The influence of picture size on recognition and exploratory behaviour in raised-line drawings

Maarten W A Wijntjes, Thijs van Lienen, Ilse M Verstijnen¶, Astrid M L Kappers
Helmholtz Institute, Utrecht University, Physics of Man [¶ Psychonomics], Princetonplein 5,
NL 3584 CC Utrecht, The Netherlands; e-mail: m.w.a.wijntjes@phys.uu.nl
Received 20 November 2006, in revised form 30 July 2007; published online 26 March 2008

Abstract. We demonstrate the influence of picture size on haptic recognition and exploratory behaviour. The stimuli were raised-line drawings of everyday objects. Participants were instructed to think aloud during haptic exploration of the pictures. We measured the delay between initial correct speculation and final correct response. The results indicate that picture size influences accuracy but not response latency: large drawings are recognised more often but not faster. By analysing video recordings of the experiment we found that two-handed exploration increases when picture size increases and that, on average, 83% of the exploration time involves the use of two hands. The thinking-aloud data showed that the average time difference between the initial correct speculation and final correct response amounted to 23% of the total reaction time. We discuss our results with respect to the design of tactile aids and the ecological validity of single-finger exploration.

1 Introduction
Embossing the lines of a drawing makes the spatial information accessible to the haptic modality. Besides serving as a learning aid for the blind, this also inspired psychologists to investigate whether the recognition of line drawings can be extended to another modality besides vision. Although all studies on this subject show that it is possible to recognise a raised-line drawing without the use of vision, the task has proved to be notoriously difficult (eg Heller 1989; Loomis et al 1991). Haptic exploration of raised-line drawings often leads to an incorrect semantic interpretation and, when recognition is finally correct, the process may well have taken several minutes. A possible way to look at the recognition of raised-line drawings is to split up the process into two components which will probably be temporally overlapping. The first component is the process of exploration, which is concerned with aspects such as stimulus properties and exploratory behaviour. The second component is concerned with how the acquired spatial information is processed and interpreted. We briefly review the literature on the recognition of raised-line drawings with respect to the exploration and interpretation components.

Two important factors have been identified as having a critical influence on the acquisition of spatial information. First, Magee and Kennedy (1980) have shown that active exploration of raised-line drawings yields lower recognition accuracy than passively guided exploration. The recognition difference between actively and passively exploring raised-line drawings can be understood by considering active exploration as a dual task where the observer is in charge of both the exploration and the recognition process, whereas passively guided observers can focus fully on recognition. Other studies report similar findings on passive superiority for the recognition of raised-line drawings (D'Angiulli et al 1998; Symmons et al 2004). However, one should be cautious to generalise passive superiority beyond haptic perception of raised-line drawings (Symmons et al 2004). Second, irrespective of whether the finger is guided, the spatial information is acquired serially, which contrasts with the parallel nature of visual perception. Loomis et al (1991) showed that, when the visual field of view is matched to the effective field of the fingertip, the quality of the visual recognisability of

line drawings decreases and becomes comparable to that of haptic recognition. It can be tentatively argued that the difference between visual and haptic perception of line drawings is to be found not in specific object recognition processes of different modalities but rather in the difference between parallel and serial information acquisition, independent of modality. Further findings concerning the exploratory component of the recognition process indicate that using five fingers significantly increased recognition accuracy compared to single-finger exploration (Klatzky et al 1993). In contrast, Symmons and Richardson (2000) found that observers spontaneously explored raised-line drawings with a single finger. In attempting to increase the recognisability of raised-line drawings, Thompson et al (2003) found that 'filled' drawings (the whole region between the lines is embossed instead of only the lines themselves) are recognised more accurately than raised-line stimuli.

An interesting question concerning the interpretation of a raised-line drawing is whether congenitally blind observers, having no experience with interpreting the projection of the 3-D environment on a 2-D surface, can recognise raised-line drawings. Heller (1989) found that late-blind observers were superior to both sighted and congenitally blind and that sighted and congenitally blind were equally poor at tactile picture recognition, whereas Lederman et al (1990) found that sighted observers were superior to the congenitally blind. The difference between sighted and congenitally blind was later replicated by Heller et al (1996) who found no differences between late-blind and sighted observers but did find that both of these groups outperformed the group of congenitally blind. These studies together suggest that visual experience is an important factor and that late-blind observers could benefit from their exploratory experience; this is in line with the increase in recognisability that occurs when observers are guided (Magee and Kennedy 1980). Further studies investigating the interpretation component of haptic line drawings showed that recognition increases when categorical information about the drawing is provided (Heller et al 1996). As suggested by Klatzky and Lederman (1987), the recognition process seems to be an inferential, hypothesis-testing procedure. This suggestion was later strengthened by Kennedy and Bai (2002) who observed that during exploration guesses are made about the identity of the referent. These guesses are labelled with a 'fit judgment' and are either rejected or accepted according to how well they correspond to ('fit') the perceived stimulus. Kennedy and Bai (2002) showed that the fit judgment that observers gave to their last guess, ie final answer, predicted the accuracy.

