisman and Gelade (1980) has come under some critique and has been thought to be an over simplified model of reality (Duncan & Humphreys, 1989; McLeod, Driver & Crisp, 1988; Nakayama & Silverman, 1986; Sireteanu & Rettenbach, 2000; Theeuwes & Kooi, 1994; Wolfe, 1998c).

One line of research has shown that some cases of conjunction search can be performed in parallel, which is in contrast to the feature integration theory.

Nakayama and Silverman (1986) found that visual search for a conjunction of stereoscopic difference and motion or colour was independent of set size. Nakayama and Silverman do, however, explain it in accordance with the theory as a parallel search in each depth plane in turn. In other words, attention can be restricted in the spatial domain.

McLeod, Driver and Crisp (1988) also found conjunction searches whose reaction times do not behave as predicted by the theory of Treisman and Gelade (1980). In their study McLeod et al. reported that visual search for a conjunction of motion and shape is independent of set size. In an experiment they asked subjects to search for a moving X from an array of moving Os and static Xs. They found that the line depicting reaction time as a function of set size was ‘essentially flat’ for both the cases of target present and target absent. McLeod et al. suggest that the human visual system can limit the search to the moving elements by arranging all the elements into two groups, a group of moving elements and a group of stationary element and then ignoring either one, in this case the stationary group.

Still another study that reported a parallel search for a conjunction of features is one by Theeuwes and Kooi (1994) where the features conjoined were contrast polarity and shape. The subjects were asked to search for an O from an array of Xs and Os of different polarity and the search slopes were found to be flat both for target present and target absent.

Whereas the studies reported above found conjunction searches that were immediately parallel, a study by Sireteanu and Rettenbach (2000) indicates that ‘under some circumstances, initially serial search tasks can become parallel with practice’. Conjunctions of colour and orientation are not, however, reported as becoming parallel with practice but do become ‘very efficient’. The learning is attributed to a different search strategy that emerges after much practice.

Finally, Wolfe (1998c) reviews in an article the results he has gathered over a period of ten years from 2500 subjects performing approximately one million search trials. Wolfe finds several results that contradict the feature integration theory. Firstly, Wolfe reports that the distribution of the search slopes is unimodal, that is, there is no clear division between serial and parallel search tasks based on the slope alone. Secondly the ratio between target absent and target present is higher than the 2:1 ratio expected for serial self-terminating search. Thirdly, even if search is categorised into feature, conjunction and spatial configuration searches the search slopes of the different categories overlap each other. This means according to Wolfe that by knowing the search task the slope can be predicted but not vice versa. He does, however, suggest that knowing the slopes and the slope ratios between target present and target absent slopes it is possible to create diagnostics tests that could categorise search tasks.

In summary, the feature integration theory has come under criticism because some searches that the theory predicts should be serial have been found to be parallel and initially serial searches can become parallel with practice. Also, search slopes are continuously distributed and the predicted 2:1 slope ratio of target absent to target present does not hold in practice.

3.1.3 Alternative theories

Alternative competing theories and models to the feature integration theory have emerged over time from different researchers. This section deals with a theory developed by Duncan and Humphreys (1989) while the next section goes through the guided search model originally presented by Wolfe, Cave and Franzel (1989).

The model proposed by Duncan and Humphreys (1989) differs from the feature integration theory in that it does not categorise searches as being either serial or parallel but is instead based on a continuum of search efficiencies. It does also not distinguish between attributes of the stimulus but rather on their relationships, the level of similarity.

The model is described as a search surface spanned by axes of level of nontarget/nontarget similarity and target/nontarget similarity. As target/nontarget similarity increases the search efficiency decreases as well as when nontarget/nontarget similarity decreases. The axes interact so that an increase in target/nontarget similarity has a relative small effect on search efficiency as long as the nontarget/nontarget similarity is high. A decrease in nontarget/nontarget similarity has virtually no effect on the efficiency of search given that the target/nontarget similarity stays low. On the other hand, when nontarget/nontarget similarity is low an increase in target/nontarget similarity greatly affects the search efficiency making it worse. The same thing results when nontarget/nontarget similarity is decreased in combination with a high level of target/nontarget similarity. This can be summarised as the more similar the target is to the distractors the more difficult the search is and the less similar the distractors are to each other the more difficult the search is. The description is still, however, rather complicated and is clearly best explained and visualised by a picture. A picture approximating the search surface proposed by the model of visual search presented by Duncan and Humphreys (1989) is shown in Figure 3 below.

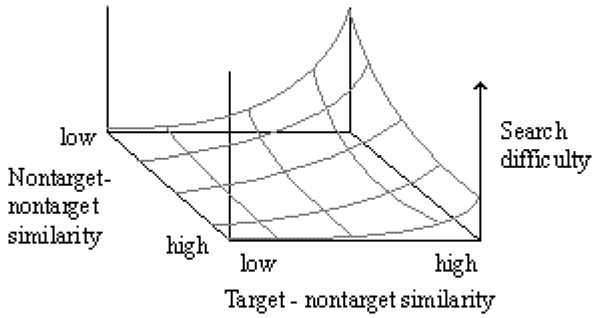

Figure 3. The search surface created by the visual search model by Duncan and Humphreys.

3.1.4 Guided search

The second alternative model of visual search is guided search presented by Wolfe et al. (1989). According to this model 'parallel processes guide the "spotlight" of attention toward likely targets'. This is very different from the feature integration theory. According to the feature integration theory, if a single feature does not define the target, the subsequent serial processes do not use the information gathered by the parallel processes.

When developing the guided search model Wolfe et al. (1989) performed a number of experiments where subjects searched for conjunctions of features. These experiments gave results that do not agree with the ones reported by Treisman and Gelade (1980), for example, the slopes of target present to target absent trials are shallower and practice seems to improve search performance for target not present cases. They also emulated an experiment performed by Treisman and Gelade and compared it to a control condition with higher stimulus salience, that is, where the stimulus was more clearly visible. The higher salience condition produced significantly faster reaction times.

The guided search model makes use of the parallel processes, or rather the feature maps generated by the parallel processes. Imagine searching for a conjunction of colour and size, for example for a big red ball from an array of big blue balls and small red balls. In this case the different feature maps for colour and size would be excited based both on bottom-up processing of stimulus features and top-down knowledge that the subject has about the target. Combining the feature maps of red and big would then produce an attention map with the biggest excitation at the location where the combination of red and big can be found. Attention is then thought to move to the point of greatest excitation moving on to the next greatest point in case no target was found. The model assumes excitation of locations with targets but Wolfe et al. (1989) point out that it works equally well in the opposite condition, inhibition of nontarget locations.

The reason conjunctions of features are not found immediately despite the combination of feature maps is that the transmission of information form the parallel stage to the serial stage is not perfect but contains noise. The amount of noise is seen to depend on the level of salience of the target, the higher the salience the lower the noise. Independent of the level of noise the signal will, however, sooner or later emerge from it due to movement of attention and the resulting updating of the feature maps. The information in the feature maps is constantly being updated as long as the stimulus is seen.

The guided search model by Wolfe et al. (1989) has later been updated as more knowledge accumulated. Wolfe (1994a) presented guided search 2.0 where the biggest changes were weighing of bottom-up activation based on element distance and similarity and top-down activation based on known properties of the target. The combination of these in creating the activation map is also weighted based on task demands, that is, more important features contribute more to the activation. The new model also states when a search is to be terminated, a point not defined by the original model. According to guided search 2.0 the targets are searched in order of decreasing activation until a target is found or no targets above a defined activation threshold remain. There is still one more update to the guided search model, guided search 3.0 (Wolfe & Gancarz, 1996). The difference between this one and the previous ones is that guided search 3.0 takes eye movements and the fact that visual processing is more detailed at the fovea into account.

As can be seen from the studies above the stimulus used is very simple and the results can thus be thought to apply only in theory. Wolfe (1994b), however, showed that the results are applicable to real world tasks. Wolfe used in his research continuous, naturalistic stimuli resembling aerial views of terrain and found that conjunction searches in these conditions were at least as good as searches in comparable tasks with isolated stimuli.

To summarise the alternative theories, the one by Duncan and Humphreys (1989) is based on a continuum of search efficiencies, not a categorisation into parallel and serial. It creates a search surface based on nontarget/nontarget similarity and target/nontarget similarity (see Figure 3); it is not dependent on the stimulus attributes. The guided search model also produces a continuum for the search efficiencies but is based on parallel processes that excite feature maps, which when combined guide the search to likely targets.

3.2 Icons

Icons and symbols have been in use for a long time and they are arguably the oldest form of written, or perhaps drawn in this case, communication. Take for example the ancient Egyptian hieroglyphs and Chinese logograms, an example of the evolution of the latter one can be seen in Figure 4. From the figure it can be clearly seen how the sign originally reminded of a fish but finding a fish in the modern one does require some imagination. Perhaps the most common form of symbols that people encounter in modern times is road symbols. While all are not directly intuitive, once learned, they effectively convey their meaning and are understood basically all over the world[2]. The reason that symbols are globally applicable is that they are language independent and to some extent even culturally independent. This study, however, uses computer icons. Icons are used in computers to represent objects and commands, for example an icon

[2] Naturally there are small variations from country to country depending on the local culture, but the general meaning stays the same.

to represent this text document or an icon representing bolding of a letter, word or part of the text.

Figure 4. The evolution of the Chinese logogram for fish through the four major phases of Chinese writing.

The use of computer icons has become more widespread during the last twenty or so years. The first computer to use icons is generally thought to be the Xerox 8010 ‘Star’ Information System unveiled in 1981 (Palo Alto Research Center Incorporated, 2002) which is based on the Xerox Alto personal computer that first became operational in 1973. However, the company often regarded, as the inventor of the graphical user interface is Apple who in 1983 released the Lisa (Apple-History, 2002). Today virtually every computer has an operating system with a graphical user interface and the use of icons is widespread. Even other devices such as mobile phones and PDAs, as mentioned in the introduction, are more and more using icon-based interfaces and the icons have become more complex as can be seen from Figure 5 and Figure 6.

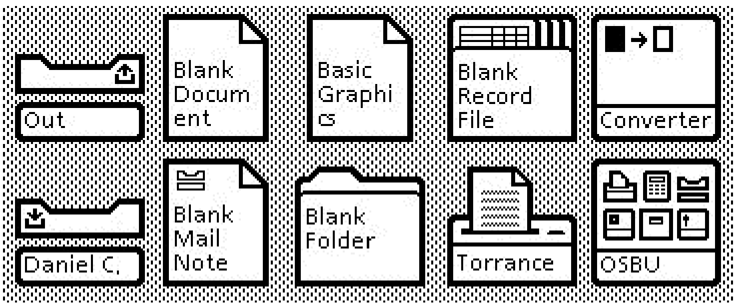

Figure 5. Icons used in the Xerox 8010 'Star' Information System.

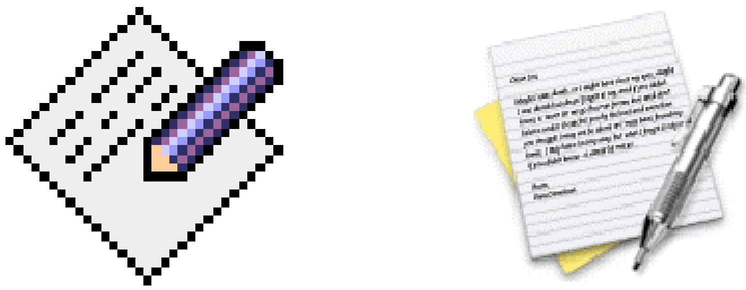

Figure 6. Two examples of Apple Macintosh icons. The icon on the left is from Mac OS 9 and earlier and the icon on the right from Mac OS X.

3.2.1 Why icons?

Reasons for using icons in computer interfaces are many and widespread (Bewley, Roberts, Schroit & Verplank, 1983; Gittins, 1986; Guastello, Traut & Korienek, 1989; Hemenway, 1982; Niemelä & Saarinen, 2000; Repokari, Saarela & Kurkela, 2002; Tanskanen, 1999; Wickens, Gordon & Liu, 1998; Wiedenbeck, 1999). Icons can be thought to be visually more distinctive from each other than words or letter strings used in character based interfaces and menus (Hemenway, 1982). At the same time related icons can be visually similar to aid the grouping of them together (Gittins, 1986; Hemenway, 1982). The icons also have the potential to represent lots of information in a small space (Gittins, 1986; Hemenway, 1982), which, however, adds the danger of making them too complex resulting in icons that are difficult to find (Byrne, 1993). It is generally acknowledged that recognition of an object is easier than recall from memory, as is the case when typing a command at a prompt (for a short review on visual memory see Wolfe (1998a)). This is perhaps one of the most important reasons icons and menus have replaced command prompts. In addition to facilitating recall, easing the burden on memory and making it easier to return to using a system after a period of absence, the use of icons and menus helps users avoid errors resulting from misspelling (Gittins, 1986). Icons, or combinations of icons and text, have also been found to be faster and easier to locate than text alone and easier to learn (Bewley, Roberts, Schroit & Verplank, 1983; Guastello, Traut & Korienek, 1989; Niemelä & Saarinen, 2000; Repokari, Saarela & Kurkela, 2002; Tanskanen, 1999; Wiedenbeck, 1999). In contrast to this, however, Benbasat and Todd (1993) did not find a performance advantage for icons over text. They concluded that 'the major advantages of iconic interfaces are not inherent to icons but properties of good design'. Thus, according to Benbasat and Todd, there is no difference between interface types as long as they are well designed. Bewley et al. conclude though in their article that 'icons were shown to be an acceptable user interface' and Wiedenbeck found in her study that learners prefer interfaces with icons to text-only interfaces, they rate higher in usability and usefulness.

3.2.2 Icon design

Designing icons is a difficult task. Icons should ideally immediately and intuitively convey their meaning, independent of the educational, professional or cultural backgrounds of the users[3]. Already a quick glance should reveal the meaning of the icon to the user. Icons representing same kinds of operations or objects should appear similar. This would make guided search possible. If for example the user knows that a red button stops an action, he/she can let red objects that appear in the interface guide the eyes. Several experiments relating to the design of icons have been made and guidelines for designing them exist (Apple, 2002; Bewley, Roberts, Schroit & Verplank, 1983; Byrne, 1993; Gittins, 1986; Goonetilleke, Shih, On & Fritsch, 2001; Hemenway, 1982; Huang, Shieh & Chi, 2002; Microsoft, 2002; Nokia, 2002; Wiedenbeck, 1999; Ximian, 2001).

While it is possible to design icons with a very high level of detail, a study by Byrne (1993) suggests that simple icons are better than complex ones. This is according to Byrne especially the case in visual search tasks with large set sizes, when the subjects have knowledge about the icons or when the filename is not known. This coincides with Gittins' (1986) opinion that simple icons are at least as usable as complex ones if not even more so.

[3] Ideally the icons should be designed for the individual user but that is an impossible task in practice. Designing different sets of icons for different regions and cultures is, however, possible, at least for large software companies. The smaller ones seldom have the recourses for this, which is why the icons need to be general enough to be applicable on a global scale.

