Visuo-Tactual Coarseness Estimation Tasks
1: Experiments, 2: Modeling

Isidore Rigoutsos
Robotics Laboratory,
New York University,
715 Broadway,
New York, NY 10012, U.S.A.
rigoutso@cs.nyu.edu

Abstract

Visual perception has little to do with the retinal image of the viewed scene: although the scene's image is 2-dimensional, the perceived objects are 3-dimensional. It has been suggested that humans as well as other primates must learn to "see." Touch was considered to be the most likely source educating vision, presumably by adding meaning to the initially meaningless plethora of retinal images. Through a series of specifically designed experiments that focus on the task of *coarseness* estimation, we examine the relation between the two senses and demonstrate the occurence of visual capture in the presence of conflicting stimuli. Last, we suggest a model in agreement with the observed behavior.

1 Introduction

Visual perception has little to do with the retinal images of the viewed scenes. Objects are perceived as 3-dimensional; however, their retinal projections are 2-dimensional. Retinal images are sampled by distinct light sensitive receptors positioned at the vertices of an almost hexagonal lattice [16] but one perceives a single object with a definite structure. A retinal image of given size can be generated by either a small nearby object or a large distant one; however the perceived spatial relationships are quite accurate [15]. It was that striking difference between sensory information and the perceived world that gave rise to the Berkeleyan school of thought [1]. 18th century philosopher George Berkeley, along with other philosophers and psychologists, proposed that *tactual perception* provides the basis for *visual learning*: one must discover the rigid characteristics of 3-dimensional objects by resorting to tactual input.

Before we proceed it is necessary to make more precise the meaning of the term *touch* as this is used in the paper. By *touch* we mean ([14]):

> "...the active tactual or haptic perception [as that would be derived from] the position of the fingers relative to one another, [the sense of position of our body parts] and so forth ...[as well as] the passive tactile perception as when a form is impressed against the skin..."

Berkeley's approach, although appealing at first, is nondefensible [14]. First of all, it advocates for the existence of innate touch sense on which visual perception is progressively built; however the argument which rules out the existence of *innate visual* sense can also be used to argue against any *innate touch* sense. There is also considerable evidence pointing to the innate basis of form perception [17]. Second, experimental evidence suggests that the visual sense is well developed soon after birth or even is innate for some nonhuman infants tested on the *visual cliff* [3]. Third, the sense of touch is not precise enough to account for the experienced perceptual accuracy of spatial relations. Finally, as experiments on size/shape perception and positioning [15] revealed, when vision and touch are in conflict, not only is vision dominant over touch but it also leads to "...the misperception of the 'feel' of the object..." ([14]).

The observed dominance leads, in the case of size and shape estimation, to the so-called *visual capture*, whereas in the case of positioning experiments it results to some form of *recalibration* of tactual perception.

Although the human ability to *visually* estimate roughness with great accuracy has been discussed in previous work [13], no studies have, to the best of our knowledge, examined the occurence or not of visual dominance during *coarseness* estimations tasks; this is precisely what we will be concentrating on in the sequel.

The remaining of this paper is structured as follows:

