

Haptics disambiguates vision in the perception of pictorial relief

M.W.A. Wijntjes^a R. Volcic^b S.C. Pont^a J.J. Koenderink^c A.M.L. Kappers^d

^a Industrial Design Engineering, Delft University of Technology, Delft, the Netherlands;

^b Psychologisches Institut II, Westfälische Wilhelms-Universität, Münster, Germany;

^c Faculty of EEMCS, Delft University of Technology, Delft, the Netherlands;

^d Physics of Man, Helmholtz Institute, Utrecht University, Utrecht, the Netherlands

ABSTRACT

In this study we demonstrate that touch decreases the ambiguity in a visual image. It has been previously found that visual perception of three-dimensional shape is subject to certain variations. These variations can be described by the affine transformation. While the visual system thus seems unable to capture the Euclidean structure of a shape, touch could potentially be a useful source to disambiguate the image. Participants performed a so-called ‘attitude task’ from which the structure of the perceived three-dimensional shape was calculated. One group performed the task with only vision and a second group could touch the stimulus while viewing it. We found that the consistency within the haptics+vision group was higher than in the vision-only group. Thus, haptics decreases the visual ambiguity. Furthermore, we found that the touched shape was consistently perceived as having more relief than the untouched the shape. It was also found that the direction of affine shear differences within the two groups was more consistent when touch was used. We thus show that haptics has a significant influence on the perception of pictorial relief.

Keywords: Visual perception, 3D shape, pictorial relief, haptics

1. INTRODUCTION

1.1 Pictorial relief

Although pictures of natural scenes are flat, we perceive the content as having a three-dimensional structures related to the original scenes. Observers might think that they perceive the same structures when viewing the same pictures. To investigate whether this is true, one needs a psychophysical method that measures the perceived structure in pictures, the pictorial relief. One of these methods, the attitude task,¹ is illustrated in Fig. 1. The attitude probe consists of a three-dimensional disk with in the middle a small rod normal to the disk. Observers adjust the orientation of the attitude probe such that the disk seems to lie on the surface of the depicted shape, as illustrated in Fig. 1A. Effectively, the local orientation of the *perceived* surface is measured. This measurement is repeated at all points of a regular grid (triangulation) over the surface, as illustrated in Fig. 1B. When the perceived local orientations are sufficiently sampled, the data can be integrated to obtain the pictorial relief, as illustrated in Fig. 1C.

This procedure thus converts a picture $I(x, y)$ into a pictorial relief $z(x, y)$ that contains the subjective depth dimension added by the observer. For an observer, the picture is a projection of an unknown scene or shape with unknown reflectance properties under unknown illumination conditions. All these ‘unknowns’ are confounded in the final image and cannot be disentangled. This makes it impossible to infer the fiducial depth without knowledge of the exact reflectance and illumination conditions. It is thus not strange that pictorial reliefs of different observers differ substantially.²⁻⁴ It is therefore a misunderstanding that observers see the same scene when viewing the same pictures. However, differences show a remarkable systematicity. In principle, any transformation of $z(x, y)$ that leaves the image $I(x, y)$ intact may describe these differences. Thus the reliefs of two observers $z_1(x, y)$ and $z_2(x, y)$ may be related by a function $f(x, y, z)$ as $z_1(x, y) = f(x, y, z_2(x, y))$. As was

Corresponding author:

M.W.A. Wijntjes: E-mail: m.w.a.wijntjes@tudelft.nl

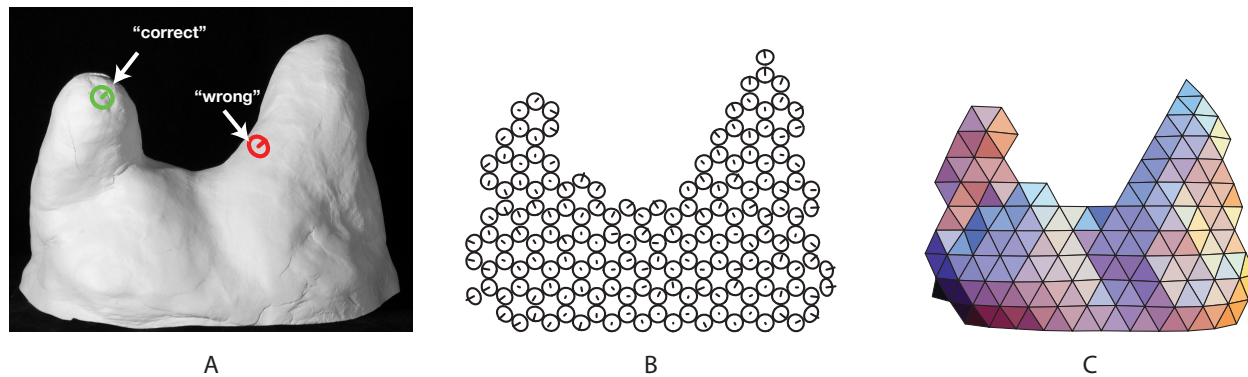


Figure 1. Illustration of the attitude task. See main text for explanation.

shown by Ref. 3, this mapping has to be linear if one assumes that planar objects are planar for each observer. Two pictorial reliefs (x, y, z) and (x', y', z') can thus be related by an affine transform:

$$x' = x \tag{1}$$

$$y' = y \tag{2}$$

$$z' = ax + by + cz + d \tag{3}$$

Here, a and b define an ‘additive plane’, c an ‘affine stretch’ in the viewing direction and d an arbitrary depth offset. In Fig. 2, the affine transform is illustrated with renderings of a 3D model. On the left, the original model (a bunny⁵) can be viewed from the top. In the middle, the object is stretched, on the right the object is sheared.

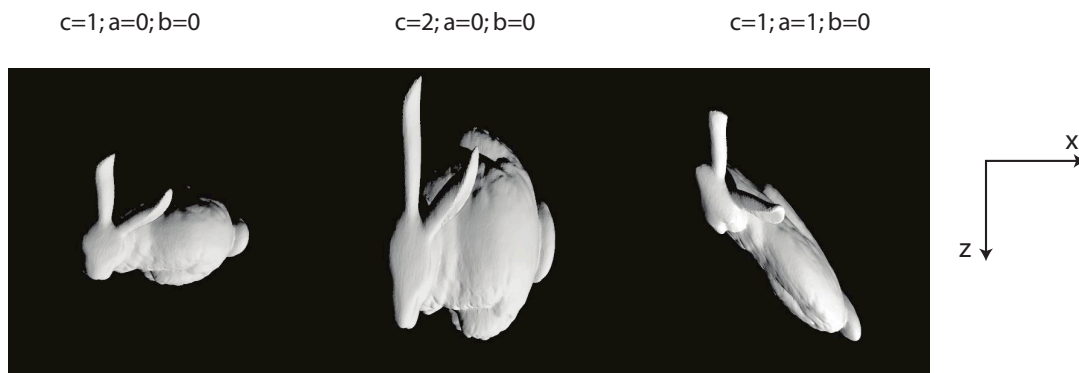


Figure 2. Illustration of affine transformation of a 3D shape.

Evidently, the affine transformation also describes differences between a pictorial relief and a fiducial relief, although there is only fragmentary evidence for this. It was shown³ that the affine transformation accounts for almost all differences between subjects. It can thus be concluded that humans fail to perceive the correct *Euclidean* structure of objects in images, whereas they correctly perceive the *affine* structure.

1.2 Vision and haptics

Although visual perception seems non-Euclidean, there is no a priori reason to expect the same for haptic perception. While the visual sense is to be ‘distal’ (it receives light rays that project the visual world onto the retina), haptics is to be ‘proximal’ (there is direct mechanical contact between the sensors and the stimulus). It is thus interesting to investigate the influence of haptics on visual perception of three-dimensional shape.