Knowing that simple icons are to be preferred over complex ones, how about concrete versus abstract icons, which are better? In an experiment conducted by Blankenberger and Hahn (1991) the effect of a measure called ‘articulatory distance’ on reaction time was measured. The icons were generated by having subjects draw icons of 17 basic word processing commands and adding to that group 17 new arbitrary icons. Afterwards 10 new subjects assigned the generated icons to the 17 word processing commands. Finally three groups, ‘near’, ‘far’ and ‘arb’, of icons were generated based on how they were assigned to the commands. The ‘near’ group consisted of icons assigned 100% correctly, and the ‘far’ group of icons assigned 70% - 90% correctly. The final icon group, ‘arb’, consisted of the arbitrary icons assigned to commands not chosen for those icons by the 10 subjects assigning them to the commands. Finally a group called ‘words’ of short descriptions was created.

The subjects’ task was to locate the icon or word corresponding to a command from within a set with reaction times being measured. Whereas the ‘word’ group had the fewest errors, the reaction times were fastest for the ‘near’ group with ‘far’, ‘arb’ and ‘word’ coming next and in that order. In other words, the experiment by Blankenberger and Hahn (1991) shows that concrete representations are better than abstract ones, at least when presentation order is randomised.

To test what effect fixing the positions of icons has Blankenberger and Hahn (1991) repeated the experiment above except that this time the icons used had fixed positions. The result of this experiment showed that the subjects established ‘local bindings’, they linked functions with positions. This result by Blankenberger and Hahn was further confirmed by the results of a study by Moyes (1994). In her study Moyes used the ‘near’ icons created by Blankenberger and Hahn as representational icons to which she added another 17 abstract icons. She used four groups of subjects in her study, the tasks of which were to search for an icon corresponding to a specific command from among distractors. After five blocks of 17 trials each one group of subjects switched from representational icons with fixed positions to abstract icons with fixed positions, one from abstract icons with fixed positions to representational icons with fixed positions, one from representational icons to representational icons in random positions and finally one from abstract icons to abstract icons in random positions. She found no significant change in search times for groups two (abstract to representational) and three (representational to random). For the two other groups, however, the search times were significantly slower. This indicates according to Moyes that subjects using representational icons rely on their shape whereas subjects using abstract icons rely on position. These studies suggest that even poorly designed icons can be used because their positions can be learned. However, this does not apply to the casual user or the beginner, they need clear, simple icons since they can not be expected to know the positions of the icons.

Based on the studies reported above, we know that icons should preferably be simple and concrete but there is surely more to icon design than that. A study by Huang, Shieh and Chi (2002) sought to investigate the opinions of GUI designers on the relative importance of design elements. They first constructed a 50-item list from literature sources that was commented and modified by two designers resulting in a 19-item list of design elements. This list was then shown to 43 GUI designers who were asked to rate the elements on a six-point scale (1 being unimportant and 6 very important). This resulted in a five-factor list of design elements (means of points shown in parenthesis): meaningfulness (4.2), locatability (3.9), message quality (3.3), styling quality (3.1) and finally metaphor (3.0). The first factor, meaningfulness, corresponds to understanding what the icon means. The factor includes the items communicativeness, reconizability and test-before-use. The second one, locatability, on the other hand contains the items familiarity and discriminability, which help users locate the desired icon. Huang et al. named the next factor message quality because ‘it included items that describe how to faithfully convey messages to the users’, items such as consistency, simplicity and feedback to name a few. Styling quality on the other hand relates to the physical form or ‘style of an icon’ and contains items like colour, layout and typography. The final factor contains only one item named the same as the factor, that is, metaphor and relates to using concepts and relations familiar to the user.

To conclude this section about icon design, when designing icons make them simple and concrete when possible[4]. As can be seen from the icon to the right in Figure 6 and the icon in Figure 7 modern icons do not necessarily follow the principle of simplicity. They should also be meaningful, locatable, consistent and give feedback. Do not, however, forget to also make them look good. The last, but definitely not the least, thing to remember when designing icons is to follow possible guidelines for the system being designed for (Apple, 2002; Microsoft, 2002; Nokia, 2002; Ximian, 2001). These guidelines control things like the use of colours, size, shape and labelling; things that need to be taken into account when starting the design process. An example of the perspective grid used to design Windows XP icons presented in the Microsoft guideline (Microsoft, 2002) for creating XP icons can be seen in Figure 7.

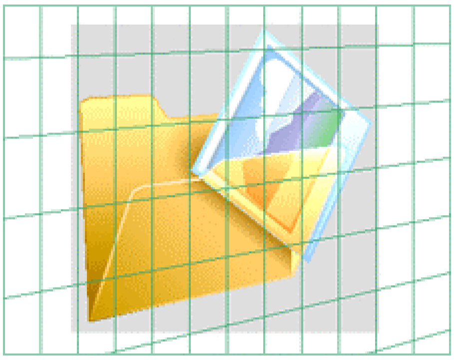

Figure 7. The perspective grid used to create Windows XP icons.

3.2.3 Visual search for icons

Now that reasons for the use of icons and the issues relating to the design of icons have been dealt with, it is time to consider visual search for icons. This is a topic that has also received some attention in scientific research (Benbasat & Todd, 1993; Bewley, Roberts, Schroit & Verplank, 1983; Byrne, 1993; Moyes, 1994; Niemelä & Saarinen, 2000; Repokari, Saarela & Kurkela, 2002; Tanskanen, 1999; Wiedenbeck, 1999).

[4] It is naturally not always possible to design concrete icons. For example, how does one represent the operation of compilation? In these cases possible standard or de facto standard icons should be used or new ones designed that make use of domain specific knowledge that the intended users might be expected to have.

In his research Tanskanen (1999) studied visual search for a target file among distractors by measuring reaction times as a function of set size and target type. The targets were normal icons found in the Apple Macintosh Plus computer containing either both icon and text, icon only, text only or an enlarged target icon. The icon only and text only versions were created from the combined icon and text icon by removing the text and icon respectively. The reaction times for the normal sized icons and text only version were found to increase serially with an increased set size with the greatest reaction times for the text only alternative followed by the icon plus text alternative with the icon only being fastest. Search times for the large target icon did not increase with increasing set size. The error rates for the target present case increased with increasing set size and were highest for text only followed by icon and icons and text. Tanskanen also asked the subjects for their subjective opinions about the target types. The large target was considered the easiest and the text only the most difficult. The icon and icon and text targets were considered equally easy or difficult.

In another experiment in the same study by Tanskanen (1999), he investigated the effect of heterogeneity. In this experiment he used two different targets, one the same as in experiment one and another one more similar to the distractors. The distractors were also modified, one group of distractors were homogenous (all icons the same, only the text varied), the second was semi-homogenous (three different types of icons) and the third heterogeneous (six different types of icons). The results of the experiment showed that reaction times increased with increased distractor heterogeneity for both targets. Another result was that reaction times for the more difficult target (more similar to the distractors) increased more with increasing distractor heterogeneity than for the easier target. Type of target and type of distractors are thus linked when it comes to reaction times.

In a somewhat similar study to that of Tanskanen (1999), Niemelä and Saarinen (2000) studied the effect of use of icons and their spatial grouping on scanning speed. In their study subjects were asked to search for a target file from among distractors. Niemelä and Saarinen had a total of three different conditions the first two of which contained standard Apple Macintosh file icons and the third textual descriptions of the file types instead of icons. In the first of the two icon conditions similar icons were grouped and in the second they were randomly placed. The results of the study show that search for grouped icons is faster than search for randomly placed icons as well as for grouped file names. In other words, using icons and grouping them improves search speed.

The studies reported above both used normal computer icons displayed on a PC; in other words, screen size and icon size did not affect the results. A study by Repokari et al. (2002) on the other hand investigated the use of icons on a small mobile phone display simulated on a PC. The size of the simulated display was 3.3 x 1.9 cm or 3.78^o x 2.18^o and icons used were 3.3 x 3.3 mm or 2.3 arc minutes of visual angle. Like the two studies above this one also used the visual search paradigm, the reaction time was measured as subjects searched for the target from amongst distractors. In line with the results reported above the study by Repokari et al. indicates that search for icons is faster than search for text and that the difference increases as set size increases. The study also investigated the effect of target / distractor homogeneity and like Tanskanen (1999) they found that search for targets among similar distractors was slower than search from among dissimilar distractors.

What can be said on the basis of these results is that search for icons seems to be faster than search for text and that the size of the display or the icons does not affect this. All of the studies reported in this section, however, had randomised the positions of the icons. As demonstrated by Blankenberger et al. (1991) and Moyes (1994), fixing the positions of icons does have an effect.

Another important result of the studies reported here (for example Niemelä & Saarinen, 2000; Tanskanen, 1999) is that they confirm the results found by Wolfe (1994b) that the theories of visual search can be applied to real world tasks. Whereas the studies used to derive the theories often had very simple stimulus, letters and shapes of different colour, size and orientation, these studies used both imaginary icons and icons found in real systems.

3.3 *Visual acuity, eye movements and the perceptual span*

Visual acuity is highest at the fovea and reduces greatly towards the periphery (Wertheim, 1894). This is due to a similar decrease of the ganglion cell density on the retina, which is also highest at the fovea (Curcio & Allen, 1990). As a result of this, in order to see information outside of the fovea clearly, the eyes need to move so that this area is brought into the fovea or close enough to it for identification.

The eyes only pick up information during fixations, that is, when the eyes are not moving (in reality there exists tremor, drift and micro saccades during a fixation (Ciuffreda & Tannen, 1995)). The average duration of a fixation is about 225 – 275 ms for reading and 275 ms for visual search (Rayner, 1998) although considerably shorter fixations (approaching 150 ms) have also been reported (Ojanpää et al. (2002); Näsänen & Ojanpää, (2002). Between every fixation there is a saccade, that is, a rapid movement of the eyes from one point of fixation to the next. Compared to fixations the duration of these saccades is much shorter, a 5^o saccade for example only lasts around 40 ms (Ciuffreda & Tannen, 1995). The velocities of saccades are very high, the reason that no information is picked up during saccades, and can reach values exceeding 600 degrees per second (Ciuffreda & Tannen, 1995). The brain, however, integrates the images at different fixations so that the observer sees a continuous image. For a comprehensive review on eye movements see Rayner.

3.3.1 Eye movements during visual search

As mentioned above, high visual acuity is limited to the fovea and in order to see objects outside the fovea clearly they need to be brought to the fovea through eye movements. This section aims to establish if there are any features of the eye movements that are specific to visual search.

A study by Jacobs (1986) investigated the search for specific letters from a long string of one type of distractor letter. The results of the study show that when the target is easy to find, the saccades are long and of variable length but the fixations are constant and short. In the opposite case of targets that are difficult to find the fixations are long and variable while the saccades are short and constant. The same study also showed that viewing distance and element spacing do not affect saccade sizes but target/nontarget similarity does. As the target becomes more similar to the nontarget (the distractor) the saccades get shorter. Element spacing was also found not to affect fixation duration. The finding by Jacobs that the more difficult the target is to discriminate from the distractors the longer the fixation becomes was confirmed by Hooge and Erkelens (1999).

Results similar to the ones reported above were found by Scialfa and Joffe (1998). They investigated the effect of target eccentricity and as expected reported that as the eccentricity increases so does the number of saccades. According to Scialfa and Joffe the fixation duration decreases as the target eccentricity becomes greater and the number of saccades increases. The decrease is, however, reported to be non-linear. A further experiment in the same study investigated the effect of practice. With practice the eccentricity effect on the number of saccades is removed, that is, as the subjects became more practised they no longer needed to make saccades. At the same time as the number of saccades decreased the fixation duration increased, they are in other words linked as the other results above also indicate.

Moving on to the effect of set size on eye movements. A study by Näsänen and Ojanpää (2002) reported an increasing number of fixations as the set size increased, but the fixation duration stayed the same. Because the number of fixations increased so too had to the number of saccades. The saccade amplitude also shows the same trend, but only up to a specific limit after which it levelled off. The reason for the increase is explained by Näsänen and Ojanpää to be that at medium set sizes the target is more likely to be close to the initial fixation point why subjects only need to make small saccades in order to find it. The point at which the saccade amplitude levelled off, 6.9^{o}, is roughly two times the width of the target and is in line with their finding that the perceptual span for faces, as reported in more detail below, is 2x2.

To summarise this section about eye movements in visual search the results show that as the target becomes easier to find the saccades become longer whereas the fixation duration becomes shorter. The opposite has also found to hold, as the targets become harder to find the saccades become shorter and fixation durations correspondingly longer. Making the target more eccentric has the effect of increasing the number of saccades at the same time as the duration of the fixations becomes shorter. Finally as the set size increases so does the number of saccades and fixations but the fixation duration stays the same. Saccade size on the other hand only increased up to a certain limit. Rayner

(1998), however, points out in his review that the nature of the visual search task influences the eye movements. In other words they vary from task to task and what applies to one task may not necessarily do so for another one.

3.3.2 Perceptual span

Perceptual span is the area around the point of fixation from which useful information can be gathered. In reading, this information can for example be letters, the length of words and word spaces. According to the review by Rayner (1998) the perceptual span extends 3 – 4 character spaces to the left of fixation and 14 – 15 character spaces to the right, that is, it is asymmetric to the right. This is, however, a characteristic of the perceptual span, that is very much dependent on culture. In cultures where text is read from left to right the perceptual span is asymmetric to the right as described, whereas in cultures where people read from right to left it is asymmetric in that direction. Rayner also reports that the size of the span is culturally dependent with readers of Japanese for example having a smaller span on average. This might be due to the fact that Japanese signs are more complex than letters.

In a study by Rayner and Fisher (1987) the perceptual span during visual search was found to be very similar to that in reading. They found the perceptual span to extend 3 – 4 characters to the left of fixation independent of target/nontarget similarity. The span extending to the right of fixation was, however, found to depend on the level of similarity between targets and distractors. When they were similar the perceptual span extended 12 characters to the right of fixation compared to 15 – 16 characters when they were dissimilar. Rayner (1998) reported in the review about another study using a moving window where subjects searched for letters or digits. This study, however, reported the perceptual span in degrees and it was found to be 5^{o}.

The studies of the perceptual span in visual search reported so far have only used strings of characters and digits and have defined its size horizontally. A study by Pollatsek, Raney, Lagasse and Rayner (1993) indicates that when reading or scanning through text horizontally subjects make hardly any use of text below the fixated line. The reason for this, they hypothesised, is that information that is not in the direction of the eye movements is not used. It can also be argued that while reading subjects have no need to locate words above or below the currently fixated line and that this is why information on them is not used.

Prinz (1984), however, came to a completely different result. In his research subjects found a target letter 3.14 rows (corresponding to 1.76^{o}) below the fixated row and 2.30 rows (corresponding to 1.29^{o}) above it. The subjects were, however, told that the target might appear below or above the currently fixated line. Another aspect through which the study by Pollatsek et al. (1993) differed from the one by Prinz was that Pollatsek et al. had subject search for words while Prinz had them search for letters. In other words, the former study represented more normal reading while the latter one represented visual search. These results by Prinz are confirmed by the results obtained by Ojanpää et al. (2002). They investigated the visual search for words in vertical and horizontal lists. The results from the search of vertical list shows that approximately 4 – 5 words (no line spacing) can be scanned in a single fixation, whereas only 2.3 words are scanned per fixation in horizontal lists. From this Ojanpää et al. draw the conclusion that the area visible during a single fixation is roughly elliptical in shape and larger in the horizontal direction. This, they point out, is consistent with the findings by Wertheim (1894) and Curcio and Allen (1990) who investigated visual acuity and the anatomy of the eye. With the stimulus used by Ojanpää et al. the extent of the area visible during a fixation is 3.9^{o} horizontally and 2.6^{o} vertically and can thus be considered to be two-dimensional.

