# Perceptual Similarities Amongst Novel, 3D Objects

**Dissertation**

**zur Erlangung des Grades eines Doktors der Naturwissenschaften**

**der Fakultät für Biologie**

**und**

**der Medizinischen Fakultät der Eberhard-Karls-Universität Tübingen**

**vorlegt**

**von**

**Theresa Cooke**

**aus Toronto, Kanada**

**2006**

| | |
|---|---|
| Tag der mündlichen Prüfung: | 15. Januar 2007 |
| | |
| Dekan der Fakultät für Biologie: | Prof. Dr. Ingo B. Autenrieth |
| Dekan der Medizinischen Fakultät: | Prof. Dr. Friedrich Schöffl |
| | |
| 1. Berichterstatter: | Prof. Dr. Heinrich H. Bülthoff |
| 2. Berichterstatter: | Prof. Dr. Hanspeter Mallot |
| | |
| Prüfungskommission: | Prof. Dr. Heinrich H. Bülthoff |
| | Prof. Dr. Hanspeter Mallot |
| | Prof. Dr. Martin Giese |
| | Prof. Dr. Dr. Hans-Otto Karnath |
| | Prof. Dr. Uwe Ilg |

# Table of Contents

# Declaration

I declare that I have produced the work entitled "Perceptual Similarities Amongst Novel, 3D Objects", submitted for the award of a doctorate, on my own (without external help), have used only the sources and aids indicated and have marked passages included from other works, whether verbatim or in content, as such. I swear upon oath that these statements are true and that I have not concealed anything. I am aware that making a false declaration under oath is punishable by a term of imprisonment of up to three years or by a fine.

Theresa Cooke
November 6, 2006
Tübingen, Germany

*This thesis is dedicated to my mother,*
*for her unfaltering love, for her gentle compassion,*
*and for her indomitable spirit.*

# Acknowledgements

First and foremost, I would like to thank Christian Wallraven for his constant support and encouragement throughout my graduate studies. This thesis would not have been possible without his creativity, openness, and readiness to help.

I am also very grateful to Heinrich Bülthoff for creating an outstanding research environment at the Max Planck Institute for Biological Cybernetics, welcoming me into it, and supporting me both financially and scientifically throughout my stay.

I would like to thank Frank Jäkel for his comments on this manuscript, his patience, dedication, and scrupulous attention to detail. Hannah Helbig also provided helpful references and comments during the preparation of this manuscript.

I am indebted to Douglas Cunningham for the many hours he generously spent teaching me about psychophysics and to Marc Ernst for his help in supporting my work on haptic device validation.

I am very grateful to Horst Herbert for his instrumental role in making my stay in Tuebingen a success and to Rainer Hirt for extremely professional assistance with the administrative aspects of my stay.

Last, but certainly not least, I would like to thank Martin Breidt, Mirko Thiesen, Walter Heinz, Bianca Arsene, Karin Bierig, Steffen Baier, and Stefan Weber, who provided invaluable technical assistance with producing stimuli, designing and building experimental setups, and running the experiments.

# Summary

Perceptual similarity is thought to play a fundamental role in human cognition. In this Ph.D. thesis, a set of parametrically-defined, novel 3D objects were used to investigate perceptual similarity from three new angles: (1) studying the effects of seeing versus touching, (2) comparing the effects of touching using different hand movements, and (3) comparing human similarity judgments to similarities computed by state-of-the-art algorithms. In all three studies, similarities were analyzed using multidimensional scaling (MDS) techniques, which provided maps of the objects in "perceptual similarity space," as well as the relative weights of perceptual dimensions. Maps and weights were then compared across conditions. In the first study, differences between judging similarity using vision, touch, or both were demonstrated and modeled by linearly rescaling a single perceptual map. The second study showed that using different hand movement patterns can also lead to changes in perceptual similarity. Furthermore, when a given hand movement allowed for the extraction of more than one object property, relative property weights varied across individuals; these individual biases remained stable over several months. In the third study, good fits were found between perceptual spaces derived from human vision and those derived from algorithms biased towards shape extraction, however none of the algorithms tested provided a good overall fit to haptic data, in which large intersubject variability was observed. Overall, the thesis highlights the role of feature extraction mechanisms in determining similarity and demonstrates how MDS techniques can be used to visualize and quantify these effects. Using MDS, differences in similarities could be accounted for by linear rescaling a single perceptual map, illustrating a potential mechanism for connecting unimodal and multimodal representations. The thesis has also led to the development of a new, similarity-based approach to perceptual validation of virtual environments and feature extraction algorithms.

# Chapter 1

# Introduction

In a recent review of research on similarity, Goldstone and Son (2005) provide a long list of cognitive abilities which depend on the ability to judge similarity, including reasoning, problem-solving, perceptual recognition, and categorization. They divide similarity research into two classes: research on conceptual stimuli, such as theories or stories, and research on perceptual stimuli, such as colours, textures, sounds, odors, and tastes. In the latter group, most studies have examined similarity judgments performed using a single sensory modality; however, judging the similarity between objects in the real world often requires that information from more than one sense be combined. For instance, comparing two pieces of fruit involves combining information about the colour and texture of their peel, the pungency of their smell, the softness of their fruit, and the sweetness of their taste. Not only does information about *distinct* object properties coming from different senses need to be combined to judge similarity; whenever a single object property can be extracted using more than one modality, the two sources may provide redundant or conflicting information. This raises the question of how disjunct as well as overlapping information provided by multiple sensory systems is combined to subserve similarity judgments.