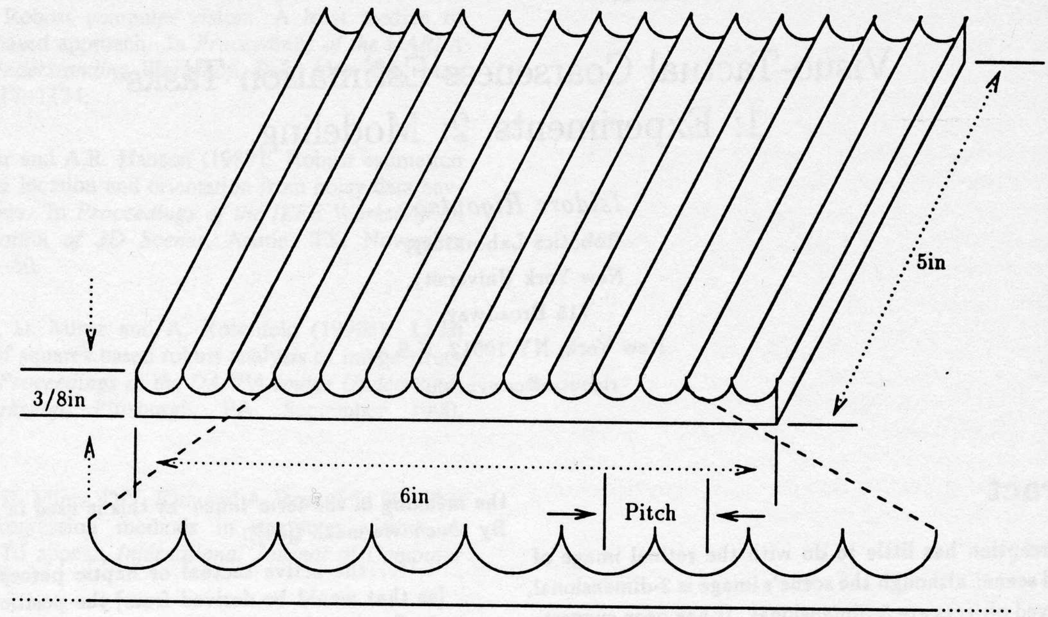

Figure 1. The plates.

in **Section 2** we elaborate on the experimental arrangements and present the designed procedures together with the underlying rationale. **Section 3** contains the experimental results followed by an analysis. In **Section 4** we propose a model in agreement with the observed experimental behavior. **Section 5** contains a discussion on the suggested model.

2 Experimental Details

2.1 The Templates

The objects used in our experiments were thin rectangular plates of industrial wax having a certain pattern carved on one of their two large facets (Fig. 1). There was no particular reason for having chosen this texture pattern other than that its generation was simple. This particular choice has no shortcomings, because as Gibson points out [14], in both visual and tactual perception it is the spatial information that is of importance and not the specific *sensory pattern* of stimulation.

In order to machine the wax plate patterns we made use of the N.Y.U. built CNC [4] milling machine. This is a programmable drilling/carving tool that allows continuous control of the 3-dimensional position of the object being carved with an accuracy as fine as 1/1,000 of an inch. The control is achieved by a program generated by a CAD/CAM package [11]. The tool provides also for interchangeable bits of different sizes and shapes. To carve our surfaces we used standard steel ball-end-mills; under program control the table on which the wax plates were laid was moved along equally spaced line segments, parallel to the short side of the plates.

Six distinct patterns were generated in the above manner: their perceptual differences originated from the fact that the values of the pitch parameter, i.e. the densities of the "ridges" per unit length, were *pair-wise* distinct. We used a total of six different ball-end-mill bits, with diameters: 1/16, 1/8, 3/16, 1/4, 5/16, and 3/8 of an inch respectively. To generate grooves with the semicircular shape shown in Fig. 1, the pitch value was set to one diameter, whereas the tip of the bit was at a depth equal to one half of the corresponding pitch. The patterns were distinguishable both visually and tactually and one pair of plates was generated for each. In what follows, we will be refering to those plates by their rank in increasing pitch order, i.e. the plate with a (1/16)" pitch value will be plate #1, the one with a (1/8)" pitch value will be plate #2 e.t.c.

2.2 A Description of the Experiments

Our experiments follow [15] and are divided in two classes; in the first class we have the following two categories:

- The subject is presented with a pair of plates, let us call them A and B respectively, and is allowed to see (but not touch) plate A while touching (without seeing) plate B, for a period of 5 seconds. We then assess the subject's perception by asking the subject to determine the relative coarseness (pitch value) of the two plates A and B.

- The same experiment as above is repeated, but this time *after* the subject has been exposed to conflicting information for an extended period (see below).

In the first category, the subjects employed always the same hand. In the second, roughly half of the subjects employed the hand they also employed during the exposure to conflicting information; the remaining subjects employed their other hand. All the template pairs used in the experiments differed by *at most* one pitch value.