The perceptual span has now been found to be two-dimensional, but still only letters and digits have been used as stimulus. A study by Näsänen and Ojanpää (2002) used faces to investigate the size of the perceptual span. In their study subjects searched for a target face in an array of distractor faces. In the first experiment they varied the set size while the size of the faces was kept constant at 3.6^{o} x 3.6^{o}. The results show that only for the smallest array, 2x2, was one fixation enough to find the target, with all the others requiring two or more. Following this result Näsänen and Ojanpää investigated if the size of the face had an effect on the number of fixations. A second experiment was thus conducted where the size of the array was kept constant (2x2) while the sizes of the faces was varied, 0.9^{o} to 14.4^{o}. The only effect the size of the faces was found to have was a decrease in search time as the size of the faces increased to 3.6^o, from which point on it was flat. In other words, the number of fixations required to identify a face was independent of face size within the limits of face sizes used in the experiment. The study thus showed that four faces can be identified during a single fixation of the eye.

In conclusion, the perceptual span in reading and visual search in text is culturally biased and also task dependent. For a reader of English, it is asymmetric to the right. In a recent study by Ojanpää et al. (2002) the area that can be seen during a fixation was also found to be two-dimensional, more precisely elliptical and horizontally elongated. The perceptual span is not only limited to characters but has been shown (Näsänen & Ojanpää, 2002) to encompass a total of four faces arranged in a 2x2 array.

3.4 Stimulus density

The effect of density on search performance has been studied to some extent (Jacobs, 1986; Jenkins & Cole; 1982; Rayner & Fisher, 1987; Staggers, 1993; Treisman, 1982; Vlaskamp, Hooge & Over, 2002) but not to as a great extent as for example eye movements as accounted for in the sections above.

Vlaskamp et al. (2002) measured search performance, defined as search time divided by number of elements. In their study subjects searched for an ‘O’ among ‘C’s. In the first experiment Vlaskamp et al. varied the number of elements in the display and the gap size of the ‘C’s. The number of elements varied from 36 to 144 and the inter-element distances correspondingly from 7.08^o to 3.39^o. The gap sizes, on the other hand, were 0.09^o, 0.18^o and 0.27^o. The results of the first experiment indicated that the search performance did not vary with inter-element distance. For the second experiment they increased the number of elements to range from 36 to 576 with the inter-element distance ranging from 0.75^o to 3.20^o. This time the gap size of the ‘C’s was kept constant at 0.18^o. The results of the second experiment in their study gave the same results for inter-element distances larger than 1.5^o as the first experiment, that is, no impact on search performance. However, for distances smaller that this, search performance ‘decreased dramatically’.

Stagger's (1993) study on the impact of screen density found that nurses working on (non-graphical, character based) hospital information systems (HIS) found what they were looking for more quickly on high-density screens than on moderately dense screens and low-density screens. In the same study the target was also found faster on the moderate-density screen than on the low-density screen. The accuracy with which the targets were found and the subjective impression of the screens were practically the same for all three screen configurations.

In the study the density was defined as 'the number of characters divided by the total number of available spaces on the computer screen'. In this fashion the density of the densest screen was calculated to be 58.5%, the two moderately dense screens were 41% and 33% and the three low density screens were 31%, 27% and 32%. The problem with this study, however, is that the two moderately dense screens were simply the dense screen split into two separate screens and the three low density screens were the high density screen split into three separate screens. The way the splitting of the high-density screen was done the local density of all the screens remained the same, that is, the moderate and low screens had groups of information of the same density as the high screen but significantly more empty white space. This in practice made the search for the target equally difficult on each screen, and the longer search times could be argued to result from toggling between screens. The author did indeed offer this solution but only for the initial trials when the subjects were practising. After this toggling between screens was no longer assumed to affect the results.

The third study (Jenkins & Cole, 1982) investigating only the effect of density reported an effect of density on the conspicuity of discs differing in luminance but not on discs differing in size. A conspicuous object was defined in the study as 'one that will, for a given background, be seen with certainty within a short observation time regardless of the location of the object in relation to the line of sight'. The authors studied three levels of density in which discs occupied 5%, 10% or 15% of the maximum area that they could occupy, with targets positioned at eccentricities of 3^o, 6^o, 10^o, 14^o, and 18^o from a fixation cross presented for 250ms. The results of the study were that increasing the density from 5% to 10% affected the conspicuity of discs differing in luminance with the target becoming less conspicuous. That is, for a target to be seen at a specific eccentricity at a higher level of density, it needed a greater difference of luminance to the distractors than at a lower level of density at the same eccentricity. Increasing the density from 10% to 15% did not have any further effect on luminance and the conspicuity of discs differing in size was not affected in any way by changes in density. The same study reports a decrease in the area of conspicuity when the background density increases. This means that the area within which the target will be seen during a fixation will become smaller as the density increases. A point of criticism for this study is that the stimulus presentation time was as long as 250 ms, which does not necessarily limit the number of fixations to one (as reported above, fixation duration has been reported to approach 150 ms). This could thus have affected the results as the subjects might have moved their eyes to locate targets at greater eccentricities.

The three studies reported above investigated solely the effect of stimulus density. There have, however, been other studies that performed experiments where density varied but where the effect of density was secondary to the research question. One like this was Treisman’s (1982), where she studied the effect of density in order to evaluate if the serial search for a conjunction of features is centrally or peripherally determined. In other words, does it require successive fixations of attention or of the eyes? In her study Treisman mentions that search for conjunctions was ‘hardly affected’ by density. Rayner and Fisher (1987) were trying to define the size of the perceptual span during visual search. They reported in their study that the type of distractor had more of an effect on performance than did array density. Strings consisting of two letters, low overall density, were searched slower than longer strings corresponding to higher overall density. Jacobs (1986) on the other hand was studying the directness of visual span control. He found that letter spacing did not have a significant impact on saccade sizes nor did it have an effect on fixation duration.

As a conclusion on stimulus density it can be said that it does not seem to have an effect on visual search. The experiments, however, are not very representative of real world task or their validity can be questioned on other grounds. The first experiment by Vlaskamp et al. (2002) used ‘O’s and ‘C’s, the second by Stagger (1993) had screens with virtually similar local densities and the third by Jenkins and Cole (1982) had so long presentation times that subjects could fixate more than once. The last three, which were not directly interested in the effect of density, all had either groups or strings of letters from which the subjects were to search for a target letter. None of the experiments used graphical objects. The most valid and interesting of the studies reported here is the one performed by Vlaskamp et al. since they truly varied global density and the subjects could search freely using whatever strategy they wished.

3.4.1 Crowding

Bouma (1970) indicated that humans are better at identifying letters presented alone than when surrounded by other letters in close proximity. For a letter presented at an eccentricity of a^o to be regarded as being alone, or visually isolated, Bouma found that no other letters should be closer than $0.5a^o$. This effect has been named crowding. The study by Vlaskamp et al. (2002) reported above showed a decrease in search efficiency when the inter-element distance was less than 1.5^o and no improvement up to that point. As an explanation for this Vlaskamp et al. offered crowding. A recent study by Chung, Levi & Legge (2001) shows that crowding is affected by the spatial frequency of the target and the flanking letters in addition to the natural effect of letter spacing. When the spatial frequencies of the target and the flanking letters coincide the threshold search time for locating the target is significantly raised and the closer the letters are to each other the more the threshold is raised. Thus increasing the spatial frequency difference decreases the threshold elevation as does increasing the letter spacing. The increase in letter spacing only has an effect on the threshold up to a certain point, however, after which further increase in letter spacing does not decrease the threshold elevation. The critical letter spacing, that is, the letter spacing after which crowding no longer has an effect, was found by Chung et al. to be 0.5^o at the fovea and 2^o at a 5^o eccentricity and is independent of the spatial frequency of the target. This could indeed explain the results obtained by Vlaskamp et al. As the inter-element distances decreased the subjects could see more and more elements per fixation but the elements were getting so close to each other that the subjects could only identify the ones at the fovea or very close to it. This is confirmed by the decrease in saccade amplitude as the inter-element distance decreased reported by Vlaskamp et al.

4 Objectives of the study

As discussed in the introduction, devices such as mobile phones and PDAs with small screens are becoming more and more commonplace. In addition, the resolution of the screens is constantly increasing and they are to an ever-greater extent utilising colour, a combination of which enables the use of more complex graphics and the display of greater detail. This in turn may increase the amount of alphanumeric and graphical information that the users need to interpret per display.

The devices are used in a wide variety of situations, ranging from the occasional use of a mobile phone for phoning a friend to the use of a PDA in the process industry for controlling complex machinery, in other words, the demands and expectations placed on them vary greatly. Whereas the occasional user of the mobile phone might be content with a sparse layout of the elements on the interface, the worker might want to receive as much information as possible with one glance in order to get an overall view of the situation. In order to get much information to fit onto a screen either the size of the elements needs to be quite small or they need to be placed with high density, or both. This may result in elements that are difficult to see and distinguish.

The research questions this study aims to answer are:

- How does information density and icon (interface element) size affect the speed of visual search?
- What is the ideal density level for icons (interface elements)?
- What is the ideal size for icons (interface elements)?
- What are the subjective impressions and personal preferences of people for different levels of icon (interface element) density?
- What is the perceptual span for icons (interface elements)?

By answering these research questions the study aims to come up with guidelines for the use of density and size of icons (interface elements) in the design of user interfaces.

5 Methods

Experiment 1 was designed to reveal any possible effects that stimulus density might have on the speed, using threshold measurements, with which information can be gathered. In parallel to this, the size of the perceptual span for icons was defined. **Experiment 2** studied the effect of stimulus size on the performance of visual search for icons.

The results of experiments 1 and 2 can be used to design effective graphical user interfaces in which the users find the sought after objects quickly and efficiently. There is, however, more to the usability of a user interface than merely the efficiency and effectiveness with which it is used, the subjective preferences and aesthetic impressions of the users also play a very big role in the overall usability. Considering interfaces designed for consumers the subjective preferences and ‘feeling’ the interface gives the user can be even more important than overall efficiency. Therefore **experiment 3** aims to find out which level of density is rated highest.

The method selected for experiments 1 and 2 applies the visual search paradigm and uses a staircase algorithm to calculate threshold search times. The selected method is considerably more advanced than the reaction time measurements traditionally used in psychophysical experiments. The reaction time measurements give two results, the reaction time and knowledge of whether the answer was correct or not. The staircase algorithm, on the other hand, gives the threshold search time for a specific level of correct answers. Using this method there exists no trade-off between speed and correctness, which is the case with reaction time measurements. In addition to the trade-off, traditional reaction time measurements also include variations in the reaction times that come from individual processing times, that is, the time from seeing the stimulus to indicating the answer, a procedure often involving clicking a button.

5.1 Subjects

Experiments 1 and 2. Three subjects participated in the experiments. Of these one had university level education and two were university students with sixth form college education. All three subjects were of the same age group with two being 25 years old and one 26 and they all had normal or corrected to normal vision. Of the three subjects one was naïve, one slightly practised and one considerably practised. Since experiment 2 took place after experiment 1 all subjects were practised by that experiment.

Experiment 3. A total of fourteen subjects participated in the experiment, six men and eight women. Of the men three were in the age range from 20 – 25, one from 31 – 35, one from 41 – 45 and one from 46 – 50. Four of the six male subjects had university level education and two had sixth form college levels with one of these two also reporting vocational school levels of education. Of the women taking part in the experiment one was in the age range from 20 – 25, two from 31 – 35, three from 41 – 45, one from 46 – 50 and finally one from 51 – 55. The educational background of the women varied somewhat more than for the men. Five had a university education, one had a vocational high school education, one had graduated from an educational institute of healthcare and one had secondary school education. All of the subjects had normal, or corrected to normal vision and were naïve as to the purpose of the study.

5.2 Stimuli

The stimuli used in **experiment 1** were generated using the ImageSeries software written by Risto Näsänen, Ph.D. The stimuli consisted of twenty-five different icon matrix configurations of five different sizes (2x2, 3x3, 5x5, 7x7 and 10x10) and five different densities (inter-element distances being zero icons (0 pixels), ¼ icon (8 pixels), ½ icon (16 pixels), one icon (32 pixels) and two icons (64 pixels)). The total size of the matrices thus ranged from 64x64 pixels for a 2x2 matrix with inter-element distance of zero icons to 896x896 pixels for a 10x10 matrix with inter-element distance of two icons. For each combination of size and density 100 different matrices containing the target and 100 matrices not containing the target were generated. The matrices were made up of 219 different icons in total, 218 distractors and one target. All icons used were found in the Windows2000 operating system and were 32x32 pixels in size. The stimuli were saved in bitmap format with a colour depth of 24 bits. The background colour used in the matrix was grey representing the RGB values of 185 for each component, corresponding to half the maximum luminance of the display. Figure 8 shows a 5x5 matrix with an icon spacing of zero icons used in the experiment. See appendix 1 for a collection of icon matrices used in the experiment.

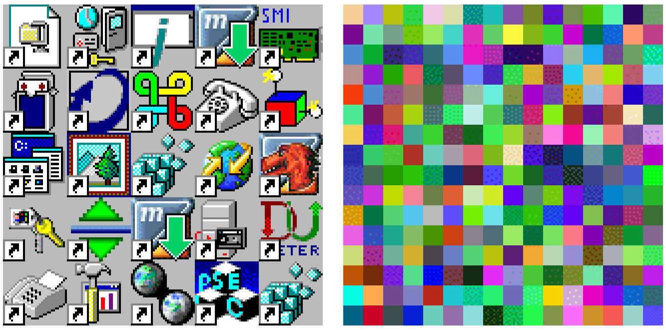

Figure 8. On the left a 5x5-icon matrix with an icon spacing of zero icons used as stumulus and on the right a mask shown before and after stimuli.

The stimuli used in **experiment 2** were the 10x10 icon matrices with an inter-element spacing of one icon from experiment 1. The reason for selecting the 10x10 matrix size was that it is large enough to allow eye movements. The icon spacing of one icon was selected because it had the best subjective impression as reported in the results of experiment 3 below.

Mask stimuli were shown before and after the stimulus matrices to mask any possible residual images on the retina that could affect the search results. The masks used were pixelated noise, with a check size of 10x10 pixels. See Figure 8 for an example of a mask. Like the stimuli, the masks were also saved in bitmap format with a colour depth of 24 bits. The software used for this was Image2.0beta written by Risto Näsänen, Ph.D.