One of the challenges in studying multimodal object similarity, at least for the case of visuohaptic similarity, is the difficulty of producing large numbers of novel stimuli with fully controllable properties, which can be both seen and touched. Until recently, visuohaptic stimuli have been created by precision-cutting (Klatzky et al., 1987; Lakatos and Marks, 1999), casting (Norman et al., 2004), hand-moulding (Garbin, 1984; James et al., 2002; Forti and Humphreys, 2005), or assembling toy bricks (Newell et al., 2001; Forti and Humphreys, 2005). In this thesis, a new combination of computer graphics and 3D printing technology was used for stimulus creation. Designing objects using computer graphics software provides full control of object properties and the opportunity to create completely novel objects with which participants have had no prior experience.

Stimuli can be easily modified, replicated, and rendered into 2D images or printed into 3D models. Furthermore, since one can change object properties incrementally in software, it is possible to create an arbitrarily large family of objects with parametrically-defined differences. The objects can then be visualized as a set of points in a multidimensional "input space" spanned by the parameters varied in software.

Having a well-defined input space is particularly advantageous for studying similarity because of a long tradition of visualizing similarity data using spatial models, in which the distance between two items is related to the similarity between them. This approach was pioneered by Richardson's characterization of a psychological colour space based on comparisons made amongst Munsell colour samples (Richardson, 1938). Research on the structure of such psychological spaces has been closely connected to the development of multidimensional scaling (MDS) techniques (seminal work by Torgerson (1952); Shepard (1962)). MDS techniques operate on pairwise similarity ratings taken over a set of stimuli and return a map of objects in which distances have been fit to similarity data. The techniques have been applied to study mental representations of a wide range of stimulus classes, either to 'discover' them or to test specific hypotheses about them. Some examples include colours (Ekman, 1954), Morse code patterns (Shepard, 1963), spices (Jones et al., 1978), textures (Hollins et al., 1993; Bergmann Tiest and Kappers, 2006), synthetic tones (Caclin et al., 2005), salts (Lim and Lawless, 2005), and facial expressions (Bimler and Paramei, 2006). MDS techniques have also been used to study how mental representations differ across individuals. For example, Bosten et al. (2005) recently used them to show that colour-deficient observers are sensitive to an additional dimension of colour variation which is not perceived by colour-normal observers. In addition to providing helpful visualizations, spatial models derived from human similarity data have proven to have significant power for predicting human performance in identification, recognition and categorization tasks and have been used to formulate a number of influential models of these tasks (Gillund and Shiffrin, 1984; Kruschke, 1992; Nosofsky, 1991; Edelman, 1999).

In this thesis, we adopted the spatial modeling approach to investigate similarity from three new angles. We began by defining a two-dimensional input space of novel, 3D objects. In a first study, we investigated how this space was perceived by humans when they explored the objects using either vision, touch, or both at the same time. In a second study, we focused on the haptic modality and investigated how changes in the type of hand movement used to explore the objects affected similarity space. We also tested the stability of perceptual similarity over time. In a third study, we compared similarity spaces generated by human vision and touch against similarity spaces based on computationally-extracted object features.

The thesis is structured as follows: we first provide an overview of the common approach used to gather and analyze similarity data, then summarize each of the three papers presented as part of the thesis, and close by considering future avenues of research in multimodal similarity.

# Chapter 2

# Summary of Methods

This section provides an overview of the approach used in the thesis. It consists of the following four steps, illustrated in Figure 2.1:

1. creating a set of parametrically-varying, novel, 3D objects over which similarities are to be measured;
2. measuring similarity data under a specific set of experimental conditions;
3. multidimensional scaling (MDS) of similarity data to generate perceptual maps of objects and corresponding dimension weights;
4. comparing perceptual maps and weights to quantify the effects of changes in experimental conditions.

## 2.1 Creating a set of novel, 3D objects

We began by creating the family of objects shown in Figure 2.2. The stimuli were designed to be novel and parametrically-related to one another. Three-dimensional models were created using a 3D graphics software package. These models could then either be rendered into 2D images (with full control of size, viewpoint, illumination, background scene, etc.) or manufactured into real, touchable models via 3D printing. Object parameters were manipulated in order to create variations in the objects' macrogeometry ("shape") and microgeometry ("texture"). We chose to manipulate geometry specifically because it is a property which can be extracted using both vision and touch. At the same time, macrogeometry is preferentially extracted by vision, while microgeometry, which can also be considered a material property, is preferentially extracted by touch (Lederman et al., 1996).