The purpose of these experiments is two-fold: the first category is designed to determine whether the 5 second interval is sufficient time for a subject to correctly determine the relative coarseness of the two plates. On the other hand, the second category provides an independent method for determining which of the two modalities, namely visual or tactual, is affected by the exposure to the conflicting stimuli. If vision were the affected sense, the choice of the hand used would be immaterial; of course this would be no longer true if touch were affected instead.

We would also like to determine whether persistence of conflicting input over a long time interval would induce a change in the tactual perception. To this end, the second class consists of four categories of experiments:

- The subject is *shown* different plates in turn and is asked to *tactually* match each one of them, from a reference set. The subject is subsequently exposed to conflicting information (using appropriately chosen pairs of plates) over the course of one half hour. Then the previous task is repeated.
- The subject is allowed to only *touch* different plates in turn and is asked to *visually* match each one of them, from a reference set. The subject is then exposed to conflicting information (using appropriately chosen pairs of plates) over the course of one half hour. At the end of the 30 minute period the previous task is repeated.
- The subject is initially *shown* a standard plate[1] and *practices* to *visually* match it, from immediate recollection, from a reference set. The subject is subsequently exposed to conflicting information (using appropriately chosen pairs of plates) over the course of one half hour. As soon as the 30 minute period is over, the subject is asked to *visually* identify the plate that has the same coarseness as he remembered the standard plate to have.
- The subject *touches* the standard plate and practices to *tactually* match it, from immediate recollection, from a reference set. The subject is subsequently exposed to conflicting information (using appropriately chosen pairs of plates) over the course of one half hour. At the end of this period the subject is asked to *tactually* identify the plate that has the same coarseness as he remembered the standard plate to have.

The purpose of the first two categories is to reveal whether the subject's perception changes after exposure to conflicting information for an extended time interval; in this way we will be able to determine the occurence of any induced change in the involved modalities. However, one still needs to determine which of the two modalities changes. Does the visual system adapt so that it matches the conflicting "tactual" input, or does the tactual system change to match the conflicting "visual" input? This answer is provided by experiments of the last two categories: if *vision* changes then the subject will not be able to correctly identify *visually* the standard square after the 30 minute interval; however, he will still be able to perform the identification *tactually*; an analogous reasoning applies to the sense of *touch*. To further confirm the findings of each individual category of this class, the experiments were in most cases followed up by those in the first class (second category).

3 Results & Analysis

The results of our experiments for the different classes have as follows:

Class: A Category: 1

Number of subjects:	19
Total number of employed pairs:	76
Times relative coarseness was correctly estimated:	70
Times coarseness of touched plate was underestimated:	1
Times coarseness of touched plate was overestimated:	5

Class: A Category: 2a

Number of subjects:	7
Total number of employed pairs:	28
Times relative coarseness was correctly estimated:	26

Class: A Category: 2b

Number of subjects:	8
Total number of employed pairs:	32
Times relative coarseness was correctly estimated:	9

The most interesting experiments were those of the second class. Thirty seven subjects participated in those experiments. The conflicting information was generated using

---

[1] This is the one created using the (3/16)" bit

plate pairs whose pitch values were one step apart: the subject was shown plates with pitch value always one larger than the plate's the subject touched. A typical sequence for those pairs was: (Shown, Touched)≡(6, 5) (5, 4) (4, 3) (3, 2) (2, 1) (3, 2) (4, 3) (5, 4) and (6, 5).

Class: B Category: 1

Number of subjects:	9
Total number of matching trials:	27
Times subject *tactually* recovered plate of next *smaller* pitch:	18

Class: B Category: 2

Number of subjects:	9
Total number of matching trials:	27
Times subject *visually* recovered plate of next *larger* pitch:	17

Class: B Category: 3

Number of subjects:	10
Times standard plate was *visually* recovered:	10

Class: B Category: 4

Number of subjects:	9
Times standard plate was *tactually* recovered:	2

Figure 2 graphically represents the results for four of the subjects. In order to be able to represent the results on the same graph, the experimental points of the second subject were shifted by one half unit in both axes.