Until now, studies on cross-modal visual-haptic perception have either focused on the representation of global shape^{6,7} or how sensory signals combine when they are simultaneously used.^{8,9} It was found that vision dominates over haptics⁸ and that this dominance can be attributed to the lower sensory noise present in the visual channel.⁹ However, these studies (see also¹⁰) use flat stimuli that are all defined in the frontoparallel plane. There is little known about how the senses combine three-dimensional stimuli. As hypothesised above, haptics can potentially encode the Euclidean structure of a shape better than vision. We wanted to investigate whether this is true by measuring the influence of haptics on the perception of pictorial relief.

2. METHODS

2.1 Materials

The stimulus was produced with clay and is depicted in Fig. 1A. It was 17 cm wide, 12 cm high and 8 cm deep. We chose this type of form, because it possessed sufficient geometric complexity (protruding parts together with convex, concave and saddle-shape regions) and could be explored easily by hand. To investigate the influence of haptics on pictorial relief, the stimulus has to be touched and viewed simultaneously and should be perceived in the same location. To achieve this, a mirror was used as depicted in Fig.3. The haptic stimulus was the original clay object and the visual stimulus was a photograph of the original. Thus, a computer screen was used to present the experimental task. The clay object, mirror and computer screen were positioned on a table. The image on the computer screen was adjusted such that it subtended the same visual angle as the real object behind the mirror. To avoid muscular fatigue, an arm rest was available that supported the elbow. The position of the head was maintained constant by using a chin rest. Furthermore, observers were viewing monocularly.

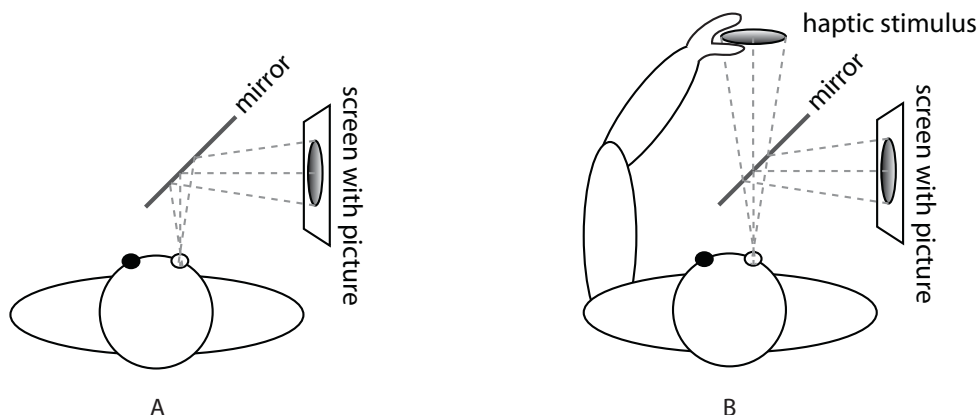


Figure 3. Top view of the experiment. Observer were looking monocularly through a mirror on a computers screen.

2.2 Participants and procedure

14 observers participated in the experiment. All had correct or corrected-to-normal vision. The participants were divided in two groups, one that performed the task with only vision and the other used both vision and haptics. Some of the observers had experience with the experimental task; they were distributed equally between the two groups. None of the participants had seen the stimulus and could not see the actual clay object before or during the experiment.

The participants were explained the procedure: adjust the orientation of the attitude probe such that it appears to lie on the surface of the object. The orientation of the probe could be manipulated with a mouse that was used with their right hand. In the haptic group the observers could explore the stimulus with their left hand. They could not see their hand because it was occluded by the mirror. It was explained that they could continuously explore the shape while doing the task.

The sample positions at which the probe appeared were acquired via a triangulation within the outline of the object, and can be seen in Fig. 1. Each position was measured once. The order of appearance was random and only one probe was visible. There was no time limit for adjusting the probe. The experiment took on average approximately 30 minutes.

2.3 Analysis

Each setting consisted of a slant and tilt pair (σ_i, τ_i) representing the local orientation of the pictorial surface at the image location (x_i, y_i) . As shown by Ref. 1 (see also Ref. 11 for more information about this procedure) these data can be used to calculate the subjective relief (x_i, y_i, z_i) where i denotes each sample point. In Fig. 1C a subjective relief is depicted. The depth data from the subjective relief, (z_i) are used for the analysis.