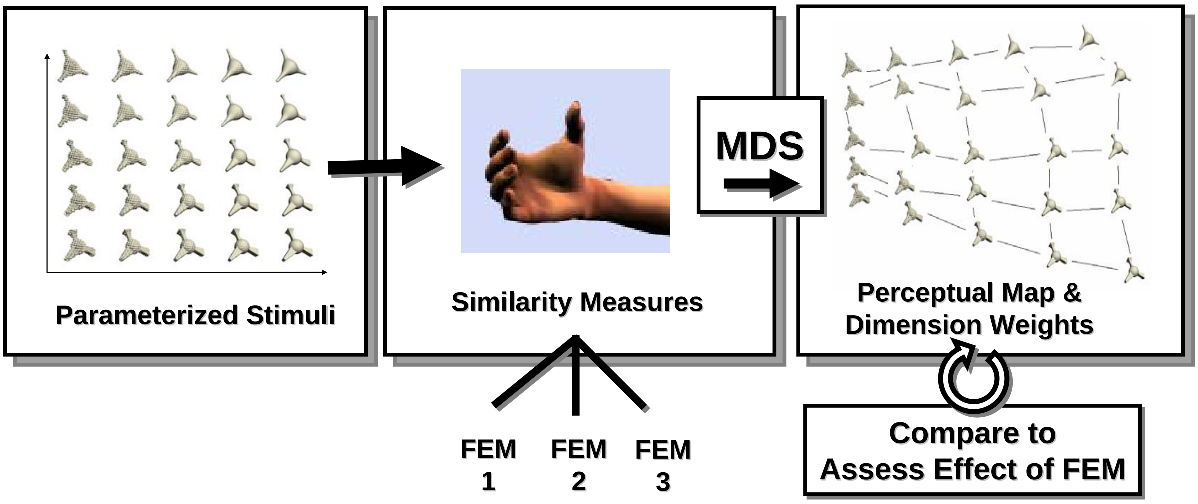


Figure 2.1: General Approach: (1) Creation of a parametrically-varying stimulus set; (2) gathering similarity data using specific feature extraction mechanisms (FEM); (3) MDS analysis of similarity data; (4) comparison of perceptual spaces to assess the effects of using different mechanisms to judge similarity.

Moving upward from the bottom row of objects in Figure 2.2, the objects' shape was gradually smoothed by "relaxing" angles in the 3D mesh via a local averaging operation. The objects' texture was generated by applying a local displacement map to the 3D mesh. Moving from left to right in Figure 2.2, texture was gradually smoothed by decreasing the amount of vertex displacement dictated by the map. Although Figure 2.2 shows the stimuli equidistantly spaced along two dimensions (the primary ones manipulated in software), it is important to note that there is no necessary relationship between these dimensions/distances and those *perceived* by participants. Characterizing the perceptual space is precisely the reason for using multidimensional scaling techniques.

## 2.2 Collecting similarity data

For human experiments, similarity data consisted of ratings provided on a seven-point scale (1 = low similarity; 7 = high similarity). For computational experiments, similarity data consisted of distance measures computed on the outputs of the various feature extraction algorithms. Each of the three studies involved a different variation in the way similarity data was generated:

- In Cooke et al. (2007a), the sensory modality used to explore the objects was varied. Object properties were extracted by the visual system, the haptic system, or both simultaneously.

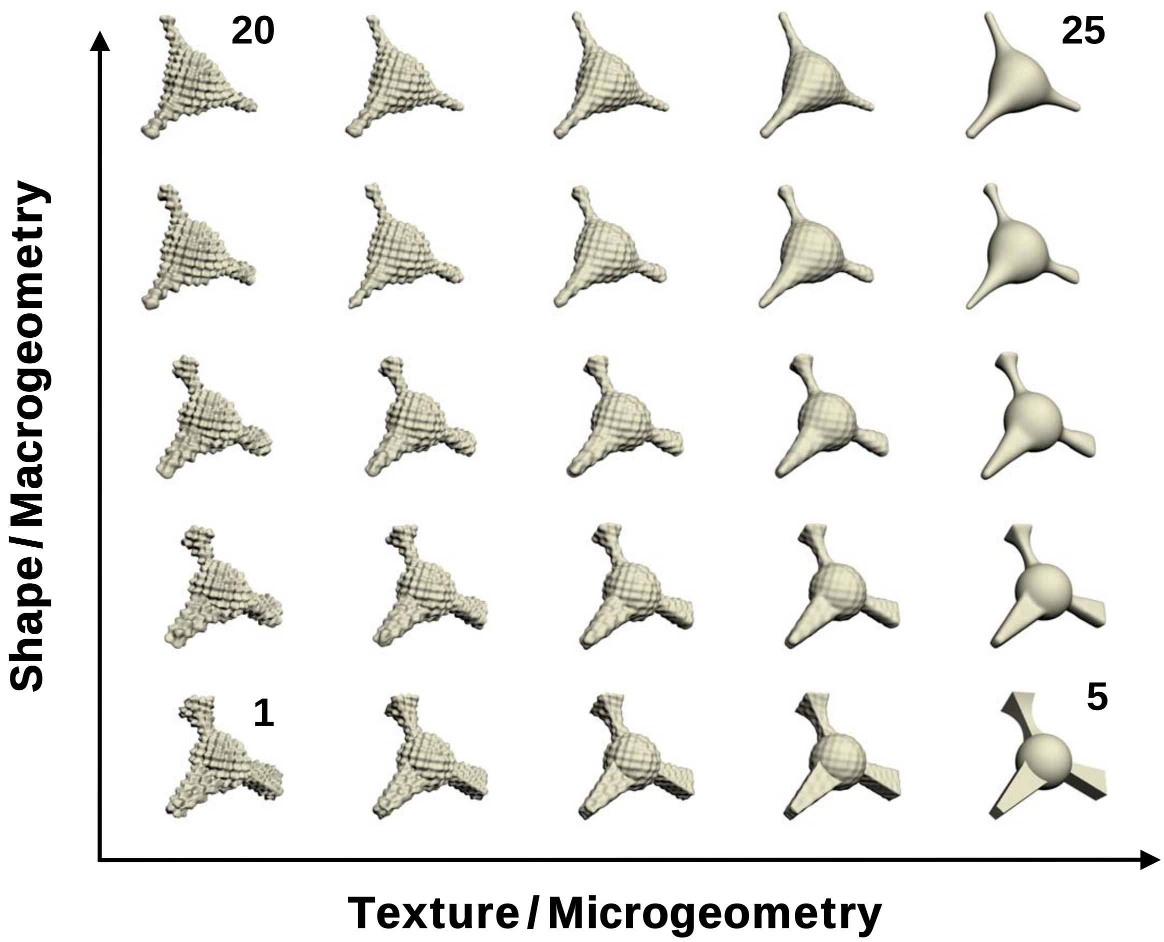


Figure 2.2: Stimuli varied parametrically in terms of microgeometry ("texture") and macrogeometry ("shape").