The implications of the above results are clear. First, it was demonstrated that 5 seconds was sufficient time for the subjects to develop a clear understanding and establish the actual relation between the physical characteristics of the objects they handled. Indeed, in 92% of the trials, the subjects had a definite unitary experience of the templates' coarseness.

In order to appreciate the results of B-1 and B-2 above, we proceed with a standard sign test analysis[10]. Let X_i (resp. Y_i) be the template number that generated the tactual (resp. visual) stimulus during the i-th matching trial. Let also

$$Z_i = \begin{cases} 1, & \text{if } Y_i - X_i > 0 \\ 0, & \text{if } Y_i - X_i \leq 0 \end{cases}$$

If $p \equiv Pr(Z_i = 1)$ then from A-1 we can see that the probability of overestimating the coarseness of the touched plate is $p \approx 5/76 = 0.0660$. Let us now hypothesize that the subjects' exposure to the conflicting stimuli does not alter their ability to tactually estimate pitch values; then Z_i is a binomial variable corresponding to a single trial of an experiment for which $p = 0.0660$ and $q = 1 - p = 0.9340$. The probability of occurence of the results in B-1 & B-2 will then be $b(18 + 17, 27 + 27, 0.0660) = 2.4 * 10^{-28}$. If we instead assumed that $p = q = 0.5$, the resulting probability would equal $b(18+17, 27+27, 0.5) = 1.0 * 10^{-2}$. In both cases the calculated probabilities are very low.

The experiments in B-3 provided evidence that vision had not changed after the exposure to conflicting information: the subjects were able in all cases to correctly identify the standard template from memory. On the other hand, only 2 out of 9 times (B-4) were subjects able to tactually recover the standard plate from memory: the templates now felt coarser than they really were. These results were further corroborated by those of A-2a & A-2b.

All of the above leads to the conclusion that the affected modality was indeed touch and not vision: the "sense" of touch underwent a change in order to match the visual input. As a result, the tactually estimated coarseness was almost always shifted by one pitch value.

In previous size and shape determination experiments [15] the conflicting stimuli were created by a distorting optical lense whose existence the subjects were not aware of. This guaranteed that they would not resort to a concious decision of which sense to rely on. An interesting observation was also made during our experiments: in a number of cases the subjects reported they suspected they were being exposed to conflicting stimuli. However, this did *not* prevent their sense of touch from adapting and eventually matching the visual input. Apparently, adaptation occurred despite their suspicion of the existence of a conflict. How can this be explained? It seems that *nature* has chosen to give more credit to the sense of vision and has consequently *hard-wired* the sensory pathway in such a way that adaptation is induced even if cognition appreciates it as unnecessary. Interestingly enough, experiments on the *visual cliff*[3] call for conclusions along the same lines: infant reactions were directly related to the role vision played to the well-being of their species.

4 The Model

With regard to the organization of the sensory pathway pertaining to the senses of vision and touch, the observed experimental behavior suggests an organisation like the one in figure 3. This model is general enough to account for the observed behavior in size/shape estimation experiments as well.

The environment provides the input to the touch[2] and vision sensors whose appropriately thresholded outputs enter the stimulus associator (an instance of a pattern associator). Normally, the associator's output is a perception of the coarseness of the object causing the given activation pattern; the associator is also able to determine the occurrence

---

[2] Recall our definition of "touch" from the introduction.