The stimuli were presented using an SGI 1600SW LCD monitor with a #9 Revolution IV digital graphics adapter connected to a PC computer with a 450 MHz Pentium III processor running Windows2000. The pixel sizes of the monitor were 0.0231x0.0231 cm^2 corresponding to 0.0231 degrees of the visual angle with an observation distance of 57 cm used in the experiment. The software used was BMP-Search written by Risto Näsänen, Ph.D.

The stimuli of **experiment 3** were printed on an HP C LaserJet 4500 PS printer with a resolution of 600 dpi on Canon High Grade 100 g/m^2 coated paper. The paper presentation was done in order to enabled the subjects to easily arrange the stimuli in the order they preferred and to view them at the same time spread out on a table. The stimuli selected for this experiment were of all five different density levels arranged in 5x5 matrices, see Figure 8 for example of one density level and appendix 1 for a collection of density levels.

5.3 Procedure

In **experiment 1** the set size and inter-element spacing were varied. Of the three subjects one performed the visual search on all combinations of set size and density. The other two subjects used all densities but only set sizes 3x3, 5x5, 7x7 and 10x10. The reason for this being the extremely low threshold measured by subject TL for set size 2x2.

Experiment 2 varied the viewing distance, corresponding to changes in icon size. The distances used were normal distance (57 cm, or the same as in experiment 1), half the normal distance (28 cm), two times the normal distance (114 cm) and four times the normal distance (228 cm). The distances corresponded to icon sizes in degrees of visual angle of 0.74, 1.48, 0.37, and 0.18, respectively. See Figure 9 for example of the different icon sizes. As in experiment 1 threshold search times were measured.

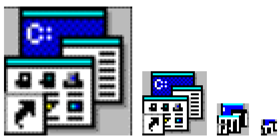

Figure 9. The change in size of the target icon. From the left sizes corresponding to 1.48^o, 0.74^o, 0.37^o and 0.18^o at a viewing distance of 57 cm.

In experiments 1 and 2 the subjects were asked to locate a target icon from a set of distractors. The target was either present or absent, both with a probability of 0.5. The subjects were asked to indicate if he/she found the target or not by clicking one of two buttons located to the left of the stimuli. If the target was present and the subject found it, he/she clicked a button with an image of the target and if not present or not found a button with the label ‘No’. A wrong selection was indicated with a sound. See Figure 10 for an example of the search display. Before and after stimulus presentation a mask, as described in the section above, was shown. The delay between clicking a selection button and showing the next stimulus was 500 ms. The subjects were not given a fixation point from which to start the search. Instead they were free to choose their own pattern of eye movements while searching for the target. The subjects were, however, instructed to immediately move their eyes to the buttons on the left-hand side of the screen when the target was found, then move the mouse cursor to the correct button and only after having moved their eyes back to the mask click the button. To ensure that the subject stayed as still as possible and that the observation distance remained constant at 57 cm a chin rest was used.

Figure 10. Example of the search display with a 10x10 matrix with inter-element distance of zero icons.

Following the procedure described above, the threshold search time, that is, the time in which the subject finds the target with a certain level of probability was measured. For every combination of density and size four different thresholds were measured and the final threshold was obtained as the mean value of these. After the subject had answered correctly on three stimuli in succession the presentation time of the stimuli was reduced by a factor of 1.26 and in the case of an incorrect answer the presentation time was increased by the same factor. The starting presentation time was 8000 ms. After eight reversals, that is, after eight changes from decreasing the presentation time to increasing it or vice versa, an estimate of the threshold at a probability level of 0.79 of correct responses was found (Wetherill & Levitt, 1965). Counting reversals was not started until the subjects had made two errors.

Experiment 3. The subjective impressions on the density of elements were examined with high quality paper print outs of different density levels in a single matrix size. Before showing the subjects the stimulus they were recorded for age, sex, education and their visual acuity was measured if necessary. Only subjects with normal or corrected to normal vision were accepted as subjects. The subjects were asked to arrange the pictures, presented initially in a randomly arranged pile, according to their subjective opinion of which looked or felt best, that is, which they would prefer to use in an actual user interface. The instructions were printed on paper to ensure consistency of instruction across subjects. The subjects were asked for their initial impressions, that is, they arranged the density levels into their preferred order without having seen the stimuli beforehand. The illumination was adjusted so that the luminance corresponded to that of the stimulus presented on the LCD screen used in experiment 1 (approximately 62 cd/m^2). After the stimuli were arranged into the preferred order, the subjects were asked to motivate why they had placed a certain density level last and another one first.

5.4 *Eye movement recordings*

Eye movements were recorded in parallel with the threshold measurements for experiment 1. No eye movements were recorded for experiments 2 and 3. The equipment used was an SMI (SensoMotoric Instruments Inc.) Eye Link video eye-tracker. The equipment enables the measurement of both eyes by illuminating them with infrared LEDs located in miniature infrared cameras positioned one below each eye. The system has a sampling rate of 250 Hz and is controlled by a separate PC running MS-DOS. It is connected to the computer presenting the stimuli by an Ethernet cable.

Software bundled with the eye-tracking system automatically analyses the eye movements recognising saccades and fixations. A sample belongs to a saccade if one of two criteria is met: acceleration exceeds $9500^o/s^2$ or velocity exceeds

35^{o}/s. If neither of these two criteria is met, the software interprets the sample as belonging to a fixation. Because measurements were made for both eyes, the results represent the mean of the results for the left and the right eye. As with the thresholds, eye data was only recorded after the subjects had made two errors.

The eye movement recordings and the presentation of stimulus were connected so that the recording of eye data began when the stimulus was presented to the subjects. The recordings were stopped either when the subjects clicked a button, moved their eyes away from the matrix to the buttons or when the stimulus presentation was ended and replaced with a mask. This was done in order to minimise recording eye movements after the subject had found the target, which would otherwise affect the results.

6 Results

6.1 Experiment 1

Figure 11 shows the combined threshold search times as a function of inter-element spacing for all subjects. Examining the threshold search times as a function of inter-element spacing shows no significant effects when combined over subjects. More specifically, a Friedman's rank test for k correlated samples (Howell, 2002) shows no significant effects for set sizes 10x10 (χ_F (4) = 9.13, p = .058, ns), 7x7 (χ_F (4) = 3, p = .558, ns) and 5x5 (χ_F (4) = 4.60, p > .331, ns). The test does, however, show an effect for the 3x3 set size (χ_F (4) = 13.52, p < .05). There exist, on the other hand, individual differences between subjects, see Figure 12. The results for subject KL show a systematic tendency of increase at low and high values of icon spacing. The results for the other two subjects do not show similar systematic effects.

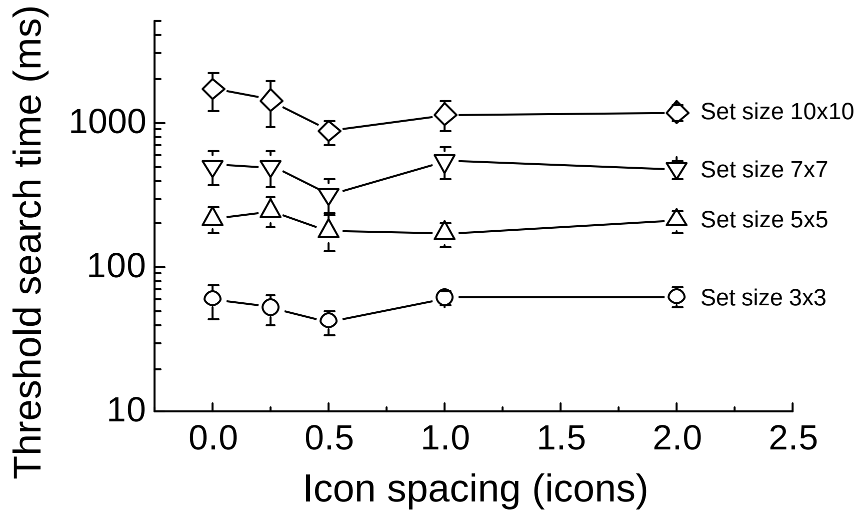

Figure 11. The combined threshold search times of all subjects as a function of icon spacing. The error bars indicate the standard error of the mean.

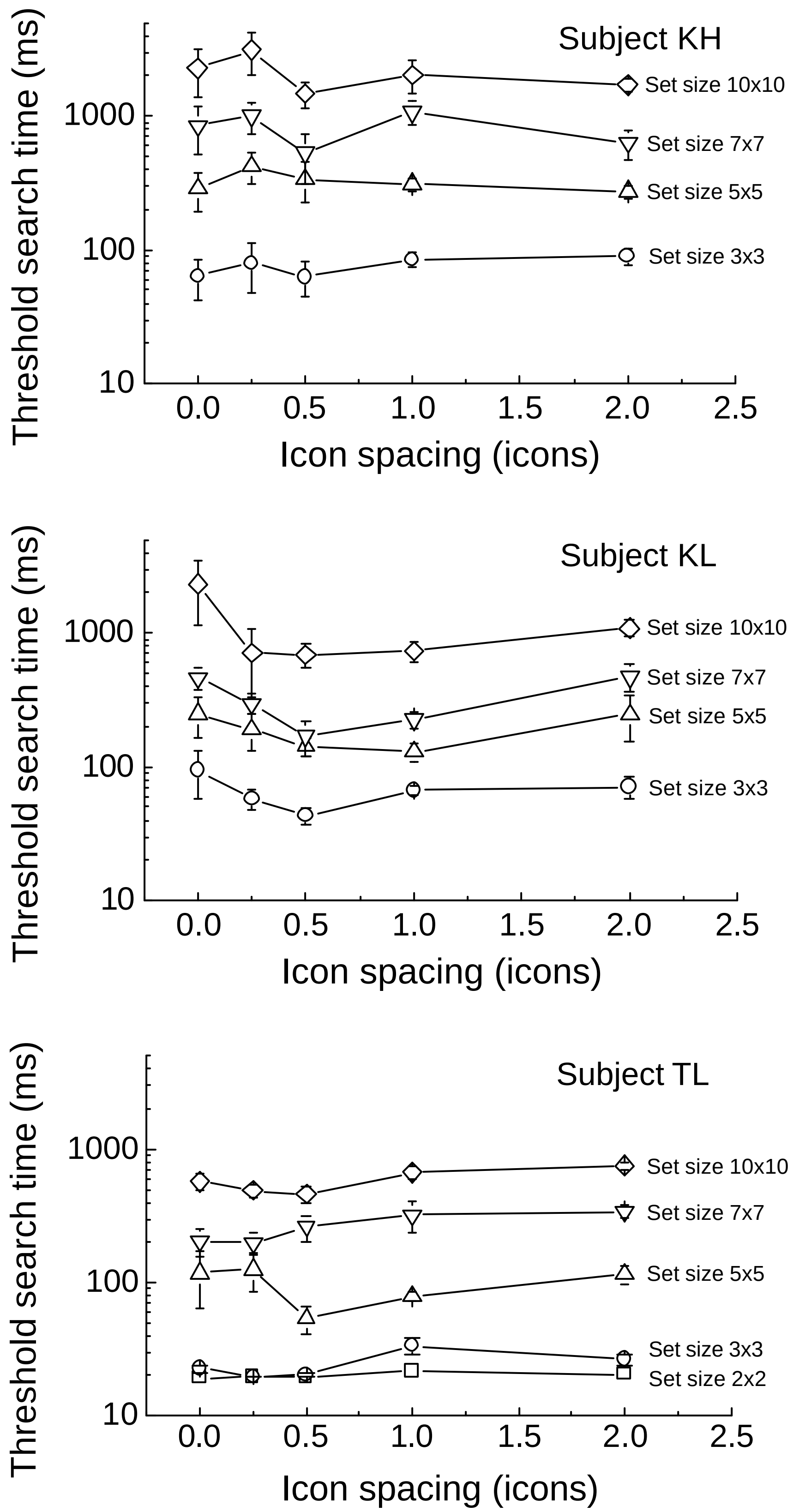

Figure 12. The threshold search times as a function of icon spacing. The error bars indicate the standard error of the mean.

Figure 13 shows the threshold search times of the subjects as a function of set size. The threshold search times can be seen to increase with increasing set size. There are, however, again differences between subjects with subject TL having clearly lower thresholds than subjects KL and KH. The experiment shows, in other words, a clear effect of set size on the threshold search times for all densities.

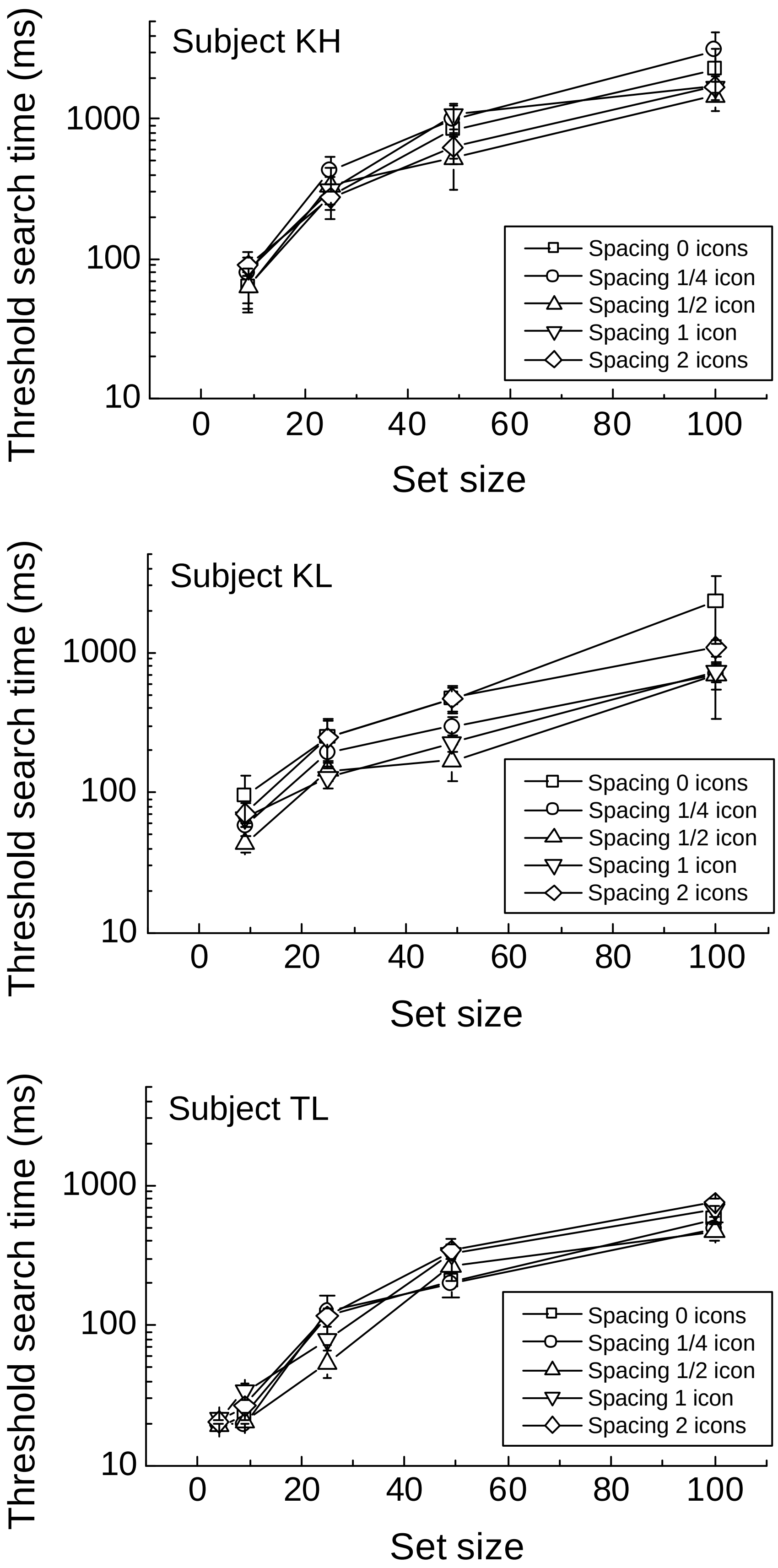

Figure 13. The threshold search times as a function of set size. The error bars indicate the standard error of the mean.