The study we present here is concerned with both the exploration and the interpretation components of recognition of raised-line drawings. Apart from the study by Thompson et al (2003), surprisingly little attention has been given to the role of stimulus properties. The influence of properties such as picture size, material, textures, line thickness, or dashing patterns has not been systematically studied up to now. It is peculiar that in previous research relatively small stimulus sizes have been used compared to the size of a typical medium on which the stimuli are printed, namely on an A4 sheet (21 cm × 29.7 cm). Drawings used in all the above-cited papers range between 4 and 15.2 cm, with one notable exception. Kennedy and Bai (2002) used stimuli ranging between 15.5 and 22.5 cm; the size was explicitly motivated by the authors' wish "to increase the likelihood that subjects would detect small features of the displays and their proportions" (page 1015). Indeed, the overall accuracy rate in their studies (61%) was higher than in earlier studies with comparable picture sets. They suggested that the difference in accuracy was caused by the use of large stimuli. The first objective in our study is to test the hypothesis that larger stimuli are more accurately recognised than small stimuli. We also present a concise review of the literature in which we relate stimulus material and picture size to recognisability.

The second objective of this study is to clarify current findings concerning spontaneous exploratory behaviour. As mentioned, Symmons and Richardson (2000) found that blindfolded sighted observers spontaneously use a single finger for the exploration of a raised-line drawing during the larger part of the exploration time. This finding may lack some generality. First, the pictures were rather small (8.5 cm) and, if according to our first hypothesis larger pictures are more easily recognised, the exploratory behaviour would have to be reinvestigated for large stimuli. Second, observers had to stand upright, a factor that might induce them to use one hand for balancing purposes and hence lead to misleading results. Besides these possibly experimental imperfections, it has been noted (eg Heller et al 2002) that blind observers do not use merely one finger and they object when the experimenter suggests they should. This might indicate that blindfolded observers use a different exploratory strategy than blind observers. It could also indicate that the findings of Symmons and Richardson (2000) are not as general as the authors believed them to be. Last, research in tactile map perception has suggested that spontaneous behaviour is often two handed (Berla et al 1976, pages 270, 272), although this is task-dependent. Together with our own observations during pilot experiments, the above-mentioned indications motivated us to reinvestigate the spontaneous exploratory behaviour with regard to raised-line drawings.

The third objective of our study focuses on the interpretation component of the recognition of raised-line drawings. As described by Klatzky and Lederman (1987) and Kennedy and Bai (2002), observers are continuously guessing about the correct referent during the exploration process. We also observed this during pilot experiments and became intrigued by the time it took for the observers to accept their last hypothesis. We presented participants with a think-aloud protocol with which we could quantify this delay between initial guess and final correct response. Besides being interested in how large a portion of the total exploration time is devoted to confirm the initial correct guess, we investigated the relationship between the delay and the difficulty of the stimulus. D'Angiulli et al (1998) found that recognition scores per picture correlated significantly with different groups of participants. This indicates that pictures contain some inherent difficulty which is invariant with respect to different observers. We first verified whether our data also revealed such an inherent difficulty by correlating recognition scores for small and large stimuli. A reconfirmation of this inherent difficulty evoked by different pictures was used to investigate whether response delay would depend on the difficulty of the task.

2 Method
2.1 Participants
A total of twenty-eight observers (seventeen males) participated. Participation was voluntary. Only twenty-one were videoed because the video facility was not available for the last seven. Participants were naive with respect to stimuli and the purpose of the study. Since some of the participants were colleagues working at the same lab as the authors, there was a possibility that they might have accidentally encountered a stimulus previously. They were asked to notify the experimenter if this occurred, whereupon that trial was excluded from the analysis.

2.2 Stimuli and materials
The set of 12 raised-line drawings used in this experiment can be seen in figure 1. The hammer, umbrella, and scissors were adopted from the picture set used by Kennedy and Bai (2002). Large and small versions of the drawings were produced on A3 and A4 Zytech swell paper (29.7 cm × 42 cm and 21 cm × 29.7 cm, respectively). The size of a stimulus is defined according to its horizontal or vertical size, whichever is the larger. The sizes of the large and small stimuli were 35 cm and 10 cm, respectively.
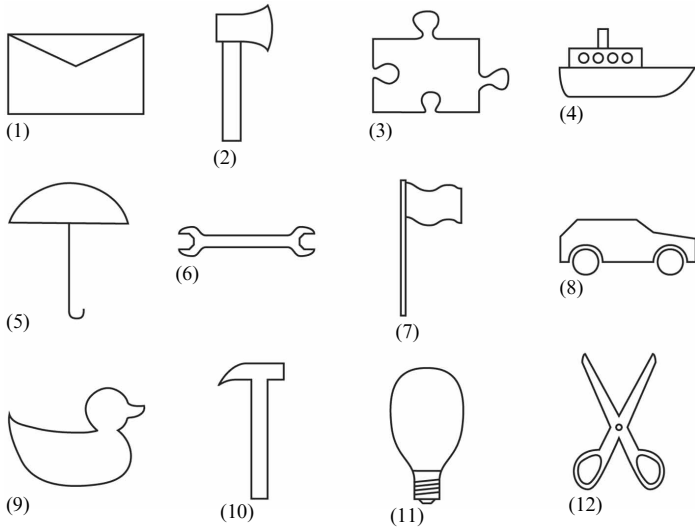
Figure 1. Stimuli used in the experiment: (1) envelope; (2) axe; (3) puzzle piece; (4) boat; (5) umbrella; (6) open-end wrench; (7) flag; (8) car; (9) bath-tub duck; (10) hammer; (11) light bulb; and (12) scissors.