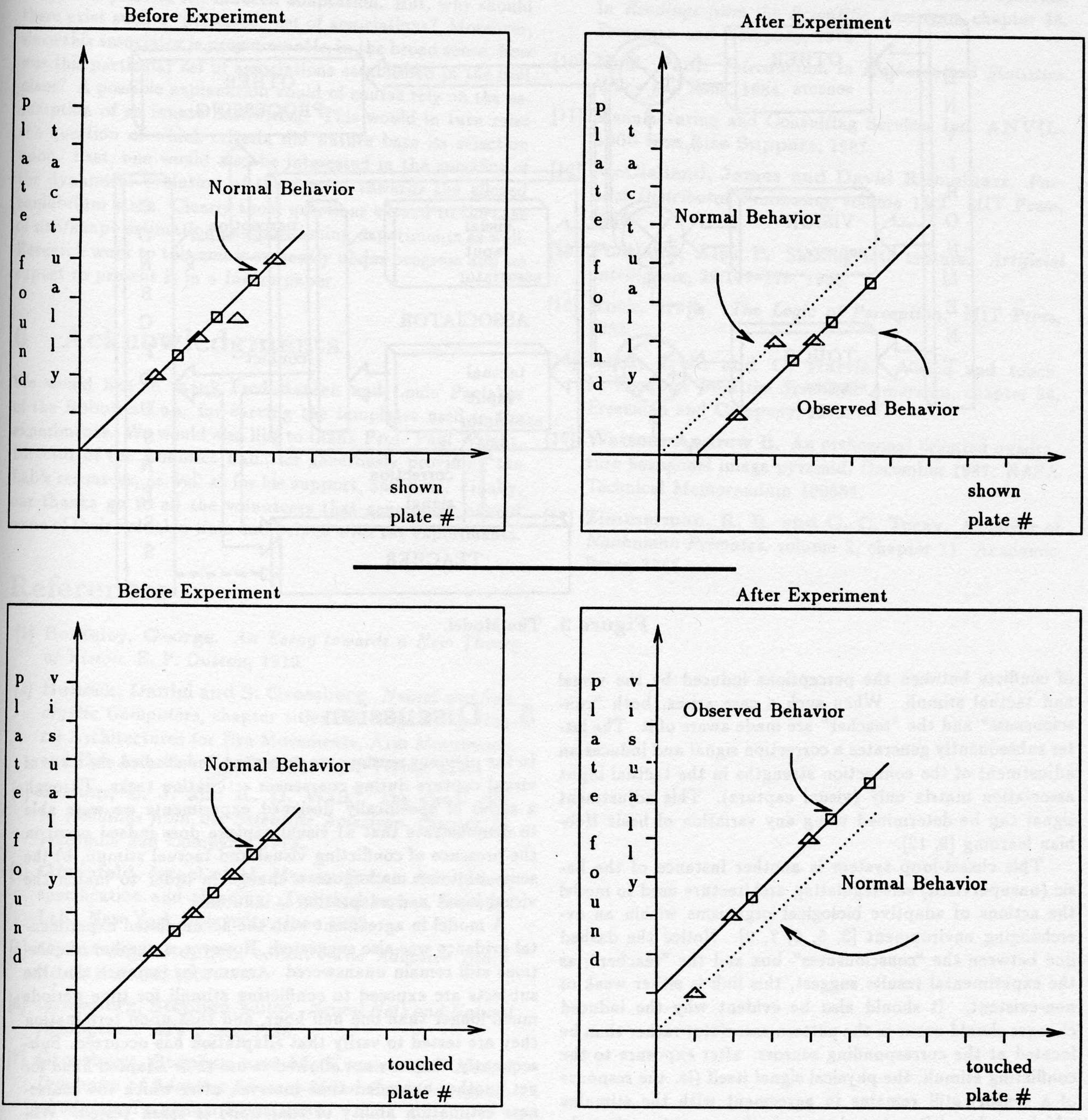

Figure 2. Class:B Categ.:1 Subjects IN (□), PB (△) Categ.:2 Subjects JM (□), SF (△) (see text).