Figure 14 below shows the fixation durations of the subjects for set size 10x10. The fixation duration for the largest set size varies considerably between subjects. Subject TL has the shortest fixations whereas subject KL has the longest, even over twice as long as subject TL. Subject TL shows an increase in fixation duration for the inter-element distance of two icons compared to the other inter-element distances which is in contrast to subjects KH and KL who shown an increase for the inter-element distance of zero icons. Subject KH also shows an increase for the icon spacing ¼ icon.

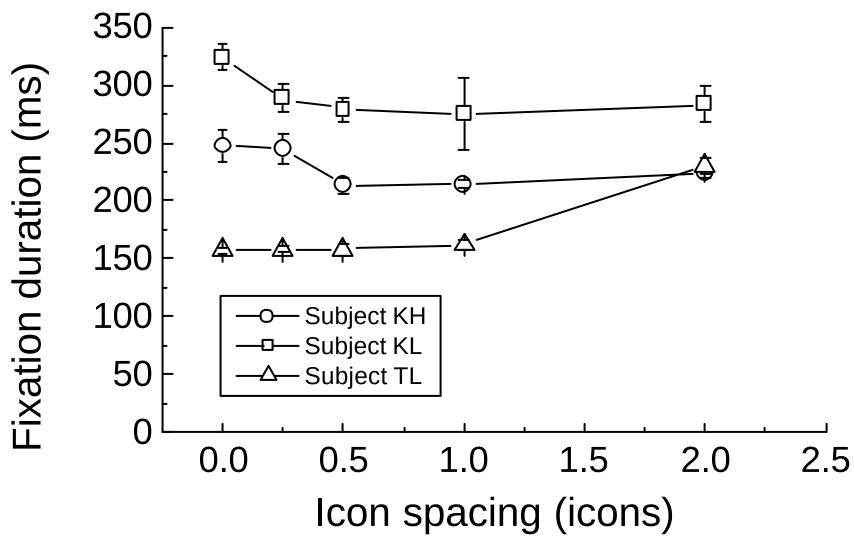

Figure 14. Fixation durations as a function of icon spacing for set size 10x10. The error bars indicate the standard error of the mean.

Figure 15 shows the numbers of fixations for specific inter-element distances as a function of set size. Whereas the threshold search time was already significantly elevated at a set size of 5x5, compared to smaller set sizes, the numbers of fixations show no such results for two out of the three subjects. The one subject, KH, that shows an elevated number of fixations for set sizes 5x5 and larger shows, however, a very moderate elevation for the 5x5 set size. Subject KH requires one fixation for the 3x3 set size and only about 1.5 fixations for set size 5x5. Subject TL, on the other hand, finds the target with only one fixation in set sizes 2x2, 3x3 and 5x5 independent of inter-element spacing. Larger set sizes, however, require more fixations and a difference between different levels of inter-element spacing at a specific set size also becomes evident. Finally, subject KL also finds the target with one fixation at set sizes 3x3 and 5x5. In addition to this, subject KL also requires only one fixation at set size 7x7 for some densities. In other words, there exist differences between subjects as expected, but the set size that can be seen with one fixation seems to be 5x5 on average.

Finally Figure 16 shows the number of fixations as a function of degrees of visual angle. The fact that set sizes of 5x5 (or even more under some circumstances) can be seen with a single fixation independent of element spacing means that the actual area, in degrees of visual angle, that is seen with a single fixation increases as the set size and inter-element spacing increases. The 5x5 set size with an inter-element distance of zero icons for example occupies an area of about 3.5x3.5 degrees squared whereas a set size of 5x5 with inter-element spacing of two icons occupies about 9.5x9.5 degrees squared. Both are still seen with one (or close to one) fixation. On the other hand, the 10x10 matrix with an inter-element distance of zero icons occupies a somewhat smaller area (7.4x7.4 degrees squared) than the 5x5 matrix with the inter-element spacing of two icons (9.5x9.5 degrees squared) but the number of fixations is significantly higher. The results thus show that the number of fixations seems to depend on the number of elements, not the area they cover.

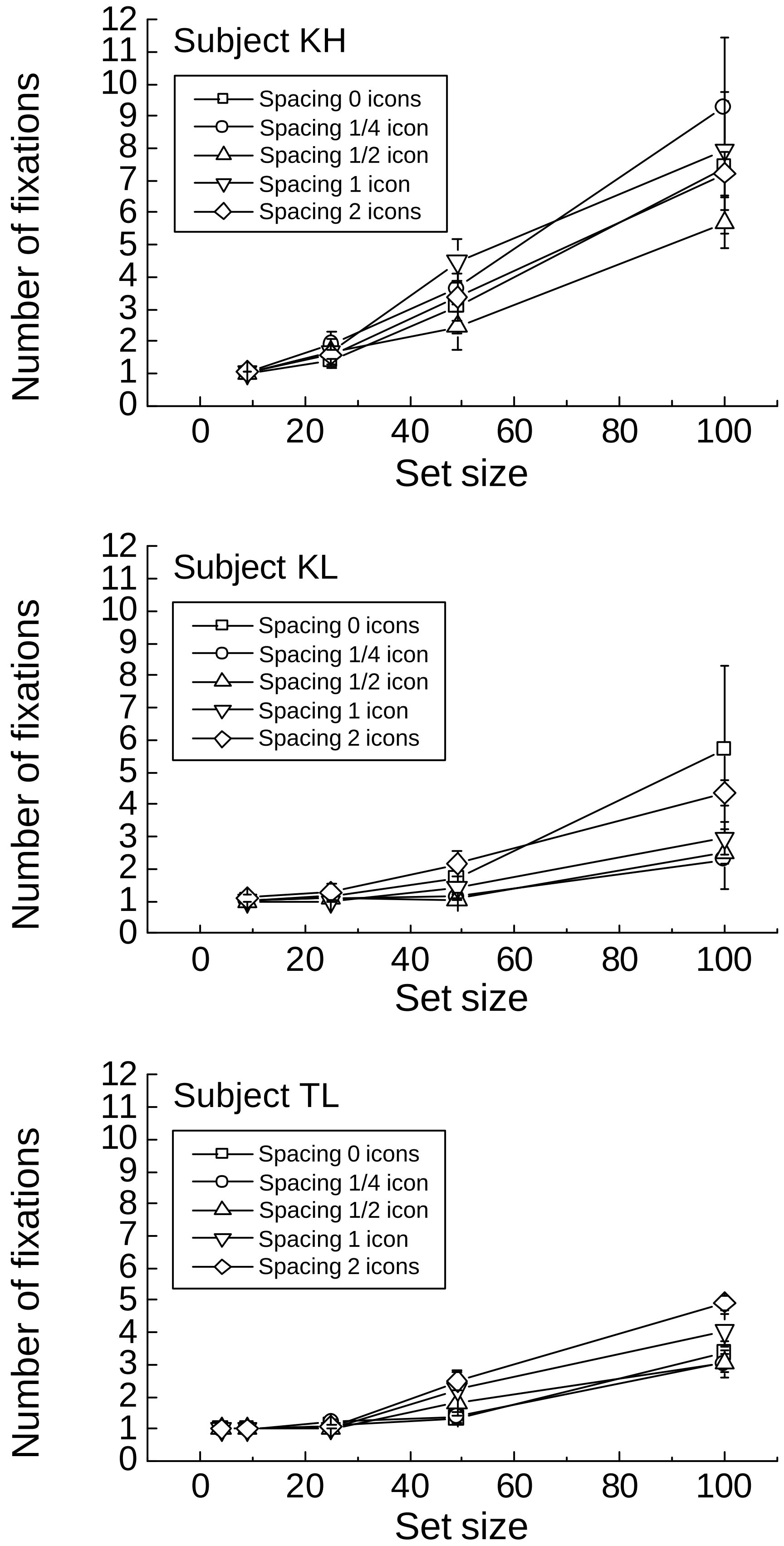

Figure 15. The number of fixations as a function of set size. The error bars indicate the standard error of the mean.

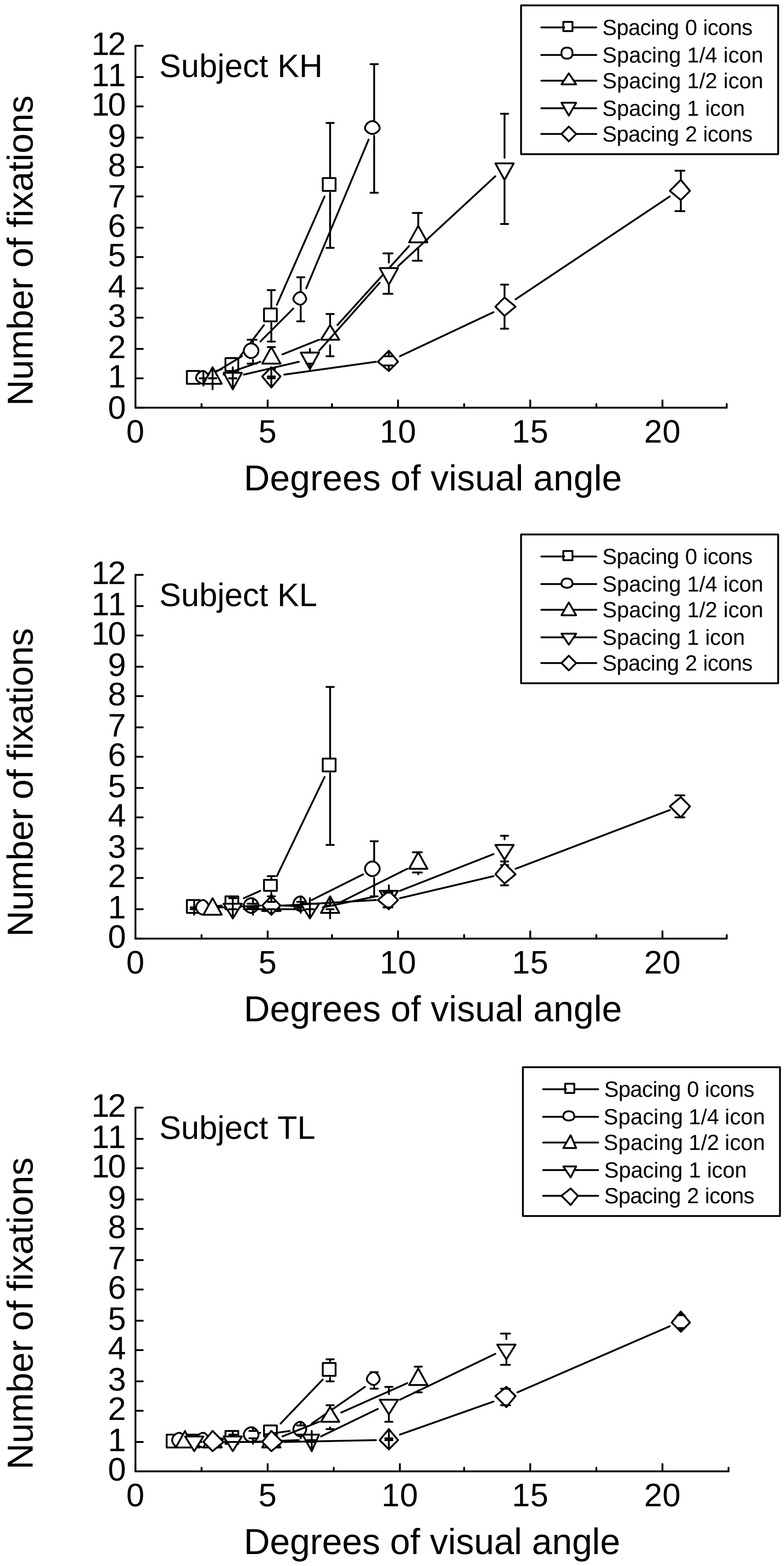

Figure 16. The number of fixations as a function of degrees of visual angle. The error bars indicate the standard error of the mean.

6.2 Experiment 2

As can be seen from Figure 17 the threshold search times increase sharply as the icon size decreases from about 0.4^o to about 0.2^o. Also the increase in the threshold search time from an icon size of around 0.7^o to about 0.4^o is clear although not nearly as drastic. The change in threshold search time from icon size 1.5^o to 0.7^o is very small. In other words, as the icon size decreases from about 0.7^o, the threshold search times increase for all subjects. As in experiment 1, small differences exist between subjects.

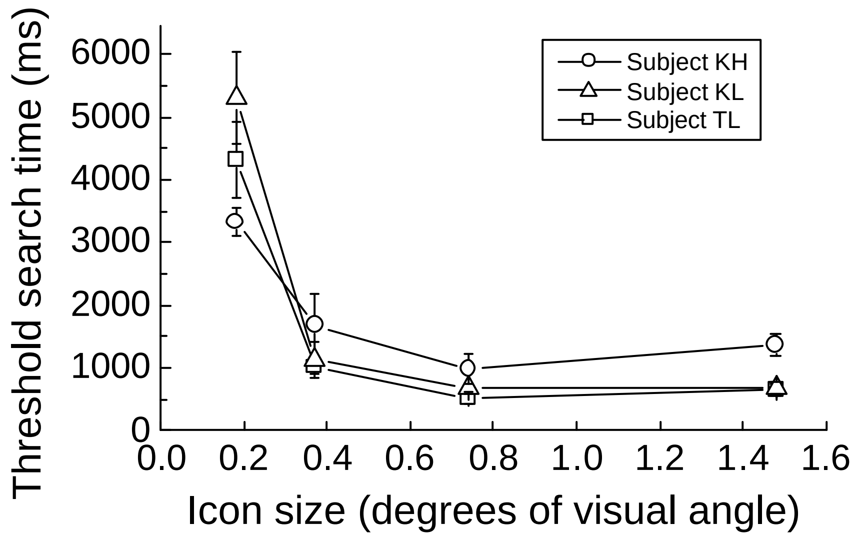

Figure 17. The effect of icon size on the threshold search time. The error bars indicate the standard error of the mean.

6.3 Experiment 3

Table 1 shows the rankings of the density levels for the different subjects. A Friedman's rank test for k correlated samples (Howell, 2002) performed on the ranking of the density levels, Equation 1, reveals that the distribution of rankings is not random but depends on the level of density.