- In Cooke et al. (2007b), the hand movement used to explore objects was varied. Object properties were extracted by following their contours, by rubbing either their centres or their extremeties, or by gripping them.
- In Cooke et al. (2006), human visual and haptic extraction were compared against machine extraction using computational algorithms which extracted features from either 2D images or 3D models of the objects.

## 2.3 Analyzing similarity data using MDS

Similarity data were analyzed using a set of algorithms referred to as multidimensional scaling (MDS) techniques. MDS techniques refer to a class of algorithms which operate on any type of pairwise proximity data collected over a set of items, such as similarity ratings, and return a spatial model of the items. When the input proximities correspond to human similarity ratings, the output spatial configuration can be interpreted as a "psychological space" (Attneave, 1950). Since we have chosen to study perceptual stimuli, we refer to this space as a "perceptual space."

Human similarity matrices were analyzed using two variants of MDS, both implemented in the SPSS ALSCAL MDS package. The first variant, replicated MDS (RMDS), allows for the simultaneous analysis of several similarity matrices. For a user-specified dimensionality, it returns a single map and a goodness-of-fit measure, referred to as "stress", which corresponds to a normalized difference between the fitted distances and the observed proximities. RMDS was used to analyze groups of similarity matrices gathered under a single experimental conditions; maps and stress values were computed for solutions of one up to five dimensions. The second variant, weighted MDS (WMDS), also referred to as individual differences scaling (INDSCAL), takes in in one input matrix of similarities per subject and returns a single stimulus configuration together with a set of weights for each subject. The weights specify how the dimensions of the perceptual configuration should be stretched in order to best fit the individual matrices. WMDS also has the advantage of specifying the orientation of the configuration relative to perceptual dimensions, which facilitates the task of interpreting the dimensions.

Together, these analyses provided the four following types of information about perceptual space:

- How many dimensions of stimulus change were used by subjects to judge similarity, i.e., *the dimensionality of perceptual space*;
- Whether or not these dimensions corresponded to the dimensions which were manipulated, i.e., *the interpretation of perceptual space dimensions*;
- What the relative importance of the dimensions used to judge similarity was, i.e., *the weights of the perceptual dimensions*;
- A visualization of similarities amongst objects as distances between them, i.e, *the configuration of stimuli in perceptual space.*

## 2.4 Assessing differences between perceptual similarity spaces

Finally, to quantify the effects of our experimental manipulations, data gathered under different conditions were compared as follows:

- The *dimensionality of perceptual space* was determined using stress values output by MDS. Although the interpretation of stress values is somewhat controversial (Lee, 2001), a sharp drop in stress values for a given dimensionality, usually resulting in stress values below 0.2, is conventionally interpreted as evidence that the output configuration provides a good fit to the data (Kruskal and Wish, 1978).

- The *interpretation of perceptual space dimensions* was done by first comparing the order of stimuli in perceptual space against the order of stimuli in the input space - finding the same order in both cases implies that the input dimensions were recovered in perceptual space - and second, by analyzing debriefing questionnaires to identify labels used by subjects in referring to the differences amongst stimuli.

- The *weights of perceptual space dimensions* output by MDS were compared by carrying out T-tests on the mean of population samples.

- The *configuration of stimuli of perceptual space* was compared by visual inspection and by computing Procrustes fit errors between two configurations.

The following section summarizes the results obtained by applying this methodology.

# Chapter 3

# Overview of Results

This section summarizes the findings reported in the three papers which comprise the thesis.

## 3.1 Multimodal similarity and categorization of novel, three-dimensional objects

This paper (Cooke et al., 2007a, Paper 1) investigated how perceptual similarities and categorization vary when different sensory modalities are used to explore objects. Subjects explored the objects (described in the Methods section) using either vision alone, touch alone, or both vision and touch. MDS techniques were used to obtain perceptual spaces arising from each of these conditions. The spaces were then compared in order to test for an effect of modality. Given the close connections between similarity and categorization (Hahn and Ramscar, 2001), we were also interested in investigating modality-specific differences in categorization; therefore, we had subjects categorize the objects at the end of the similarity rating experiment.

We found similarities as well as differences amongst the representations recovered from haptic, visual, and visuohaptic exploration. Regardless of modality, subjects referred to the dimensions used to judge similarity as "shape" and "texture" exclusively (Figure 7 in Paper 1, centre column). The same ordinal relationships amongst stimuli in the input space were found in MDS spaces reconstructed from visual, haptic, and bimodal exploration, i.e., subjects were able to recover these relationships regardless of the modality used. Given the complexity of the measurement space, this is not a trivial process, as demonstrated by the difficulty of performing the same task using state-of-the-art computational algorithms (Paper 1, Figures 6 and 7).