2.3.1 Straight regression

A straight forward way to compare the reliefs of two participants is to correlate them. This is similar to a linear regression $z' = \alpha z + \beta$ (greek symbols are used for the straight regression to avoid confusion with the affine regression) and calculate the coefficient of determination. In the regression, z' and z are depth values for *two different observers*. The calculation of adjusted r^2 is used to quantify how well two reliefs correlate. This analysis is performed for each participant pair within the two groups. Thus, 21 correlation coefficients are calculated per group. These can then be compared to see whether the haptic group produces reliefs that correlate better with each other than the vision-only group. Besides the similarity within the two groups, it is also possible to assess the difference between the experimental conditions.

2.3.2 Affine regression

As noted in the introduction, it has been shown that the affine structure of the pictorial reliefs of different observers is more alike than the Euclidean structure. To assess whether this is true, an affine regression can be performed in a similar way as described in the previous paragraph but now according to equation (3). It can be tested whether the affine regression explains the pair-wise data significantly better than the straight regression. Furthermore, the parameters of the affine regression, $\{a, b, c\}$ from equation (3) reveal the affine relation between the two reliefs. The parameter c reveals whether the relief differ in 'affine stretch' or, in other words, whether it is either flatter or more relieved than the other. This parameter was analysed between observers from the different conditions. Thus we can quantify whether haptics make the shape have more or less relief. The parameters $\{a, b\}$ define the 'additive plane' that reveals whether there is a shift in 'mental viewpoint'³ or in other words, whether the object is seen in a different affine orientation. This analysis was conducted within and between the different groups.

3. RESULTS

As an example, straight and affine regressions of two participants are shown in Figure 4A. As can be seen, the data correlate much better for the affine regression. Before discussing this further, we will first look at the straight regression data. As can be seen in Figure 4B, the correlation is higher in the haptics+vision group than in the vision-only group. The difference in r^2 between the vision-haptics group (0.74) and vision-only group (0.58) was significant ($t_{20} = 2.81, p < 0.01$).

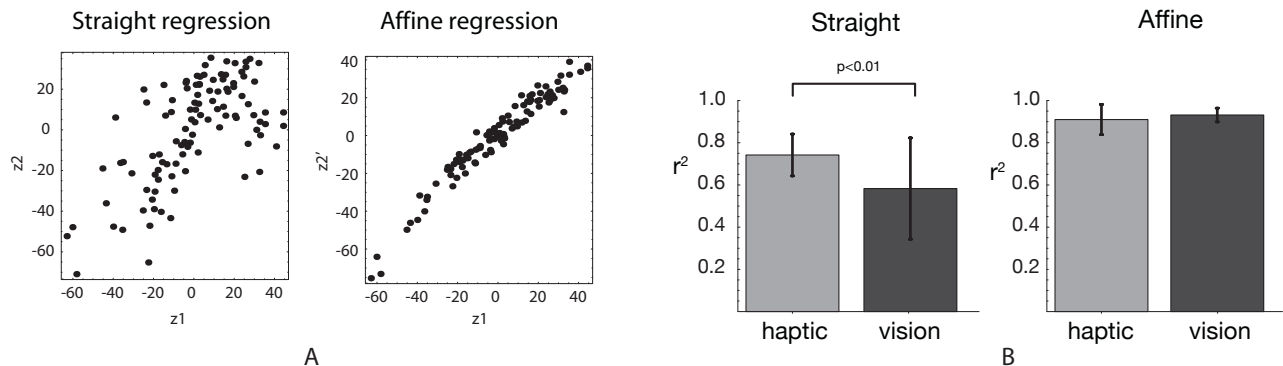


Figure 4. A. The depth values of two participants are compared. On the right, an affine transformation has been performed. B. Average correlation results expressed in adjusted r^2 . On the left, a clear difference can be seen between the haptics+vision and vision-only group. On the right, this difference is resolved after performing affine regressions.