So as not to have representations of objects of unbalanced prototypical sizes of the depicted objects which might induce a size effect, we selected drawings such that half of the stimulus set consisted of depicted objects which have typical sizes that are nearer to 10 cm than 35 cm (envelope, puzzle, open-end wrench, bath-tub duck, light bulb, and scissors). Scaling the length of the umbrella and puzzle piece drawings to 35 cm resulted in a width which exceeded the width of an A3 sheet. Therefore, the size of the umbrella was reduced to 26 cm and the size of the puzzle piece was reduced to 30 cm. The width of the lines was 1 mm and the approximate height of the lines for both small and large stimuli was 0.5 mm.

For each of the two sizes a stainless steel mould was made in which the stimuli sheets could be fixed. The appropriate mould was placed on the table, in front of the observer, and was fixed with respect to lateral movement. A microphone was positioned on the table, and above the mould a video camera was suspended. The setup can be seen in figure 2. The video recordings were performed on a VCR video recorder to which both the microphone and camera were connected.

2.3 Procedure

Participants received written instructions. They were informed that their primary task was to recognise the raised-line drawing both as accurately and as fast as possible while blindfolded. They were also asked to express their thoughts and guesses verbally during the recognition process. The participants were supposed to explicitly announce their final answer and both the time of their first correct guess and the time of their final answer were noted by the experimenter. Participants were given unlimited exploration time. A trial ended either with a correct or incorrect final answer or could be broken off by the participant if he or she wished. Participants were not told whether their response was right or wrong. The written instructions informed the observers that they were free in the way they explored the stimuli. After that a raised-line drawing of a fork was presented as a practice trial in order to familiarise participants with the recognition of raised-line drawings.

During the experiment, participants were presented with small drawings alternating with large drawings, but were never presented with both the small and large versions of the same picture. The caption of figure 1 describes the order of the drawings (1 to 12)
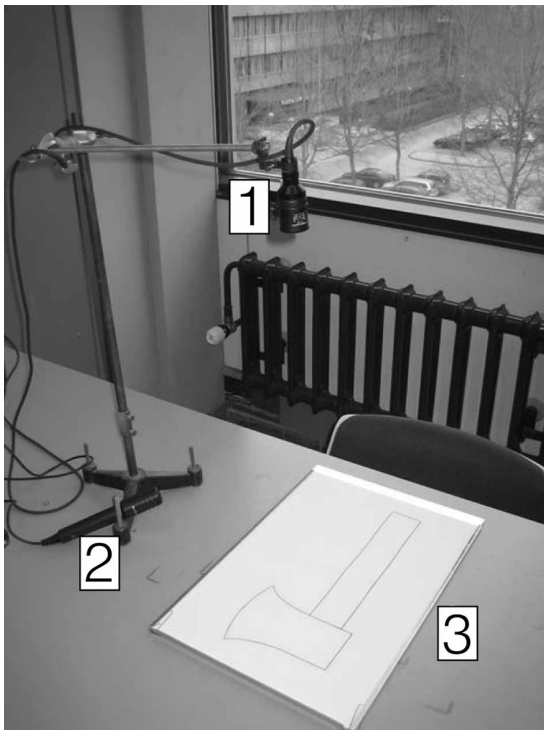
Figure 2. Picture of the setup. (1) Camera; (2) microphone; (3) stainless steel mould with an A3 stimulus sheet.

as used in the experiment. By presenting the drawings in reverse order and by starting alternately with a large and a small stimulus we were able to create four different stimulus sets which were equally distributed among the twenty-eight participants.

2.4 Recognition analysis

To statistically analyse the response latency and accuracy we used what one could call a within-pictures design. The rationale behind this design is as follows. In general, when performance of the same observers is measured under different conditions, one uses a within-subjects (repeated-measures) design. The rationale behind this design is that it cancels out intersubject variability. Intersubject variability is caused by individual differences, such as IQ, which are supposed to be unrelated to the experimental manipulations. One could apply the same reasoning to pictures: each picture presents an individual difficulty which is independent of the experimental manipulation. However, before applying this reasoning, one first needs to prove that pictures present an individual difficulty. D'Angiulli et al (1998) found that pictures correlated with different conditions, which proves that pictures present individual difficulties. To ascertain that also in our experiment the pictures fulfilled the requirement for a repeated-measures design, we tested whether the average recognition scores of the small pictures correlated significantly with the average recognition scores of the large pictures. It should be noted that the observers never experienced a picture twice. The data which were used for analysis were the average recognition latency and accuracy per picture per condition. These data points thus reflected the average scores of a group of fourteen participants.