Despite sharing common dimensions and ordinal relationships, there were two clear differences amongst modality-specific perceptual spaces. First, the *relative weights* of shape and texture dimensions differed: on average, shape dominated texture when objects were seen, while shape and texture were roughly evenly-weighted when objects were either touched, or both seen and touched (Paper 1, Figure 4, left). This finding agrees with the notion that vision is specialized for the extraction of object macrogeometry (Lederman et al., 1996). Interestingly, the same pattern of weights was observed in the categorization task (Paper 1, Figure 6), indicating that a relationship between similarity and categorization exists not only for stimuli perceived visually, but also for those perceived haptically and visuohaptically. Second, compared to the visual condition, we observed larger individual differences in similarity weights used in the haptic condition and even greater differences in the visuohaptic condition (Paper 1, Figure 4, right).[1]

A final result is particularly relevant vis-à-vis the question of multimodal integration: fitting similarity data from all three modality conditions using a *single* map with subject-specific weights was as good as fitting data from each modality condition with three separate, modality-specific maps. This implies that, for this data set, differences across modalities can be accounted for by simply linearly scaling the dimensions of a common map. Furthermore, when both similarity and categorization weights were averaged according to modality, bimodal weights turned out to be values between unimodal weights (Paper 1, Figures 4 and 6), indicating that a weighted average of unimodal weights may used in the multimodal condition.

## 3.2 Multidimensional scaling analysis of haptic exploratory procedures

This paper (Cooke et al., 2007b, Paper 2) took a closer look at perceptual similarities in the haptic modality and, in particular, the role of hand movements or "exploratory procedures" (EPs) in shaping similarity space. Lederman and Klatzky (1987, 1993) classified EPs into six types (lateral motion, pressure, static contact, holding, enclosure, and contour following) and showed that the ease of extracting object properties varies according to the EP used. For instance, texture is extracted best by lateral motion, a back-and-forth rubbing motion of the fingers over a surface, whereas this EP provides little or no information about object shape. In contrast, gripping objects in the hand allows for fast extraction of global shape and texture, whereas detailed shape information is harder to perceive by gripping. Following an object's contour, though time-consuming, allows for extraction of both texture and exact shape. A natural question, therefore, is whether

[1] Individual variability in haptic weights was also observed in in Papers 2 and 3. This issue is further discussed in Paper 1, p.5, Paper 2, p.13-16, and Paper 3, p.18.

these differences in extraction capabilities affect perceptual similarities. To investigate this, we had subjects provide similarity ratings after exploring the objects using one of four EPs (contour-following, gripping, tip-touching, or lateral motion on the objects' centres). In addition, to test whether perceptual similarities were stable over time, subjects repeated the experiment several months later.

The results showed that the specific type of hand movement used to explore the objects does indeed affect perceptual similarity. Both the number of dimensions needed to represent similarity data, as well as the specific stimulus configurations in perceptual space, varied as a function of hand movement. One expected result was that the dimensions used to judge similarity are critically dependent upon the dimensions which can physically be extracted using a specific hand movement. Specifically, lateral motion on the objects' centres, which does not provide any information about global shape changes, yielded one-dimensional perceptual representations in which the single dimension corresponded to texture. In contrast, two perceptual dimensions, shape and texture, were needed to explain similarity data when subjects gripped the objects, followed their contours, or touched their tips. For these EPs, we also observed a clear difference in the spatial layout of the configurations themselves: shape differences were roughly perceptually equidistant when subjects explored object tips or followed object contours, but when objects were gripped, objects with smoother macrogeometry were grouped apart from those with sharp angles in the macrogeometry (Paper 2, Figure 3).

In addition, the perceptual importance of shape and texture varied markedly from individual to individual when both properties could be extracted. Thus, the individual variation observed in the other two studies, in which only contour-following was used, was *not* specifically due to the use of the contour-following EP. Interestingly, weights used by the same individuals were significantly correlated (Paper 2, Figure 6, right), indicating that subjects may have been imposing an inherent preferred tradeoff value across all conditions. What could be the source of these biases? In the grip condition, photographs of subjects' grip positions allowed us to correlate shape/texture weight with the number of object tips contacted by subject-specific grip positions (Paper 2, Figures 8 and 9). However, for the other two EPs (contour-following and lateral motion on the objects' tips), the biases could not be explained. Further studies are required to test whether differences could be related to subtle, yet systematic differences in the way subjects perform the EPs, or to individual cognitive biases arising from individual preferences and/or experience.

When subjects returned to repeat the experiment several months later, the same patterns of weights were found, i.e., individual differences remained stable over time (Paper 2, Figures 6 and 7). Further work is needed to determine whether stability is due to an influence of memory (of the objects and/or task) or whether stability could be explained by the fact that both experiments involved the same inputs being processed by the same perceptual system, i.e., that subjects used "the world as an external memory store" (Simons, 1996; O'Regan and Noe, 2001).

The fact that haptic similarities *were* so variable across individuals is an important finding for the designers of haptic devices. The paper demonstrates how MDS techniques can be used to quantify individual differences and, as a result, makes it possible to calibrate systems to compensate for these differences. The paper concludes by presenting a similarity-based approach for validating haptic devices (Paper 2, Figure 10).