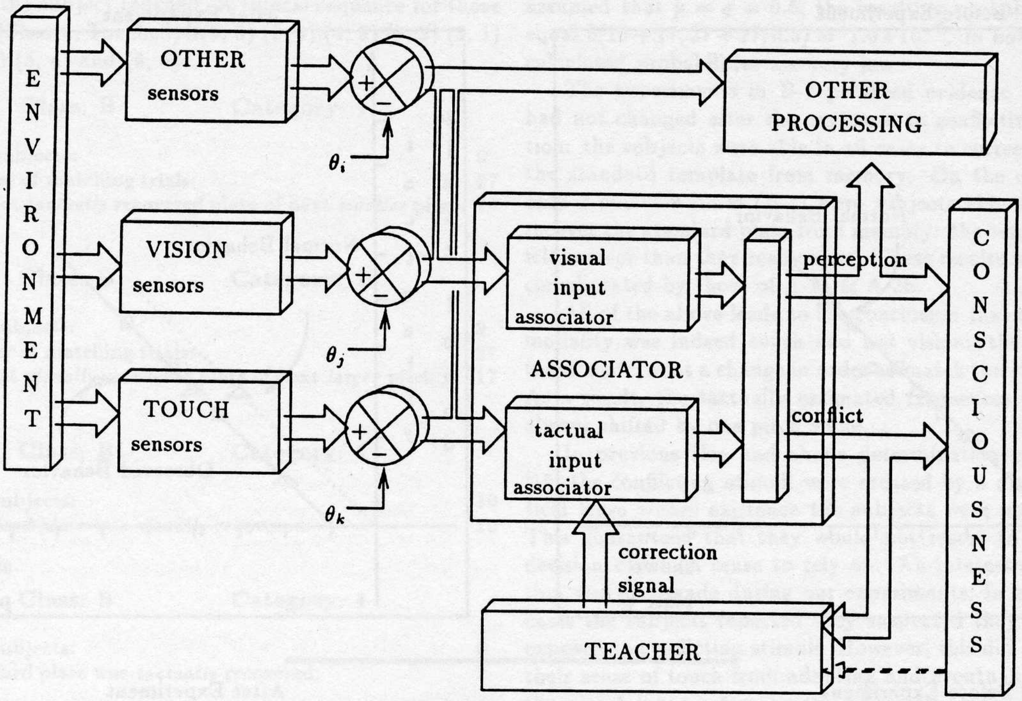

Figure 3. The Model.

of conflicts between the perceptions induced by the visual and tactual stimuli. When such a case arises, both "consciousness" and the "teacher" are made aware of it. The latter subsequently generates a correction signal and induces an adjustment of the connection strengths in the tactual input association matrix only (visual capture). This adjustment signal can be determined using any variation of basic Hebbian learning [8, 12].

This closed-loop system is another instance of the basic (unsupervised) self-regulating architecture used to model the actions of adaptive biological organisms within an everchanging environment [2, 5, 6, 7, 9]. Notice the dashed line between the "consciousness" box and the "teacher"; as the experimental results suggest, this link is either weak or non-existent. It should also be evident why the induced changes should occur in the pattern associator rather than be located at the corresponding sensors: after exposure to the conflicting stimuli, the physical signal itself (ie. the response of a sensor) still remains in agreement with the stimulus which produced it; what changes is the *perception* the subject now associates with the given stimulus. The suggested model, although simple, accounts well for the observed behavior.

5 Discussion

In the previous sections we presented and studied the issue of visual capture during coarseness estimation tasks. Through a series of specifically designed experiments we were able to demonstrate that a) visual capture does indeed occur in the presence of conflicting visual and tactual stimuli, b) the sense of touch undergoes a change in order to match the visual input and adaptation is induced.

A model in agreement with the accumulated experimental evidence was also suggested. However, a number of questions still remain unanswered. Assume for instance that the subjects are exposed to conflicting stimuli for time periods much longer than one half hour, and that upon termination they are tested to verify that adaptation has occurred. Subsequently, they are not allowed to use their adapted hand for yet another extended time interval, after which the coarseness estimation ability of that hand is again tested. Will adaptation still be observed? Or, will the tactual pattern associator have "unlearned" the associations induced by the exposure to the conflicting stimuli? The first alternative is not an interesting one. On the other hand, if the associator has unlearned, this would imply the existence of a *quiescence*