The equation for the Friedman's rank test can be written as (Howell, 2002):

$$c_F^2 = \frac{12}{Nk(k+1)} \sum_{i=1}^{k} R_i^2 - 3N(k+1) \qquad (1)$$

where N is the number of subjects, k the number of density levels and R_i the sum of the ranks for the different density levels. Entering the values into

Equation 1 gives $c_F^2 = 35.6$ on 4 *df*. Comparing this to $c_{0.001}^2(4) = 18.467$ H_0, the null hypothesis that the rankings are randomly distributed, can be rejected and as mentioned above, the rankings depend on the density level.

Table 1. The ranking of the different icon spacing levels by subjects.

	Icon spacing				
Subjects	*0 icons*	*¼ icon (8 pixels)*	*½ icon (16 pixels)*	*1 icon (32 pixels)*	*2 icons (64 pixels)*
Subject 1	5	4	2	1	3
Subject 2	5	4	3	1	2
Subject 3	5	4	3	1	2
Subject 4	5	4	2	1	3
Subject 5	5	3	2	1	4
Subject 6	5	4	3	1	2
Subject 7	5	4	2	1	3
Subject 8	5	4	3	1	2
Subject 9	5	3	2	1	4
Subject 10	5	4	3	1	2
Subject 11	5	1	2	3	4
Subject 12	5	4	2	3	1
Subject 13	5	1	2	3	4
Subject 14	4	2	1	3	5
Sum	69	46	32	22	41

From the ranking sums it can be seen that the one icon spacing is the most preferred while the zero icon spacing is clearly the least preferred. Following the one icon spacing the next best density level, measured by the sum of the rankings, is the ½ icon spacing. The last two icon spacing levels, ¼ icon and two icons, are quite closely matched. The two icon spacing level is judged to be only marginally better than the ¼ icon spacing level.

Turning from the ranking to the most and least preferred density levels measured by the number of first and last places respectively. Figure 18 below shows that of the fourteen subjects ten held the one icon spacing as the most preferred with an icon spacing of ¼ icon getting two votes and both ½ icon and two icons getting one vote.

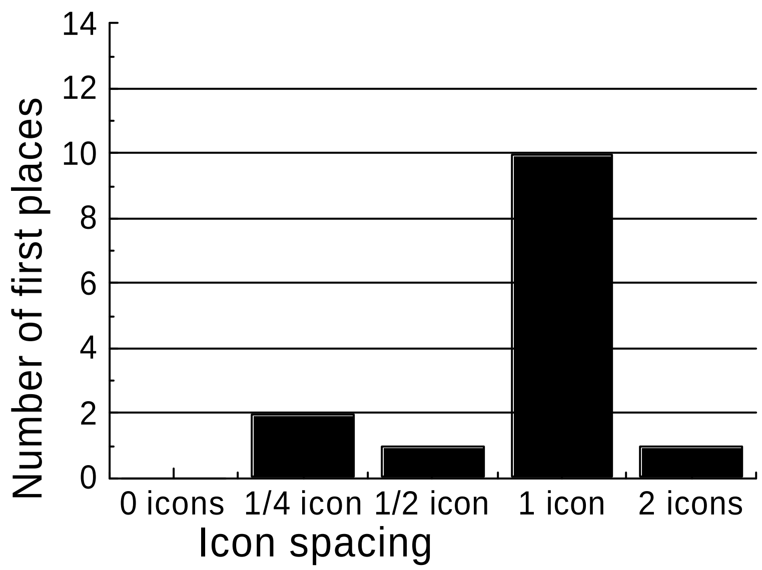

Figure 18. The icon spacing most preferred by the users.

The motivations for selecting the icon spacing of one icon, as the most preferred were that "all the information can be seen at a glance" and that "the icons were adequately spaced resulting in it being easy to find targets". "They were easily distinguishable". One comment about the density level was that "it was manageable and easily perceived". The image was in one word, clear. The same comments apply even to the other icon spacing levels that were reported to be most preferred with the additional comment on the ¼ spacing of it being "calm, not uneasy or nervous". It would seem these are the attributes of a good layout, subjects just find them in slightly different levels of density.

As can be seen from Figure 19, thirteen of fourteen subjects found an icon spacing level of zero icons the least preferable. The only other icon spacing to be preferred the least is 2 icons.

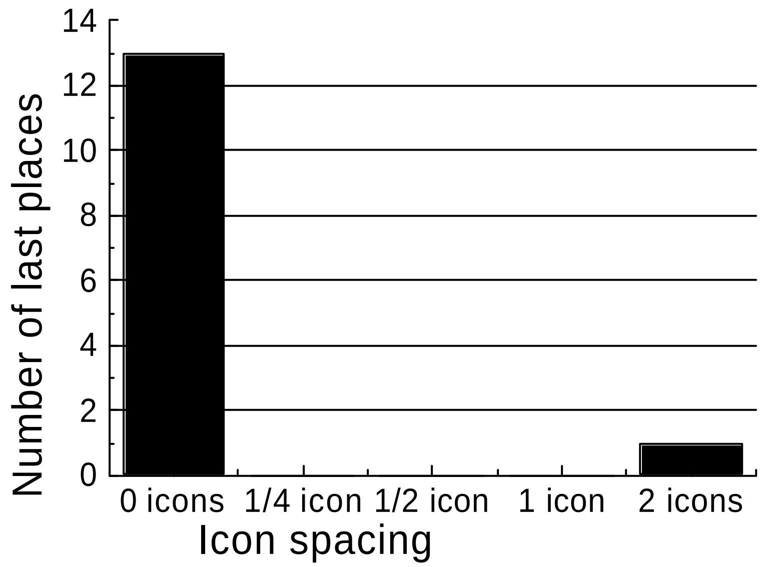

Figure 19. The icon spacing least preferred by the subjects.

Because the density levels preferred the least are the ones from both extremes, the comments for why they were not liked do not match. The subject who found a spacing of two icons to be the worst said the reason to be that “they were spaced too far apart, the visual field became too large”. The subjects reporting the zero icon spacing level to be the least preferred to a large extent commented on it being “too small and crowded”. “The individual icons became hard to find”, “they were getting mixed together” and “it was difficult to tell where one ended and the next began”. One subject said it was “oppressive“ and one said she “got irritated looking at it” and said it was “uneasy”.

7 Summary of findings

The objective of the study was to come up with guidelines for the use of density and element size in the design of interfaces. One objective was also to define the size of the perceptual span for icons.

Experiment 1 showed that the threshold search time for icons increased with increasing set size. When the threshold was plotted as a function of density, however, there was no trend. The smaller arrays (2x2, 3x3 and 5x5) showed no impact of density on an individual level whereas the larger arrays (7x7 and 10x10) showed only a small increase in threshold as the inter-element spacing increased. Independent of the inter-element distances the smaller arrays only needed one fixation whereas the larger required more. The experiment also showed that the perceptual span for icons seems to be limited by attention, not by target eccentricity. The actual size of the perceptual span for icons was found to be 5x5.

Experiment 2 measured the effect of element size on performance. The results of the experiment were quite clear, as the element size decreased the threshold search time increased. The critical size from which point on search performance is considerably degraded was found to be about 0.7^{o}. For icons smaller in size that this the importance of good design is further emphasised, simpler is better.

Because efficiency is only one aspect of usability, **experiment 3** investigated the subjective preferences and aesthetic impressions of users on the density of icons. The results of experiment 3 were also quite clear-cut. Based on the results the inter-element spacing not to be used is zero icons and the one to be used is one icon.

8 Discussion

8.1 About the results

Experiment 1 showed that the threshold search times increase as the set size increases. This is in line with the theory by Treisman and Gelade (1980), which states that search for a conjunction of features is serial. The target icon could not be detected based on any one single feature, in other words it did not pop-out but required instead the deployment of attention before it could be found from among the distractors.

The threshold search times were shown by the Friedman's rank test for k correlated samples (Howell, 2002) not to depend on inter-element spacing with one exception, the 3x3 set size. The results of the experiment would thus indicate that crowding (Bouma, 1970) does not play a part when icons are concerned. It is, however, possible that larger inter-element distances would have a negative effect on search times but the results by Vlaskamp et al. (2002) would indicate otherwise. The dependency shown for the 3x3 set size can, however, be explained by the very small variations compared to the other set sizes. Furthermore, even though the threshold search time was shown to depend on the inter-element spacing for that set size, the actual value of the threshold varies very little and the number of fixations at the set size is one for all subjects independent of inter-element spacing. Due to these reasons the dependence shown for that set size has no significance in practice.

The arrays which on an individual level were unaffected by the change in density only required one fixation to tell whether the target was present or not. The two large arrays on the other hand required more than one fixation. Additionally, regarding the results by subject KL showing an increase in threshold search times for decreasing inter-element distance, they can to a large extent be explained by the learning effect. The experiments were set-up so that the subjects started with set sizes with an inter-element spacing of zero icons, which is why the results by subject KL show rising curves at high densities. As mentioned previously, subject KL was not practised. As a result of this, the only actual trend shown on an individual level for threshold search times, as a function of inter-element spacing is a small increase as the inter-element distances increase for set sizes 7x7 and 10x10.

Returning to the study by Vlaskamp et al. (2002), the results obtained in this experiment do not support it. Vlaskamp et al. found that the search time per element increased as the inter-element distance decreased below 1.5^o. In the current experiment all inter-element distances were smaller than 1.5^o with the smallest being 0^o but no decrease in performance was found as inter-element distances decreased. The reason for the difference in results might be the completely different stimuli used. Whereas Vlaskamp et al. used Landolt 'C's the current experiment used considerably more complex icons which may have been easier to discriminate due to larger variations among them. The results do, however, corroborate the finding by Motter and Belky (1998) that the perceptual span could have a diameter of five elements. The study by Motter and Belky was performed on rhesus monkeys and is thus not necessarily directly compatible with results obtained by human subjects.

The important finding regarding the perceptual span for icons is that it seems to be limited by attention, not by target eccentricity. The limit in number of elements where one fixation still was enough was on average about 25 arranged in a 5x5 array independent of inter-element distance. Larger arrays, that is, more elements, could cover a smaller area, in degrees of visual angle, than a 5x5 array but still require more than one fixation.

Another interesting finding that the eye movement recordings reveal is a difference in search strategies between the subjects. Subject TL for example uses many short fixations whereas subject KL uses few long fixations. The differences in performance for the 10x10 set size for subjects KL and TL are not very different, however, indicating that different search strategies may be equally efficient.

In **experiment 2** the reason for the decreased search performance as the size of the icons decreased might be the low-pass filtering performed by the human visual system. As a result of this any high-frequency information (fine detail) in the icons is filtered away and the discrimination of the icons has to be made based on low-frequency information (areas with low detail). As a result of this the icons used in the experiment become more difficult to tell apart, resulting in longer search times.

From this it follows that small icons should be designed separately from large ones. Whereas large icons may utilise more high spatial frequency information (more fine detail) small icons should be designed so that they are easy to recognise based on low spatial frequency information (low detail).

The experiment only used icons as large as 1.48^{o} and can thus not draw conclusions about the possible effect on search time of even larger elements. However, an icon size of 1.48^{o} is already twice as large as the normal icon size on a Windows desktop viewed at normal distance. Taking this into consideration, the use of even larger elements in applications designed for the average user is unlikely. Applications designed for devices with small displays are on the contrary more likely to use even smaller icons in order to save space.

The implications of the results obtained in the experiment are that the size of the elements used in interfaces is important and can have a highly significant impact on performance. This fact should be taken into consideration in the design of applications. Likewise, the spatial frequencies of the icons do have a difference. Small icons should use low spatial frequencies relative to icon size, or in other words, small icons should be simple.

From the results of **experiment 3** it is very clear that the inter-element distance of zero icons is not suitable for use in applications if people’s subjective opinions are taken into account. Put simply, people do not seem to like it when the interface is too crowded and complex. The display becomes nervous and even oppressive to some.

The inter-element distance that is preferred by the subjects is one icon. If for some reason the one icon spacing is not usable the results reported above suggest that the next best icon spacing is ½ icon. These alternatives seem to be best balanced between a too crowded display with too much information and a sparse one where the interface elements are too far apart and the interface contains too little information.

One thing that must be taken into account, however, is that as mentioned in the methods section the subjects were asked for their initial impressions of the density levels. Given this, the results might be significantly different if the subjects had been using actual applications with the different density levels. They might for example have found levels that initially looked good to be inefficient in use and vice versa.

The results of experiment 1, however, showed that inter-element spacing does not have an effect on threshold search times. This means that there is no disagreement between the results of experiment 1 and the subjective impressions of people. In other words, independent of what inter-element spacing the user prefers, it does not have a negative effect on efficiency of use measured in threshold search time. Ideally then, users should be able to modify the inter-element spacing to their own liking.

8.2 In general

The set-up of the experiments does not reflect real use situations. The subjects were wearing eye-tracking equipment and the viewing position was restricted. Also the fact that the subjects were being measured might have some effect. All these effects would, however, probably be equal for all measurements and thus not affect any possible trends. Without eye movement recordings, differences in search strategies would not have been discovered.

The icons used in the experiments were set up in symmetrical arrays (see appendix 1 for examples). The set-up can be considered to represent a fairly typical situation. Oftentimes the icons on the desktop, the mobile phone or PDA are set up in this way by default. The actual icons themselves were real icons found on a computer running the Windows2000 operating system.

Despite these restrictions on the authenticity of the experiment set-up the results obtained can be used to advantage in real life applications and situations.

Even though the study used icons as stimuli the results are applicable to other forms of graphical interface elements as well. The guidelines listed in the section below should thus be used every time a new interface is designed or an existing one is modified. The guidelines can be seen as a complement to the ISO 9241-12 standard (ISO 9241-12) for information presentation and they state how the density and size of interface elements is to be taken into consideration when designing graphical user interfaces. Following the guidelines should result in applications that are more efficient to use while retaining a high level of user satisfaction. The results obtained in the study are also applicable to all areas of interface design ranging from applications controlling complex industrial processes to mobile applications used occasionally by novices.

The results obtained in the experiments are all usable as such on their own. Experiment 1 defined the size of the perceptual span, which can be used in the design of interfaces. It also showed that density does not have an effect on search times. The optimum element size can be read from the results of the second experiment and subjective impressions of humans from the third experiment. The best results are, however, undoubtedly obtained when all these results are viewed together as a whole.

8.3 Further research

The subjects participating in experiments 1 and 2 were fairly young, 25 and 26 years old. The experiments could be performed with older subjects to see whether the results generalise over different age groups.

The experiment investigating the effect of size on search performance used normal 32x32 pixel icons designed for the Microsoft Windows2000 operating system for use on a normal computer screen. These icons contained much high-frequency information. The results showed a significant decrease in performance as the size of the icons decreased. Further research could investigate whether the decrease in performance is independent of the icons used or if small icons can be used as long as they are well designed. A new set of smaller icons could be designed that only use low-frequency information. Using these icons the search performance at normal viewing distance could be evaluated and compared to the results obtained in this study.

The experiment investigating the effect of size only used one level of density and set size, or vice versa, the experiment investigating the effect of density only used one level of element size. The combination of different set sizes, inter-element distances and element sizes could also be investigated to see if they reveal any new phenomena.