Since our accuracy scores were sometimes near 100% correct, we used a non-parametric test, because data near 100% cannot be normally distributed, whereas a binomial distribution around 50% is nearly similar to a normal distribution. Also the latencies were analysed non-parametrically since reaction times cannot be normally distributed. When the picture correlations were significant, we used a paired Wilcoxon signed rank test, and when there were no correlations we used a Mann – Whitney test.

We also reported the correlation values. We predicted an increase in accuracy with increasing picture size (Kennedy and Bai 2002), and therefore we used a one-tailed p value. We made no prediction about the response latency.

2.5 *Video analysis*
The manual explorations by the first twenty-one participants were videotaped (as noted previously, the video facility was not available for the last seven participants). Raters were instructed to categorise the movement of the participants' hands in one of three strategies: (i) use of a single hand: the other hand does not touch the drawing; (ii) two-handed use: one hand moves while the other hand rests on the drawing; (iii) simultaneous use of both hands. In the discussion we will comment more elaborately on these choices. Figure 3 illustrates these strategies. We chose these three categories because video recordings gave the impression that they were unequivocal and would thus be objectively rateable. We categorised by hands and not by individual fingers because the movement of individual fingers was too complex to rate. We used strategy (i) to compare our data with the single-finger exploration data of Symmons and Richardson (2000).
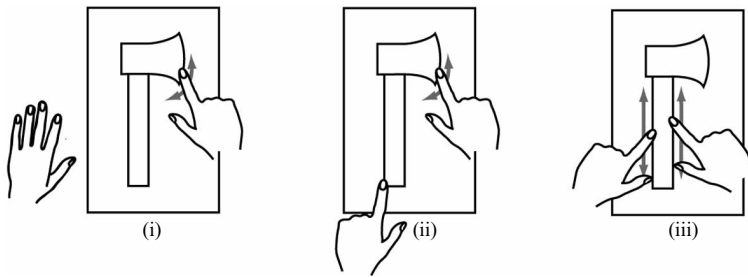
Figure 3. Sketches of the three exploration strategies. (i) Single-hand use, the other hand does not touch the drawing. (ii) Two-handed use where a single hand moves while the other hand rests on the drawing. (iii) Two hands are moving simultaneously.

While watching the video recordings the raters pressed a keyboard key specifically to denote a particular kind of movement strategy. A program made in LabView recorded the times at which a key was pressed and released. For a given trial (raised-line drawing) the total exploration time for each strategy was calculated. All video data were analysed by a naive rater who was reimbursed for her participation. The first author rated the video data of the first six participants in order to calculate the interrater reliability.

We calculated rater reliability using intraclass correlations as described by Shrout and Fleiss (1979). For the reliability analysis we used the raw data which are the absolute times of each strategy per trial. To analyse the hand use, we used proportions of each strategy per trial to prevent the data from being weighted by trial time. The data of slow observers would otherwise weight more than those of fast observers.

3 Results
In total, during four trials some participants commented that they recognised the picture because they had encountered it previously in the lab. These trials were omitted from the analyses. We later verified that all possible outcomes of these trials did not have any effect on the final results. Also dismissing these subjects from the analyses did not influence the outcome. Two stimuli incidentally caused confusion: the hammer was sometimes mistaken for a pick-axe which was counted as a correct response; the exact name of the open-end wrench was occasionally not known whereupon the

trial was counted correct if observers were able to correctly describe the execution associated with an open-end wrench.[1]

3.1 *Response latency and accuracy*
The response latency is defined as the time the participant needed from the start of exploration until the final correct response. Response latencies ranged from 5 to 337 s (average scores are presented in table 1). Mean response latencies (standard errors) taken over all trials were 48.8 (4.1) and 47.5 (3.9) s for large and small pictures, respectively. The median response latencies for large and small pictures were 33 and 37 s. The overall mean proportion of correct responses (standard error) for large and small pictures was 0.84 (0.03) and 0.77 (0.03), respectively.

Table 1. Recognition data per picture. The accuracy shows the proportion of correct responses.

Picture	Response latency/s		Accuracy	
	small	large	small	large
Envelope	68	53	0.79	0.86
Axe	49	64	0.71	0.71
Puzzle piece	62	92	0.71	0.79
Boat	65	43	0.93	1.00
Umbrella	18	24	1.00	1.00
Open-end wrench	52	36	0.79	0.64
Flag	35	47	0.79	1.00
Car	52	32	0.93	0.93
Bath-tub duck	32	90	0.43	0.64
Hammer	52	28	0.85	0.93
Light bulb	26	57	0.69	0.79
Scissors	52	43	0.62	0.79
Mean[a]	47	51	0.77	0.84
Median[a]	*52*	*44*	*0.79*	*0.82*
Minimum[a]	*18*	*24*	*0.43*	*0.64*
Maximum[a]	*68*	*92*	*1.00*	*1.00*

[a] Since these statistics are calculated using the means per picture they can differ from the overall scores reported in the text.