### 3.2.1 A similarity-based approach to haptic device validation

One variant of the similarity-based approach for validating haptic devices proposed in Paper 2 was demonstrated in a collaboration with ETH Zürich.[2] The target of the study was a haptic rendering algorithm for simulating tissue samples, to be used as part of a surgical training system. At this stage of the development process, the engineering goal was to have an algorithm capable of rendering a set of virtual objects which are perceived to vary only in terms of stiffness. This study tested two versions of such an algorithm. Users were presented with mixed pairs of virtual objects and real objects. The real objects consisted of rubber samples manufactured to have different physical stiffness values. Two types of virtual objects were used; each type was rendered with a different version of the rendering algorithm. The first version, referred to as "higher-fidelity," was expected to yield more realistic stimulation than the second version, referred to as "lower-fidelity." The setup was designed such that subjects would be unaware whether a given object was real or virtual.

MDS analysis revealed that when virtual objects were rendered using the lower-fidelity algorithm, users made similarity judgments based on stiffness differences, but *also* distinguished between real and virtual objects along an additional perceptual dimension. When the higher-fidelity algorithm was used, stiffness alone sufficed to account for similarity data. This result can be interpreted as "validating" the higher-fidelity algorithm, since it led to the desired perceptual relationships amongst the objects (i.e., no significant perceptual difference between real and virtual objects), and "invalidating" the lower-fidelity algorithm, since it led to an additional degree of perceptual separation between real and virtual objects.

This study provides proof-of-concept for a similarity-based approach to haptic device

[2]This collaboration resulting in the second-author publication (Leskovsky et al., 2006).

validation. There is a growing need for such methods given the increasing number of haptic devices being developed for teleoperation, training, assistance, and entertainment (for an overview, see Fisch et al. (2003)). Furthermore, the method can be applied to validate *multimodal* interfaces, such as visuohaptic surgical simulators, which require that multiple sources of information about the virtual environment be successfully integrated (Oviatt, 1999).

## 3.3 Object feature validation using visual and haptic similarity ratings

This paper (Cooke et al., 2006, Paper 3) compared similarity spaces generated by human haptic and visual perception against similarity spaces generated using state-of-the-art computational feature extraction algorithms. Our motivation for doing this was two-fold: to validate computational features by identifying algorithms which yield similarity spaces akin to those produced by humans and to identify computational mechanisms which may underlie feature extraction processes in the human haptic and visual systems. Nine algorithms were implemented: five operated on 2D images of the objects and four operated on 3D object models (*2D:* raw image subtraction, edge detection, Gabor jets, the Visual Difference Predictor (VDP), and the Structural Similarity measure (SSIM)); *3D:* raw 3D position, perimeter, curvature, and local shape histograms). In addition, the algorithms were run on a set of higher-resolution object data as well as a set of lower-resolution data in order to investigate the effect of scale variation on computed similarity space representations.

As found in Paper 1, human similarity spaces varied according to the modality used to explore the objects, with shape dominating for vision and properties being equally weighted for touch on average.[3] The structure of similarity spaces extracted from computational measures also varied according to the type of computational algorithm used, both in terms of interstimulus distances (Paper 3, Figures 6 and 7) and relative dimension weights (Paper 3, Figure 8). Based on these representations, the algorithms were divided into two groups: a group yielding texture-dominated representations (3D curvature and perimeter) and a group yielding shape-dominated representations (2D and 3D raw subtraction, VDP, SSIM, and 3D shape histograms).

[3] The similarity ratings experiments conducted for Paper 3 differed from the ones conducted for Paper 1 in a number of ways: in Paper 1, visual stimuli were real objects presented under natural lighting conditions, whereas in Paper 3, they were photographs presented in darkness on a computer screen; in Paper 1, haptic stimuli were mounted upright on a stand, whereas in Paper 3, they were laid flat on a table. Despite these differences, we found consistent patterns of property weights across modality conditions, including greater intrasubject variability in the haptic condition than in the visual condition.

The map generated from each computational measure was fit to each human subject map and mean fit errors were calculated, once for all subjects in the haptic condition and once for all subjects in the visual condition.[4] A computational measure was deemed to be "perceptually valid" relative to either vision or touch if the mean fit error was not significantly different from the error generated by fitting all human subject maps to one another.

When fit to visual data, the 3D shape, 2D and 3D subtraction, VDP, and SSIM measures met this criterion at the fine scale of object data, while Gabor jets, SSIM, and both subtraction measures met it at the coarse scale (Paper 3, Figures 9 and 10, top left). In general, shape-dominated measures provided good fits to human visual maps because they shared the common characteristic of strongly separating objects according to shape differences. However, the algorithms had more difficulty ordering the stimuli according to texture, which humans did very well. The main effect of the scale manipulation was that algorithms had more difficulty recovering texture from the coarser data set. This affected texture-dominated measures (curvature and perimeter) most adversely. An interesting extension of this research would be to compare similarity data gathered from human perception of downsampled object data against computed similarities.

When fit to human haptic data, no single computational measure met the criterion we had established for perceptual validity. This was due to the fact that the measures we implemented responded strongly to *either* shape or texture, while many haptic subjects relied on both properties to make similarity judgments. Good fits were obtained between the curvature measure and haptic subjects who were strongly texture-biased, and between VDP and 2D subtraction and a subject who was strongly shape-biased. Combinations of features could be used to model the results of haptic subjects who weighted shape and texture more evenly. Contrary to our expectations, we did not find that measures computed on 3D data provided better fits to haptic data than measures computed on 2D data, which could be taken as an indicator that 2D features suffice to support haptic object representations. However, studies using objects with greater 3D variation are needed to test whether this is indeed the case.