9 Conclusions

Up to now no guidelines or standards for the density or size of interface elements have existed. For example the ISO 9241-12 standard (ISO 9241-12) for information presentation only states that ‘information should be located to meet user expectations and task requirements’. Furthermore it says that different states of graphical objects should be indicated in some way and that if identical icons are used a unique text label should differentiate them.

Based on the results of the experiments, however, guidelines can now be given to interface designers on the density and size of icons (interface elements) used in graphical user interfaces:

- The preferred spacing to use between icons (interface elements) is ½ - one icon (interface element) width.
- Icons (interface elements) should not be placed next to each other without any spacing between them.
- The minimum size of icons (interface elements) should not be less than 0.7 degrees of visual angle, which corresponds to about 0.5 cm at a viewing distance of 40 cm, and about 0.9 cm at a viewing distance of 70 cm. If smaller icons (interface elements) need to be used they should be designed carefully using mostly low spatial frequency information (low level of detail).
- The perceptual span for icons is 5x5, that is, they can be seen with one fixation. This knowledge can be used to minimise the number of fixations for an interface.

10 References

Apple Computer. (2002). Aqua human interface guidelines. Retrieved 1.11.2002 from http://developer.apple.com/techpubs/macosx/Essentials/AquaHIGuidelines/index.html

Apple-History. (2002). The Lisa / Lisa 2 / Mac XL. Retrieved 1.11.2002 from http://www.apple-history.com/lisa.html

Benbasat, I., & Todd, P. (1993). An experimental investigation of interface design alternatives: Icon vs. text and direct manipulation vs. menus. *International Journal of Man-Machine Studies, 38,* 369 – 402.

Bewley, W.L., Roberts, T.L., Schroit, D., & Verplank, W.L. (1983). Human factors testing in the design of Xerox's 8010 "Star" office workstation. *Proceedings of 1983 Conference on Human Factors in Computing Systems,* 72 – 77.

Blankenberger, S., & Hahn, K. (1991). Effects of icon design on human-computer interaction. *International Journal of Man-Machine Studies, 35,* 363 – 377.

Bouma, H. (1970). Interaction effects in parafoveal letter recognition. *Nature, 226,* 177 – 178.

Byrne, M.D. (1993). Using icons to find documents: Simplicity is critical. *Proceedings of the Conference on Human Factors in Computer Systems,* 446 – 453.

Chung, S.T.L., Levi, D.M., & Legge, G.E. (2001). Spatial-frequency and contrast properties of crowding. *Vision Research, 41,* 1833 – 1850.

Ciuffreda, K.J., & Tannen, B. (1995). *Eye movement basics for the clinician.* St. Louis, Missouri, USA: Mosby.

Curcio, C.A., & Allen, K.A. (1990). Topography of ganglion cells in human retina. *The Journal of Comparative Neurology, 300,* 5 – 25.

Duncan, J., & Humphreys, G.W. (1989). Visual search and stimulus similarity. *Psychological Review, 96,* 433 – 458.

Gittins, D. (1986). Icon-based human-computer interaction. *International Journal of Man-Machine Studies, 24,* 519 – 543.

Goonetilleke, R.S., Shih, H.M., On, H.K., & Fritsch, J. (2001). Effects of training and representational characteristics in icon design. *International Journal of Human Computer Studies, 55,* 741 – 760.

Guastello, S.J., Traut, M., & Korienek, G. (1989). Verbal versus pictorial representations of objects in a human-computer interface. *International Journal of Man-Machine Studies, 31,* 99 – 120.

Hemenway, K. (1982). Psychological issues in the use of icons in command menus. *Proceedings of the First Major Conference on Human Factors in Computer Systems,* 20 – 23.

Hooge, I.T.C., & Erkelens, C.J. (1999). Peripheral vision and occulomotor control during visual search. *Vision Research, 39,* 1567 – 1575.

Howell, D.C. (2002). *Statistical methods for psychology (fifth ed.).* Pacific Grove, USA, Duxbury.

Huang, S., Shieh, K., & Chi, C. (2002). Factors affecting the design of computer icons. *International Journal of Industrial Ergonomics, 29,* 211 – 218.

ISO 9241-12. Ergonomic requirements for office work with visual display terminals (VDTs) – Part 12: Presentation of information.

Jacobs, A.M. (1986). Eye-movement control in visual search: How direct is visual span control? *Perception and Psychophysics, 39,* 47 – 58.

Jenkins, S.E., & Cole, B.L. (1982). The Effect of the density of background elements on the conspicuity of objects. *Vision Research, 22,* 1241 – 1252.

Legge, G.E., Pelli, D.G., Rubin, G.S., & Schleske, M.M. (1985). Psychophysics of reading – 1. Normal vision. *Vision Research, 25,* 239 – 252.

Legge, G.E., Rubin, G.S., & Luebker, A. (1987). Psychophysics of reading – V. The role of contrast in normal vision. *Vision Research, 27,* 1165 – 1177.

McLeod, P., Driver, J., & Crisp, J. (1988). Visual search for a conjunction of movement and form is parallel. *Nature, 332,* 154 – 155.

Microsoft Corporation. (2002). Creating Windows XP icons. Retrieved 1.11.2002 from http://msdn.microsoft.com/library/default.asp?url=/library/en-us/dnwxp/html/winxpicons.asp

Motter, B.C., & Belky, E.J. (1998). The zone of focal attention during active visual search. *Vision Research, 38,* 1007 – 1022.

Moyes, J. (1994). When users do and don't rely on icon shape. *Proceedings of the CHI '94 Conference Companion on Human Factors in Computing Systems,* 283 – 284.

Nakayama, K., & Silverman, G.H. (1986). Serial and parallel processing of visual feature conjunctions. *Nature, 320,* 264 –265.

Niemelä, M., & Saarinen, J. (2000). Visual search for grouped versus ungrouped icons in a computer interface. *Human Factors, 42,* 630 – 635.

Nokia Corporation. (2002). Nokia Series 60 UI style guide. Available from Forum Nokia web site, http://www.forum.nokia.com

Näsänen, R., Karlsson, J., & Ojanpää H. (2001). Display quality and the speed of visual letter search. *Displays, 22,* 107 – 113.

Näsänen, R., & Ojanpää, H. (2002). How many faces can be processed during a single eye fixation? (P95). *11th European Conference on Eye Movements,* August 22-25, 2001, Turku, Finland. (Abstract).

Näsänen, R., Ojanpää, H., & Kojo, I. (2001). Effect of stimulus contrast on performance and eye movements in visual search. *Vision Research, 41,* 1817 – 1824.

Ojanpää, H., Näsänen, R., & Kojo, I. (2002). Eye movements in the visual search of word lists. *Vision Research, 42,* 1499 – 1512.

Palo Alto Research Center Incorporated. (2002). PARC history. Retrieved 1.11.2002 from http://www.parc.com/company/history/

Pollatsek, A., Raney, G.E., Lagasse, L., & Rayner, K. (1993). The use of information below fixation in reading and visual search. *Canadian Journal of Experimental Psychology, 47,* 179 – 200.

Prinz, W. (1984). Attention and sensitivity in visual search. *Psychological Research, 45,* 355 – 366.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124,* 372 – 422.

Rayner, K., & Fisher, D.L. (1987). Eye movements and the perceptual span during visual search. In J.K. O'Reagan, & A. Levý-Schoen (Eds.), *Eye movements: From Physiology to Cognition* (pp. 293 – 302). Netherlands: Elsevier Science Publishers B.V.

Repokari, L., Saarela, T., & Kurkela, I. (2002). Visual search on a mobile phone display. SAICSIT 2002 Enablement Through Technology Annual Conference of the South African Institute of Computer Scientists & Information Technologists

Scialfa, C.T., & Joffe, K.M. (1998). Response times and eye movements in feature and conjunction search as a function of target eccentricity. *Perception & Psychophysics, 60,* 1067 – 1082.

Sireteanu, R., & Rettenbach, R. (2000). Perceptual learning in visual search generalizes over tasks, locations and eyes. *Vision Research, 40,* 2925 – 2949.

Staggers, N. (1993). Impact of screen density on clinical nurses' computer task performance and subjective screen satisfaction. *International Journal of Man-Machine Studies, 39,* 775 – 792.

Tanskanen, T. (1999). *Visuaalinen haku graafisessa käyttöliittymässä.* Pro gradu, University of Helsinki, Helsinki.

Theeuwes, J., & Kooi, F.L. (1994). Parallel search for a conjunction of contrast polarity and shape. *Vision Research, 34,* 3013 – 3016.

Treisman, A. (1982). Perceptual grouping and attention in visual search for features and objects. *Journal of Experimental Psychology: Human Perception and Performance, 8,* 194 – 214.

Treisman, A. (1986). Features and objects in visual processing. *Scientific American 254,* 106 – 115.

Treisman, A. (1998). Feature binding, attention and object perception. *Philosophical Transactions of the Royal Society of London – Series B: Biological Sciences, 353,* 1295 – 1306.

Treisman, A.M., & Gelade G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12,* 97 – 136.

Treisman, A., & Gormican, S. (1988). Feature analysis in early vision: evidence from search asymmetries. *Psychological Review, 95,* 15 – 48.

Treisman A., & Souther J. (1985). Search asymmetry: A diagnostic for preattentive processing of separable features. *Journal of Experimental Psychology: General*, *114,* 285 – 310.

Vlaskamp, B., Hooge, I., & Over, E. Stimulus density does not affect saccadic search performance. *Perception, 31,* supplement. 25th European Conference on Visual Perception, Glasgow, 25 – 29 August, 2002. (Abstract).

Wertheim, T.H. (1894). Über die indirekte Seheschärfe. *Zeitschrift für Psychologie und Physiologie der Sinnesorganen, 7,* 172 – 187. Translated by Dunsky, I.L. (1980). Peripheral visual acuity. *American Journal of Optometry & Physiological Optics, 57,* 915 – 924.

Wetherill, G.B., & Levitt, H. (1965). Sequential estimation of points on a psychometric function. *The British Journal of Mathematical and Statistical Psychology, 18,* 1 – 10.

Wickens, C.D., Gordon, S.E., & Liu, Y. (1998). *An introduction to human factors engineering.* New York: Addison-Wesley Educational Publishers Inc.

Wiedenbeck, S. (1999). The use of icons and labels in an end user application program: an empirical study of learning and retention. *Behaviour & Information Technology, 18,* 2, 68 – 82.

Wolfe, J.M. (1994a). Guided search 2.0 a revised model of visual search. *Psychonomic Bulletin & Review, 1,* 202 – 238.

Wolfe, J.M. (1994b). Visual search in continuous, naturalistic stimuli. *Vision Research, 34,* 1187 – 1195.

Wolfe, J.M. (1998a). Visual memory: What do you know about what you saw? *Current Biology, 8,* 303 – 304.

Wolfe, J.M. (1998b). Visual search. In H. Pashler (Ed.), *Attention* (pp. 13 – 73). East Sussex, UK: Psychology Press Ltd.

Wolfe, J.M. (1998c). What can 1 million trials tell us about visual search? *Psychological Science, 9,* 33 – 39.

Wolfe, J.M., Cave, K.R., & Franzel, S.L. (1989). Guided search: An alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 419 – 433.

Wolfe, J.M., & Gancarz, G. (1996). Guided search 3.0: A model of visual search catches up with Jay Enoch 40 years later. In V. Lakshminarayanan (Ed.), *Basic and Clinical Applications of Vision Science,* (pp. 189 – 192). Dordrecht, Netherlands: Kluwer Academic.

Ximian Inc. (2001). Gnome icons. Retrieved 1.11.2002 from http://developer.ximian.com/articles/tutorials/icons/output/

Appendix 1: Icon matrices

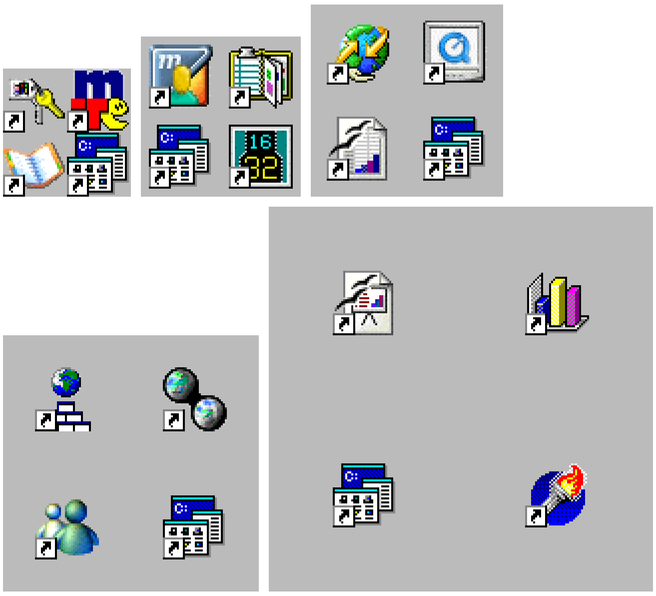

Figure 20. 2x2 matrices.

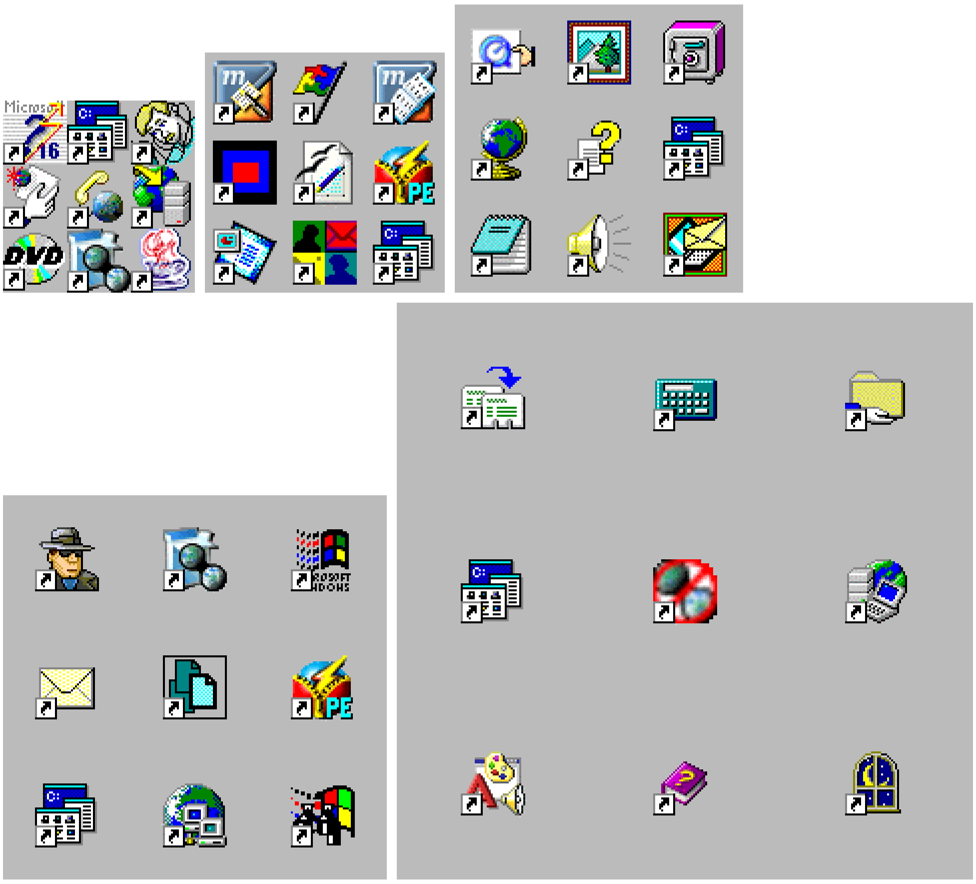

Figure 21. 3x3 matrices.