　Whereas the response latency did not show significant correlation between pictures of different sizes ($r = -0.02$, $p = 0.96$), the accuracy showed highly significant correlation ($r = 0.76$, $p = 0.004$). Response latency did not show a significant effect as shown by the Mann–Whitney test (n_1, $n_2 = 12$, $U = 71$, $p = 0.954$, two-tailed) whereas the accuracy did show a significant effect as shown by the Wilcoxon test ($W_+ = 39$, $W_- = 6$, $N = 9$, $p = 0.028$, one-tailed). Furthermore, we did not find a significant correlation between accuracy (averaged per picture) and reaction time ($r = -0.23$, $p = 0.28$).

[1] In the Netherlands, the name for an open-end wrench is not always known whereas the purpose is known. Often people incorrectly use the word screwdriver for this object. We thereupon ascertained that they really meant an open-end wrench and not a screwdriver by asking them to describe its purpose. It is debatable whether one can speak of recognition if the object cannot be named correctly. Since the experimenters got the impression that observers very vividly knew and described the object, we decided to count this as recognition. It should be noted that the open-ended wrench stimulus was the only stimulus which showed higher recognition for the small-sized version. This means that, if we would have excluded this picture from the data, the *p*-value would have been lower and thus would have supported our findings even stronger.

3.2 Response accuracy compared with other studies

Since the response accuracy we found was rather high compared to previous studies, a reviewer suggested a control experiment in which we tested five new naive subjects on the large stimuli. The methods for this experiment were similar to those in the main experiment except that we used only large stimuli and the observers were not asked to think aloud. We found that all five observers recognised 11 of the 12 stimuli, ie a response accuracy of 91.7%. Four of the five unrecognised pictures were unique. To compare our data with literature and to facilitate the discussion of the differences we present an overview in table 2.

Table 2. Review of recognition data in a selection of literature.

Reference	Pictures	Material	Size/cm	Accuracy/%
Magee and Kennedy (1980)				12.5
Heller (1989)	own set	Swedish kit	4–11	12.9
Lederman et al (1990)	Snodgrass	Braille printer	12.7–15.2	33.5
Loomis et al (1991)	Snodgrass	swell paper	15.9	43.8[a,b]
Klatzky et al (1993)	own set	swell paper		23[a,b]
Heller et al (1996)	Snodgrass	Swedish kit	7.5–10	24.7
Kennedy and Bai (2002)	own set[c]	Swedish kit	15.5–22.5	61
Thompson et al (2003)	Snodgrass	swell paper	13–15	56[a]
This research (small)	own set	swell paper	10	77
This research (large)	own set	swell paper	35	84

[a] The accuracy data have been (partly) extracted from a graph.
[b] Accuracy has been averaged over the single and multiple finger data.
[c] Partly based on Heller (1989) and D'Angiulli et al (1998).

From all studies we took the accuracy scores from adult sighted observers. When several experiments were reported, we used the data from the first experiment. In order to compare the different studies we examined what kind of pictures were used, what kind of material was used, and how large the pictures were. The different picture sets which were used were either subsets of Snodgrass and Vanderwart (1980) or pictures which were inspired by the study of Heller (1989). The stimulus material used in most cases was either the Swedish raised-line drawing kit or swell paper which was used in the current study. The drawing kit uses plastic sheets of paper which are very sensitive to pressure strokes with a pen and produces thin but clear tangible lines. We have also plotted in figure 4 the accuracy scores of these studies against the average size of the stimuli (when available).

3.3 Video analysis

On the basis of the video data for the first six participants we calculated the intraclass correlation coefficient (Case 2 model—Shrout and Fleiss 1979) which showed that the rating procedure was highly reliable [$ICC(2, 1) = 0.966$, $F_{215, 216} = 57.468$, $p < 0.001$]. For the analysis we used the video rating data of the naive rater for the first twenty-one participants.