The affine regression analysis showed higher correlation coefficients (adjusted r^2) than the straight regression, as can be seen in Figure 4B. This appeared to be true for both conditions although the vision-only condition seems to gain more. For each participant pair within a group we tested whether the correlation improved significantly when an affine transformation was performed. For the haptics+vision group, the improvement was significant ($p < 0.05$) for 10 out of 21 pairs, whereas for the vision group the improvement was significant for all 21 pairs. Secondly, the depth gains c between members of the two conditions were analysed. On average, the depth was 1.28 times larger ($t_{48} = 3.46$, $p < 0.005$) for observers who had touched the stimulus. Furthermore, we wanted to see what kind of affine shear differences could be found within and between the groups. This is shown in figure 5. The polar plot is explained in figure 5A. The orientation of the affine shear can be expressed in slant ϕ and tilt θ . The slant denotes how much the affine plane deviates from the viewing direction z ; the tilt denotes the orientation with respect to the image plane. As can be seen, the haptics+vision group seems to show more consistency than the vision only group, although it is difficult to test this statistically. The mean slants are 7.9° and 9.7° , for the haptics+vision and vision only group, respectively, although this difference is not significant ($p=0.1$). Furthermore, there is no indication of a preferential direction in the affine shear difference between the two conditions.

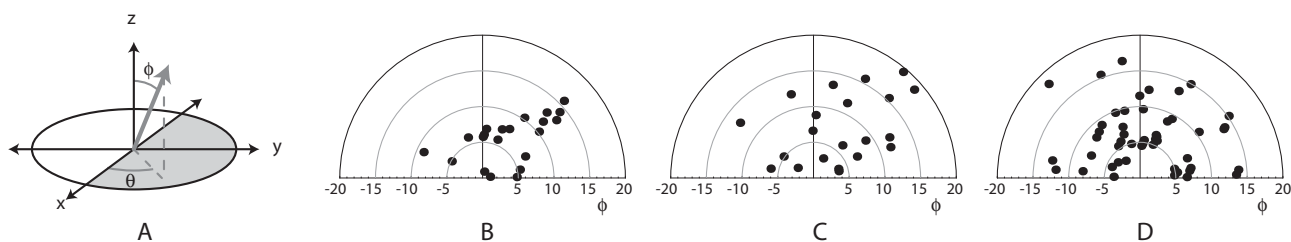


Figure 5. A. Explanation of the polar plots. The affine shear is parametrised by the magnitude of the shear ϕ and the direction with respect to the viewing direction θ . B. Haptics+vision, within group shears. C. Vision only, within group shear. D. Shears between the two groups.

4. DISCUSSION

Previous research in cross-modal visual-haptic perception has mainly focussed on three-dimensional object representational issues or combination of sensory input for two-dimensional stimuli. The research presented here provides new insight in how the senses combine when the stimulus is a three-dimensional object. The ambiguity that is often found in visual perception of pictorial reliefs is decreased by haptic input. This ambiguity is not solved completely, since the affine transform also improves the correlations for the haptics+vision group for most participant pairs. The affine shear analysis showed a similar tendency: affine shears are also present in the haptics+vision group, although there seemed to be more ‘directionality’ in the data, the directions of shears θ are more consistent than in the vision-only group. These findings raise questions both in the field of multisensory perception and in the application of visual displays. Maybe surprisingly, we found that a shape that is explored haptically, is perceived as having more depth. There could be two explanations for this effect. Firstly, haptics could induce more relief *in general*. It has been previously shown that viewing a real 3D shape with one eye decreases the relief as opposed to binocular vision.¹² A similar mechanism could be causing the haptic depth gain. The other explanation could be that with haptics, the percept becomes more fiducial and that the pure visual percept underestimates the depth relief. Since we do not have the exact geometrical shape description of our stimulus, we cannot resolve whether the haptics+vision group indeed perceives the shape more veridical. However, recent evidence¹³ indicates that the influence of haptics on visual shape perception indeed results in a more veridical shape percept.