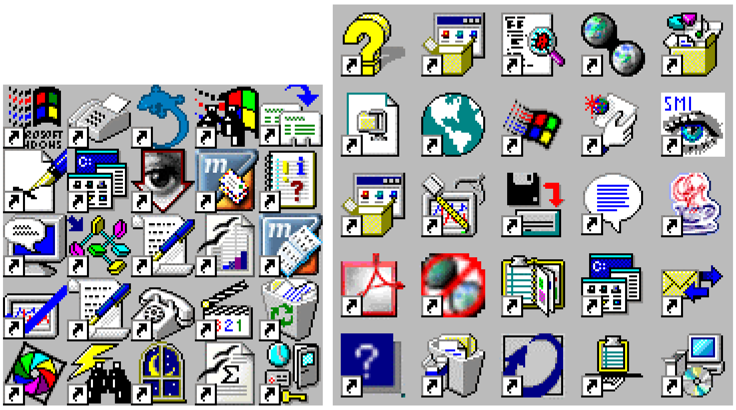

Figure 22. 5x5 matrices with inter-element distances of 0 icons and ¼ icon.

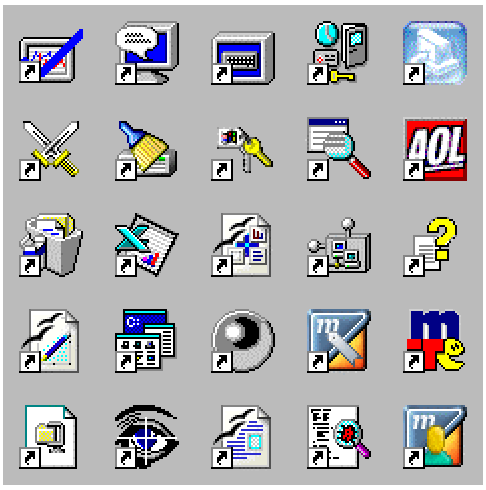

Figure 23. 5x5 matrix with inter-element distance of ½ icon.

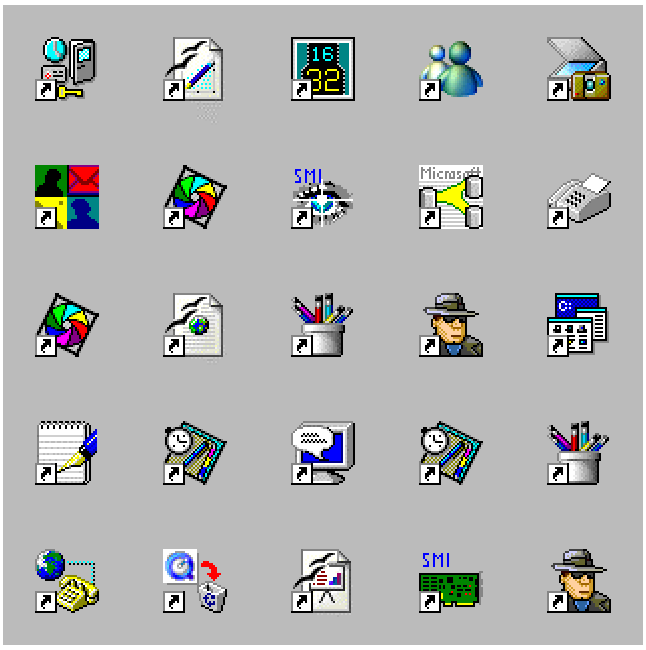

Figure 24. 5x5 matrix with inter-element distance of 1 icon.

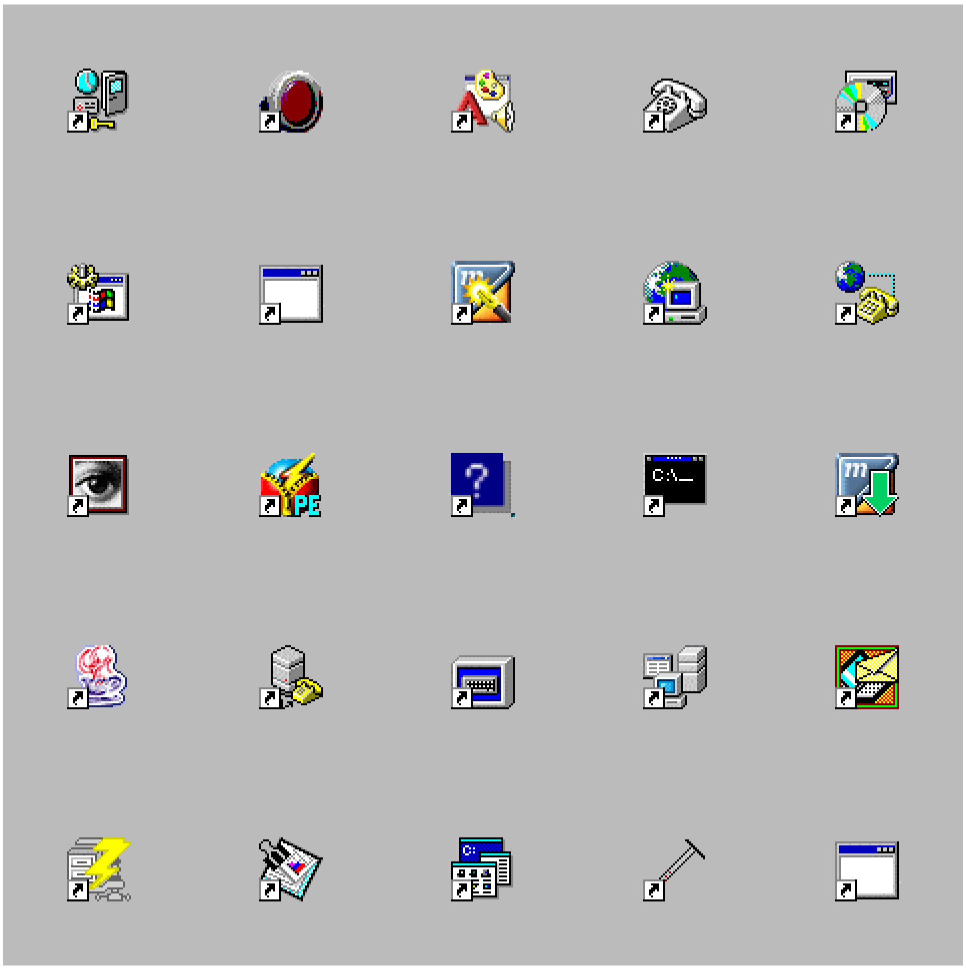

Figure 25. 5x5 matrix with inter-element distance of 2 icons.

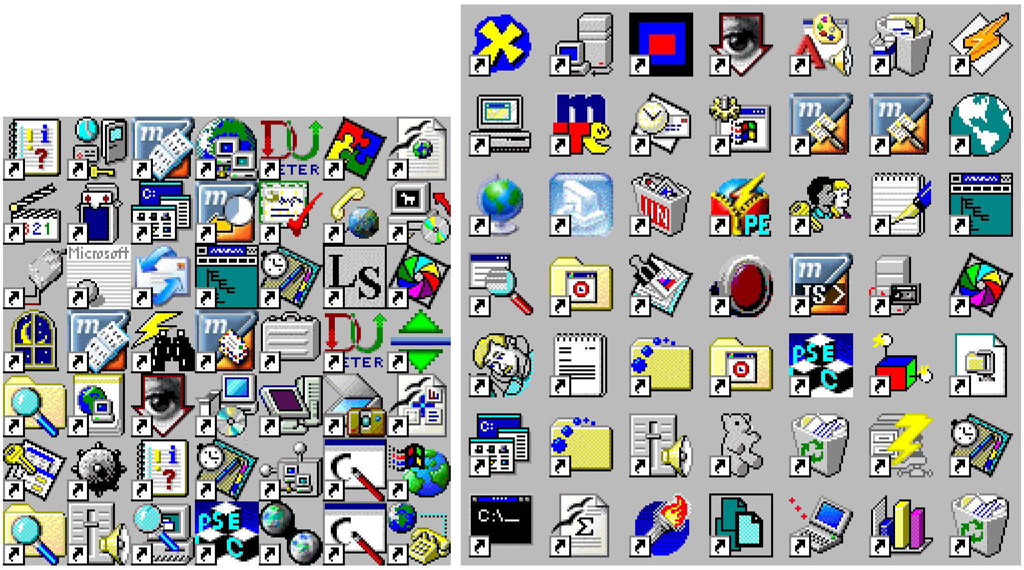

Figure 26. 7x7 matrices with inter-element distances of 0 icons and ¼ icon.

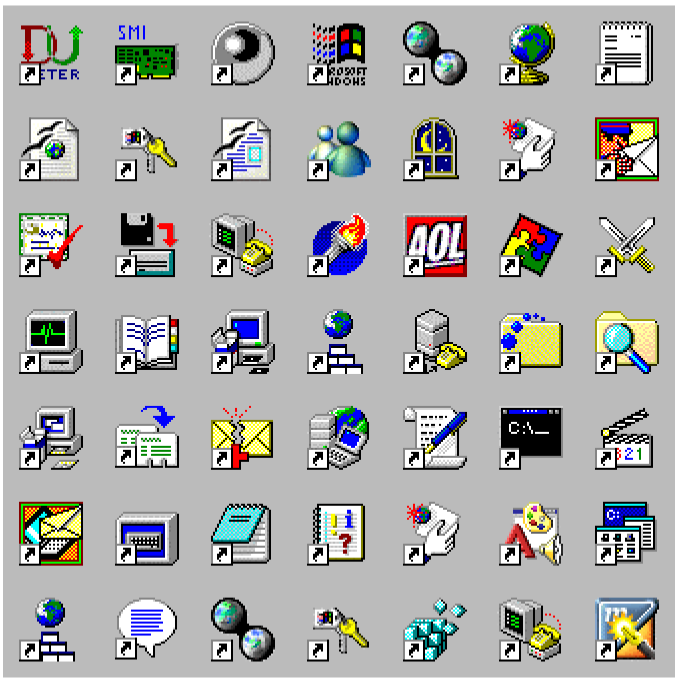

Figure 27. 7x7 matrix with inter-element distance of ½ icon.

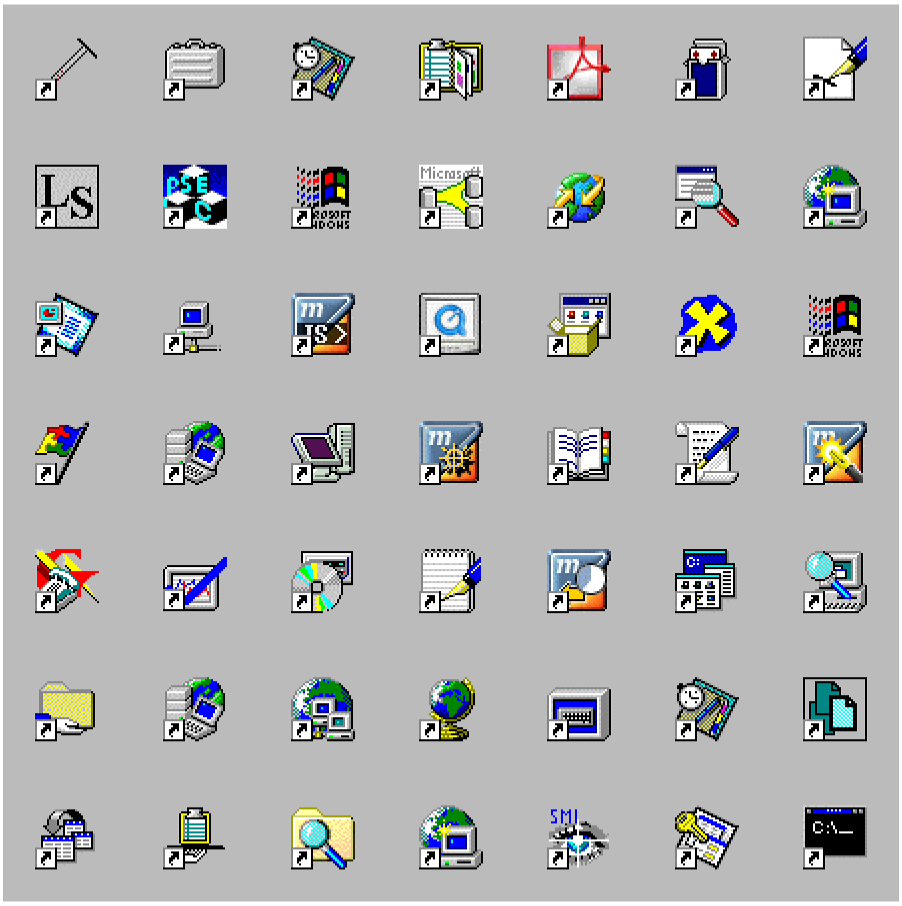

Figure 28. 7x7 matrix with inter-element distance of 1 icon.

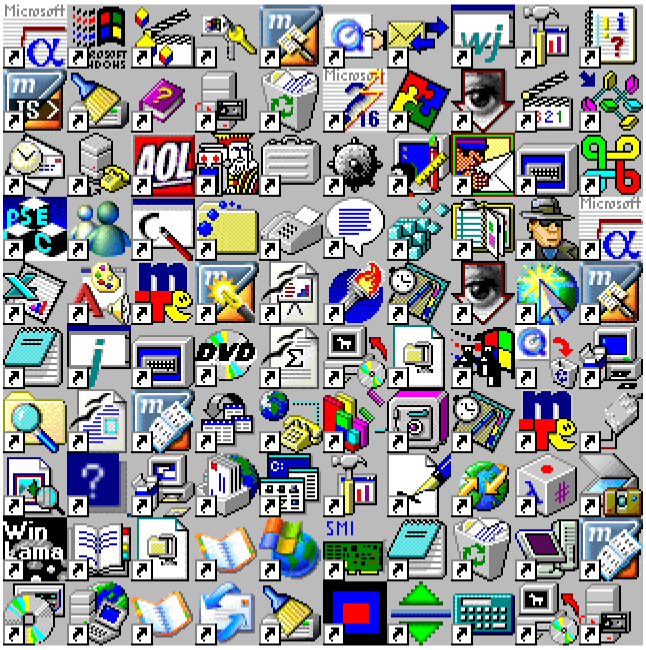

Figure 29. 10x10 matrix with inter-element distance of 0 icons.

Figure 30. 10x10 matrix with inter-element distance of ¼ icon.

Figure 31. 10x10 matrix with inter-element distance of ½ icon.