The time participants spent using each strategy was normalised per experimental trial to ensure that each trial would have equal weight. This resulted in $21 \times 12 = 252$ sets of proportions for the three conditions. The average proportions of movement strategies per picture size can be seen in figure 5. On average, for 17% of the exploration time observers used the single-hand strategy, 21% used two-handed exploration with one hand static, and in 62% of the time the two hands were both moving. It can be seen that the single-hand strategy is used more often for small pictures, but the opposite is the case for both two-handed strategies. A 2×3 repeated-measures ANOVA was performed on the average data per participant with size and strategy as factors.
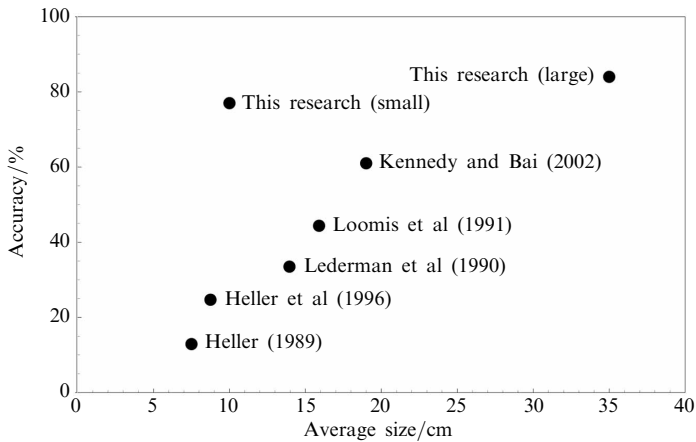
Figure 4. The average recognition accuracy plotted against the average stimulus size from previous studies together with the current results.
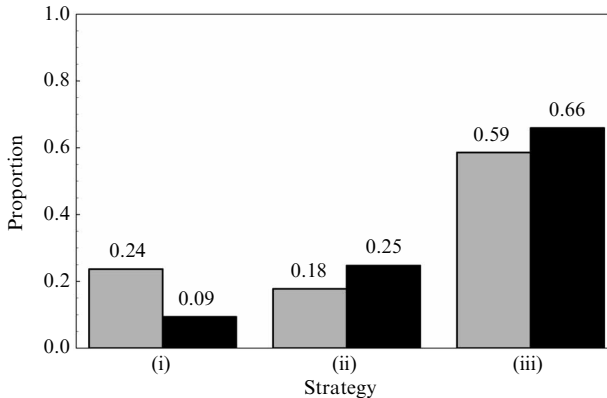
Figure 5. The overall proportions of movement strategies per size condition: grey bars denote small pictures; black bars denote large pictures. Strategy (i) single-hand movement, (ii) use of two hands one of which is static, and (iii) use of both hands dynamically. Because of rounding, the proportions do not add up to exactly one.

Owing to the normalisation of the exploration strategy, the factor 'size' yields equal means and will therefore not be reported. The degrees of freedom for the strategies were adjusted with the Greenhouse–Geisser correction because the data failed to show sphericity. Both the main effect of strategy ($F_{1.3, 26.1} = 25.057$, $p < 0.0005$, $\varepsilon = 0.652$) and the interaction between strategy and picture size ($F_{2, 40} = 7.483$, $p = 0.002$) were significant. The latter effect reflects the change in strategy when the size of the picture changed as shown in figure 5. Our results further suggest a possible relationship between two-handed exploration and accuracy since they both increase with increasing picture size. To explore this possibility we calculated the correlation between accuracy and two-handed exploration proportions per picture. We did not find any correlation between these two measures ($r = 0.06$, $p = 0.86$).

3.4 Delay
For 30% of the correct responses there was no record of a preceding first answer; these data points were included in the analysis and scored as a delay of 0 s. The average time difference between the first and final correct response (standard deviation) was 11.4 (22.5) s which equals 23% of the average final response times. Correlating the first response time with the delay yielded $r = 0.07$ ($p = 0.23$), indicating that the

delay does not depend on the initial recognition time. To investigate the relationship between delay and difficulty we calculated the correlation between all nonzero delay trials and the average accuracy obtained for the individual stimuli. We did not find a significant correlation ($r = -0.04$, $p = 0.60$).

4 Discussion

We have shown that picture size influences recognisability. Large pictures elicit higher response accuracy, but response latency is not influenced. According to the hypothesis of Kennedy and Bai (2002), larger pictures should be recognised more easily because they contain more accurate spatial information. To investigate this, it should first be noted that a hypothetical gain in spatial resolution can, in principle, be both cutaneous and kinaesthetic, and should thus be analysed separately.

The minimal distances between the lines of the small pictures of our stimulus set were never less than 3 mm (the spiral lines on the bottom of the light bulb) which is above the typical tactile spatial resolution of 1 mm (Van Boven and Johnson 1994). Although the lines can therefore be distinguished during a controlled discrimination experiment, this might be different when high line density regions form only a small part of a large and spatially complex stimulus. When visually inspecting our stimulus set, we realised that only a few stimuli may contain these high line density regions. Although visual inspection is speculative, this indicates that the cutaneous resolution is not the critical factor responsible for the size effect.

Assessing how the resolution of kinaesthetic input is influenced by size is not straightforward. The spatial information contained in the raised lines can be divided into positions and path elements. For the haptic encoding of position we can assume that the signal contains constant uncertainty over the whole workspace, which would mean that the relative kinaesthetic accuracy, defined by the ratio between discrimination threshold and, for example, length, increases with scale. This assumption about the kinaesthetic encoding of position can be compared with measuring two lengths with a ruler: the measurement error of the ruler is constant and thus the relative error will decrease with increasing length. It is, however, unknown how the kinaesthetic accuracy of the path elements is related to size. A law which often applies to discrimination thresholds of quantities such as weight, brightness, and sound is Weber's law which states that thresholds increase linearly with intensity. If Weber's law holds for line lengths and curvatures this implies that the relative accuracy is invariant with respect to scale. Since research relating spatial resolution to scale is very sparse and does not exist at all for raised-line materials, it is impossible to infer whether the kinaesthetic encoding of line paths gains accuracy when larger stimuli are used. There are, however, no indications that accuracy is negatively influenced by size.