The paper concludes by presenting a two-stage approach to perceptual feature validation (Paper 3, Figure 11). In contrast to Paper 2, which proposed a similarity-based approach for validating haptic devices, i.e., systems which provide artificial *input* to human senses, this paper shows how the same ideas can be used to validate systems which perform artificial *perception* in real-world environments. In the past, computational object features have been evaluated based on computational efficiency or in terms of how well they can discriminate, recognize, or categorize objects. As an alternative criterion,

[4]For a complete description of the fitting procedure, see Paper 3, Section 2.6.

we propose that "valid" measures are those which are capable of mimicking human similarity judgments on a set of objects.

### 3.3.1 The *Perceptual Feature Toolbox*: a collection of methods for benchmarking psychophysical stimuli

The work presented in Paper 3 has recently generated interest amongst researchers who would like to use the features we implemented to automatically compute distances amongst their own stimuli sets. As pointed out in the Methods section, parameters manipulated in software to vary object properties may not correspond to the dimensions used by subjects to judge similarity. Furthermore, stimuli created using equidistant steps in a parameter defined in software cannot be assumed to be *perceptually* equidistant. Perceptual space can, of course, be characterized by MDS analysis of experimentally-measured similarity data, but this is time-consuming for a large stimulus set and may need to be repeated after any changes to the stimulus set. An alternative is to identify a computational measure which generates good fits to a sub-sample of perceptual similarities and then to process all remaining stimuli using that measure. Similarities can also be easily recomputed when the stimulus set is modified. To assist researchers in identifying potentially suitable measures, a *Perceptual Feature Toolbox* (PFT) for MATLAB has been developed, which currently implements 12 computational features.[5] It allows researchers to run multiple features detection algorithms simultaneously over 2D or 3D data sets, compute similarities, and perform MDS analyses to visualize corresponding similarity spaces. MDS maps generated by the algorithms can then be fit to those based on human similarity ratings, as done in Paper 3, to identify the best-fitting measure. Computational results can also be compared against other kinds of behavioural proximity data, such as neuronal firing rates, discriminability, naming times, or category decisions (see, for example, Laws et al. (2003); Kayaert et al. (2005)). Such investigations are central to understanding the connections between physical quantities available in real-world stimuli and the information extracted by human perception.

## 3.4 Outlook: Towards a better understanding of multimodal similarity

Taken together, the studies presented in this thesis demonstrate that perceptual similarity is not only defined by physical relationships between objects, but also by the specific

[5]This work was done in collaboration with S. Kannengiesser, F. Steinke, M. Siepmann, and C. Wallraven.

mechanisms used to extract object properties. One of the most compelling examples is that seeing and touching objects led to different perceptual weighting of object properties. The sensory modality used to extract object properties appeared to bias the dimensions of perceptual representation towards the feature it could most readily extract, i.e., shape in the case of vision and texture in the case of touch. The variability of property weights across individuals also changed according to modality: greater variability was observed when exploration was haptic or visuohaptic. Evidence was found that individual differences in haptic weights remain stable over time.

The existence of differences in visual and haptic similarity judgments raises the fundamental questions of how disparate haptic and visual information is merged within a given individual, and how social consensus is reached amongst individuals with different biases. With respect to the first question, results from Paper 1 suggested that vision and touch could feed into similar or even shared representations, with differences being accounted for by linearly rescaling the map's dimensions using a modality-specific factor. In the case of bimodal perception, we found evidence that the rescaling factor could be a weighted sum of unimodal rescaling factors. These findings demonstrate that spatial models can be used to characterize modality-based differences in perceptual similarity and suggest a simple mechanism for linking mental representations derived from unimodal and multimodal experience.

To further test the plausibility of linearly-weighted spatial representations for the case of visuohaptically-perceived objects, the work presented in this thesis could be extended in several ways. The following paragraphs outline studies on three topics: (1) tests of whether spatial models are capable of representing a broader range of stimuli explored under more natural conditions; (2) tests of specific models for computing dimension weights; and (3) a search for the neural underpinnings of multimodal spatial representations.

A first series of experiment should test whether linear spatial models can account for similarity data measured when objects are explored in a more natural setting, for instance freely and/or bimanually, and over longer periods of time. A broader range of stimulus sets also need to be tested, including:

- different viewpoints/rotations/sizes of the same objects;
- objects belonging to the same family, but created using different parameters in the space;
- objects with the same general structure, but with a larger number of changing geometrical properties, such as part size, type, number, and configuration;
- completely different classes of visuohaptic objects, including natural classes such as

shells (Vermeij, 1996).

Second, for dimension reweighting to be a plausible way of combining unimodal sources of information about objects, one needs to specify a scheme for determining the weights. One method is to use the modalities' relative statistical reliabilities, for instance how reliably each modality can be used to estimate the magnitude of a certain object property (Ernst and Bülthoff, 2004). Experiments are needed to test for correlation between reliability estimates and subject-specific dimension weights in similarity space. It would also be interesting to test whether higher group variability in haptic weights correlates with overall lower reliability of the haptic system for a given task.

Third, functional magnetic resonance imaging (fMRI) studies could be helpful in testing whether/how spatial representations of visuohaptic objects are implemented in the brain. A number of studies have already made significant progress in identifying areas involved in the convergence of visual and haptic object information (see James et al. (2006) for a review); here, we suggest how similarity-based approaches presented in this thesis could be applied to shed further light on this issue. An important first task would be to design a system which enables automatic, pseudorandom presentation and exploration of sufficiently large numbers of 3D visuohaptic stimuli in the fMRI scanner, without causing problematic motion artifacts. As a first study, brain areas activated during unimodal and multimodal similarity judgments could be localized and compared to identify areas where unimodal information converges. One could also test for an effect of property type on activation patterns. Second, similarity ratings could be performed on objects presented to two different modalities; for trials in which identical stimuli were presented to different modalities, neural adaptation patterns could be analyzed to identify brain areas for which the cross-modal stimuli are the "same" or "different" (Grill-Spector and Malach, 2001). Perceptual maps generated from behavioural and neural data could be compared and used to determine whether identical objects presented separately to different modalities are co-located or separated in these spaces.