The dominance often attributed to vision is based on two-dimensional stimuli.^{8,9} Introducing an extra dimension makes the research more ecological but also more complex. Our research showed that haptics can influence visual perception of shape, but the reverse remains unanswered. One of the reasons is that there is currently no method to investigate the structure of a haptic three-dimensional shape perception. A haptic version of a technique such as the attitude task is difficult to conceive. Future studies could explore the development

of paradigms with which this can be investigated. Furthermore, we hypothesised that in the context of sensory combination, vision could profit from haptic perception because haptics supposedly has better access to the Euclidean structure of the shape. As shown in a different study¹³ this indeed appears to be true. It is assumed that the sensory signals combine according to the maximum likelihood estimation model, as was shown for length⁹ and 2D shape perception.¹⁰ This model attributes more weight to the sensory signal that possesses less noise and assumes that the signals are unbiased. In our study, we cannot be so sure whether the visual percept is unbiased. It has been shown that observers perceive different reliefs, but are very consistent over multiple repetitions. Future studies could explore the possibilities of using the maximum likelihood estimation model to describe cross-modal perception of 3D shapes while taking into account these new challenges.

Independent of the underlying mechanisms, the data show that the inherent ambiguity present in visual images can be reduced by haptic perception. This could be important for applications that make use of visual images of three-dimensional objects. Using a haptic interface could improve the perception of shapes, although the haptic interface should provide sufficient shape information. Commercial interfaces generally provide only zeroth order shape information (position of the surface boundary) while it has been shown recently that first order shape information (local orientation of the surface) is the dominant shape cue.¹⁴

ACKNOWLEDGMENTS

This research was supported by grants from the Netherlands Organisation for Scientific Research (NWO) and a grant from the EU (FP7-ICT-217077-Eyeshots).

REFERENCES

- [1] Koenderink, J. J., van Doorn, A. J., and Kappers, A. M. L., "Surface perception in pictures," *Perception & Psychophysics* **52**, 487–496 (1992).
- [2] Todd, J. T., Koenderink, J. J., van Doorn, A. J., and Kappers, A. M. L., "Effects of changing viewing conditions on the perceived structure of smoothly curved surfaces," *Journal of Experimental Psychology-Human Perception and Performance* **22**(3), 695–706 (1996).
- [3] Koenderink, J. J., van Doorn, A. J., Kappers, A. M. L., and Todd, J. T., "Ambiguity and the 'mental eye' in pictorial relief," *Perception* **30**(4), 431–448 (2001).
- [4] Todd, J. T., "The visual perception of 3d shape," *Trends in Cognitive Sciences* **8**(3), 115–121 (2004).
- [5] [Source: Stanford University Computer Graphics Laboratory], <http://graphics.stanford.edu/data/3Dscanrep/>.
- [6] Newell, F. N., Ernst, M. O., Tjan, B. S., and Bulthoff, H. H., "Viewpoint dependence in visual and haptic object recognition," *Psychological Science* **12**(1), 37–42 (2001).
- [7] Norman, J. F., Clayton, A. M., Norman, H. F., and Crabtree, C. E., "Learning to perceive differences in solid shape through vision and touch," *Perception* **37**, 185–196 (2008).
- [8] Rock, I. and Victor, J., "Vision and touch: An experimentally created conflict between the two senses," *Science* **143**(3606), 594–596 (1964).
- [9] Ernst, M. O. and Banks, M. S., "Humans integrate visual and haptic information in a statistically optimal fashion," *Nature* **415**(6870), 429–433 (2002).
- [10] Helbig, H. B. and Ernst, M. O., "Optimal integration of shape information from vision and touch," *Experimental Brain Research* **179**, 595–606 (2007).
- [11] Nefs, H., "Three-dimensional object shape from shading and contour disparities," *Journal of Vision* **8**(11) (2008).
- [12] Koenderink, J. J., van Doorn, A. J., and Kappers, A. M., "Depth relief," *Perception* **24**(1), 115–126 (1995).
- [13] Wijntjes, M. W. A., Volcic, R., Pont, S. C., Koenderink, J. J., and Kappers, A. M. L., "Haptics disambiguates visual perception of 3d shape," *Experimental Brain Research* (in press).
- [14] Wijntjes, M. W. A., Sato, A., Hayward, V., and Kappers, A. M. L., "Local surface orientation dominates haptic curvature discrimination," *IEEE Transactions on Haptics* (in press).