We believe that the gain in spatial resolution is likely to be found in the kinaesthetic encoding, at least in the position information. Our conclusion is supported by Magee and Kennedy (1980) who found that the spatial information is largely kinaesthetically encoded and that the cutaneous input merely serves to guide the fingers along the lines.

Another candidate that could be responsible for the size effect is the differential hand use for the two different picture sizes. Although both two-handed exploration and recognition accuracies increase with scale, no correlation was found between these measures. Correlating the accuracy with exploration strategies, however, might not be the appropriate technique for confirming that the use of two hands has a positive influence on accuracy. Since multiple strategies are used within a trial we do not know which exploration strategy is the most informative for the observer. A study that systematically controls hand use could resolve this matter.

A striking aspect of the results is that our overall recognition scores are so high when compared to other studies on the recognition of raised-line drawings. As can be seen from table 2, this does not seem to be due to different materials used to produce the raised-line drawings. Another possible difference between our study and previous studies is the use of different observers. Since all studies, including ours, were performed on university campuses and the majority of the observers were students, we do not think that the difference is due to this factor. However, our picture set might be a realistic candidate responsible for our high recognition scores. Only three of our stimuli have been used in previous studies which makes our set to a great extent new. We can thus only hypothesise that the use of different pictures is a (partial) explanation for our high recognition scores. This is surprising since, as can be seen in table 2, various studies have used different picture sets, and this has never before seemed to be an important factor. Although this issue transcends the scope of the current research, it is an interesting finding and it could be worthwhile to investigate whether our set is indeed better recognisable than that used by Kennedy and Bai (2002). Furthermore, our high recognition scores support the theory that assumes that touch is capable of accurate picture perception.

When we collected the accuracy data from literature and plotted these against the average stimulus sizes, a clear relation emerged. As reported by Kennedy and Bai (2002), and confirmed by the current research, the size effect was already present in the literature. The fact that the accuracy we found for small pictures does not fit this trend can be explained by the overall high accuracy we found. Although the interpretation of the graph may be tentative, it certainly strengthens our finding that size affects recognisability.

The second objective of our study was to reinvestigate recent findings concerning the spontaneous behaviour observed in the recognition of raised-line drawings. As noted in the introduction we had sufficient reason to doubt whether the findings of Symmons and Richardson (2000) were reproducible. Indeed, we found different spontaneous behaviour. As can be seen in figure 5, a single hand is used in only 9% and 24% of the total exploration time for large and small pictures, respectively. This conflicts with the notion of single-finger dominance. Even if Symmons and Richardson (2000) meant *dynamic* single-finger movement, which would imply that our strategy (ii) ranks as single finger use. If we use only our small stimuli to compare our results with those of Symmons and Richardson (2000) (who used 8.5 cm stimuli) then we still observe the dominance of dynamic two-handed use (59%, see figure 5). One possible reason why we obtained different results is that in our study observers were seated, whereas in the study of Symmons and Richardson (2000) they were standing at a table. In conclusion we can here say that we do not share the idea of ecological validity of single-finger exploration. Not only did we find two-handed exploration to be dominant for small pictures, but we also observed that single-hand exploration decreases to 9% of the total exploration time for large pictures, which were also recognised more accurately. The fact that spontaneous behaviour appears to be two-handed does not degrade single-finger studies. Our results merely indicate that the studies of the recognition of raised-line drawings should not restrict exploration when the experimenters desire an ecologically valid method.

Besides reinvestigating the earlier results of Symmons and Richardson (2000), a rationale behind our study of the exploratory behaviour was inspired by the exploratory procedures (EPs) proposed by Lederman and Klatzky (1987). That study showed that different EPs are used for distinct object properties when exploring 3-D objects. From the total of eight EPs, three might theoretically apply to the exploration of raised-line drawings: lateral motion, static contact, and contour following. The other five EPs are inherently connected with the three-dimensionality of real objects.

When we watched the videos of the experiments, it occurred to us that almost all exploration would be categorised as contour-following. The only clear difference was that subjects used either one or two hands and the two-handed exploration can be split up into our strategies (ii) and (iii). Aside from the advantage that these three strategies are unequivocally rateable, we also had the impression that these strategies serve different purposes. Using two hands simultaneously (strategy iii) facilitates symmetry detection, which may be cognitively efficient. Using one hand as an anchor point while exploring with the other hand (strategy ii) constitutes a different reference frame than using only a single hand. Using a reference frame with a fixed origin (strategy ii) probably encodes the spatial information more accurately than without origin (strategy i). As one reviewer suggested, using different reference frames for the small and large pictures may also explain why larger pictures are better recognised. The use of two hands also decreases illusory percepts as was shown by Heller et al (2005): when two index fingers were used, the Müller-Lyer illusion became weaker in comparison to single-finger exploration.

The third objective of our study was to contribute to an understanding of the mental process during the recognition of line drawings. Our finding that, on average, 23% of the total response latency is used to confirm the initial correct idea suggests that hypothesis-testing forms a substantial part of the recognition process. It was already mentioned by Klatzky and Lederman (1987) and Kennedy and Bai (2002) that the recognition process would resemble a hypothesis-testing procedure. Kennedy and Bai (2002) observed that the 'fit judgment' which observers made was related to accuracy. We did not find that the delay predicted the difficulty of the stimulus. It is likely, therefore, that relations exist which we cannot detect because we have delay data only for correctly recognised stimuli. Therefore, we cannot investigate the delay on a trial-to-trial basis.

As for the applications in the area of tactile aids, our results point in one clear direction. Larger pictures are recognised more often. It is certainly advisable to increase the size of the picture as much as possible. Combining our results with those of Thompson et al (2003) would yield a design recipe where the size is maximised and the region between the lines is also embossed, although further research is needed to investigate the interaction between these two factors. Although we and Thompson et al (2003) used sighted observers, there is no reason why the effect of size would not work for blind observers. It has been theorised that similar interpretation principles underlie the recognition of raised-line drawings (D'Angiulli et al 1998). Furthermore, we did not find a direct relation between hand use and accuracy but, taking into account that blind observers disapprove of single-hand use (Heller et al 2002), we suggest that observers should not be encouraged to use just a single finger.

References
Berla E, Butterfield L, Murr M, 1976 "Tactual reading of political maps by blind students: a videomatic behavioral analysis" *Journal of Special Education* **10** 265 – 276
D'Angiulli A, Kennedy J M, Heller M A, 1998 "Blind children recognizing tactile pictures respond like sighted children given guidance in exploration" *Scandinavian Journal of Psychology* **39** 187 – 190
Heller M A, 1989 "Picture and pattern perception in the sighted and the blind: the advantage of the late blind" *Perception* **18** 379 – 389
Heller M A, Brackett D D, Wilson K, Yoneyama K, Boyer A, Steffen H, 2002 "The haptic Müller-Lyer illusion in sighted and blind people" *Perception* **31** 1263 – 1274
Heller M A, Calcaterra J A, Burson L L, Tyler L A, 1996 "Tactual picture identification by blind and sighted people: Effects of providing categorical information" *Perception & Psychophysics* **58** 310 – 323

Heller M A, McCarthy M, Schultz J, Greene J, Shanley M, Clark A, Skoczylas S, Prociuk J, 2005 "The influence of exploration mode, orientation, and configuration on the haptic Müller-Lyer illusion" *Perception* **34** 1475 – 1500

Kennedy J M, Bai J, 2002 "Haptic pictures: Fit judgments predict identification, recognition memory, and confidence" *Perception* **31** 1013 – 1026

Klatzky R L, Lederman S J, 1987 "The Intelligent Hand", chapter in *The Psychology of Learning and Motivation* Ed. G Bower (New York: Academic Press) pp 121 – 151

Klatzky R L, Loomis J M, Lederman S J, Wake H, Fujita N, 1993 "Haptic identification of objects and their depictions" *Perception & Psychophysics* **54** 170 – 178

Lederman S J, Klatzky R L, 1987 "Hand movements: A window into haptic object recognition" *Cognitive Psychology* **19** 342 – 368

Lederman S J, Klatzky R L, Chataway C, Summers C D, 1990 "Visual mediation and the haptic recognition of two-dimensional pictures of common objects" *Perception & Psychophysics* **47** 54 – 64

Loomis J M, Klatzky R L, Lederman S J, 1991 "Similarity of tactual and visual picture recognition with limited field of view" *Perception* **20** 167 – 177

Magee L, Kennedy J, 1980 "Exploring pictures tactually" *Nature* **283** 287 – 288

Shrout P E, Fleiss J L, 1979 "Interclass correlation: Uses in assessing rater reliability" *Psychological Bulletin* **86** 420 – 428

Snodgrass J, Vanderwart M, 1980 "A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity" *Journal of Experimental Psychology: Human Learning and Memory* **6** 174 – 215

Symmons M, Richardson B, 2000 "Raised line drawings are spontaneously explored with a single finger" *Perception* **29** 621 – 626

Symmons M, Richardson B, Wuillemin D, 2004, in *Active Versus Passive Touch: Superiority Depends More on the Task than the Mode* Eds S Ballesteros, M A Heller (Madrid: Universidad Nacional de Educacion a Distancia)

Thompson L J, Chronicle E P, Collins A F, 2003 "The role of pictorial convention in haptic picture perception" *Perception* **32** 887 – 893

Van Boven R W, Johnson K O, 1994 "The limit of tactile spatial resolution in humans: Grating orientation discrimination at the lip, tongue, and finger" *Neurology* **44** 2361 – 2366