These extensions would help to further our understanding of how our multiple sensory systems extract information about objects in the world and how this information is combined when we make similarity judgments. Having this knowledge not only satisfies our intellectual curiosity about ourselves, but it is also a basic requirement for designing technologies which are capable of efficiently delivering information through our sensory channels. As Oviatt (1999) concluded in her overview of human-computer interfaces: "The development of (robust multimodal systems) will not be achievable through intuition alone. Rather, it will depend on knowledge of the natural integration patterns that typify people's combined use of different input modes."

# Bibliography


Attneave, F. (1950). Dimensions of similarity. *American J Psychol*, 63:516–556.

Bergmann Tiest, W. and Kappers, A. (2006). Analysis of haptic perception of materials by multidimensional scaling and physical measurements of roughness and compressibility. *Acta Psychol (Amst)*, 121(1):1–20.

Bimler, D. and Paramei, G. (2006). Facial-expression affective attributes and their configural correlates: Components and categories. *The Spanish Journal of Psychology*, 9:19–31.

Bosten, J., Robinson, J., Jordan, G., and Mollon, J. (2005). Multidimensional scaling reveals a color dimension unique to 'color-deficient' observers. *Curr Biol*, 15(23):950–952.

Caclin, A., McAdams, S., Smith, B., and Winsberg, S. (2005). Acoustic correlates of timbre space dimensions: A confirmatory study using synthetic tones. *Acoustical Society of America Journal*, 118:471–482.

Cooke, T., Jäkel, F., Wallraven, C., and Bülthoff, H. (2007a). Multimodal similarity and categorization of novel, three-dimensional objects. *Neuropsychologia*, 45(3):484–495.

Cooke, T., Kannengiesser, S., Wallraven, C., and Bülthoff, H. (2006). Object feature validation using visual and haptic similarity ratings. *ACM Transactions on Applied Perception*, 3(3):239–261.

Cooke, T., Wallraven, C., and Bülthoff, H. (2007b). Multidimensional scaling analysis of haptic exploratory procedures. *submitted*, -:–.

Edelman, S. (1999). *Representation and recognition in vision*. MIT Press.

Ekman, G. (1954). Dimenions of color vision. *The Journal of Psychology*, 38:467–474.

Ernst, M. and Bülthoff, H. (2004). Merging the senses into a robust percept. *Trends Cogn Sci*, 8(4):162–169.

Fisch, A., Mavroidis, C., Melli-Huber, J., and Bar-Cohen, Y. (2003). *Biologically-Inspired Intelligent Robots*, chapter Chapter 4: Haptic Devices for Virtual Reality, Telepresence, and Human-Assistive Robotics, page 24. SPIE Press.

Forti, S. and Humphreys, G. (2005). Cross-modal visuo-tactile matching in a patient with a semantic disorder. *Neuropsychologia*, 43(11):1568–1579.

Garbin, C. (1984). Visual and haptic perception of three-dimensional solid forms. *Percept Psychophys*, 36:104–110.

Gillund, G. and Shiffrin, R. (1984). A retrieval model for both recognition and recall. *Psychol Rev*, 91(1):1–67.

Goldstone, R. and Son, J. (2005). *The Cambridge Handbook of Thinking and Reasoning*, chapter Similarity, pages 13–36. Cambridge University Press, New York.

Grill-Spector, K. and Malach, R. (2001). fMR-adaptation: a tool for studying the functional properties of human cortical neurons. *Acta Psychol*, 107(1-3):293–321.

Hahn, U. and Ramscar, M., editors (2001). *Similarity and categorization*. Oxford University Press.

Hollins, M., Faldowski, R., Rao, S., and Young, F. (1993). Perceptual dimensions of tactile surface texture: A multidimensional scaling analysis. *Percept Psychophys*, 54(6):687–705.

James, T., Humphrey, G., Gati, J., Servos, P., Menon, R., and Goodale, M. (2002). Haptic study of three-dimensional objects activates extrastriate visual areas. *Neuropsychologia*, 40:1706–1714.

James, T., James, K., Humphre, G., and Goodale, M. (2006). *Touch and Blindness: Psychology and Neuroscience*, chapter Do Visual and Tactile Object Representations Share the Same Neural Substrate?, pages 139–155. Lawrence Erlbaum Associates, London.

Jones, F., Roberts, K., and Holman, E. (1978). Similarity judgments and recognition memory for common spices. *Percept Psychophys*, 24:2–6.

Kayaert, G., Biederman, I., and Vogels, R. (2005). Representation of regular and irregular shapes in macaque inferotemporal cortex. *Cereb Cortex*, 15(9):1308–1321.

Klatzky, R., Lederman, S., and Reed, C. (1987). There's more to touch than meets the eye: The salience of object attributes for haptics with and without vision. *Experimental Psychology: General*, 116(4):356–369.

Kruschke, J. (1992). ALCOVE: an exemplar-based connectionist model of category learning. *Psychol Rev*, 99(1):22–44.

Kruskal, J. and Wish, M. (1978). *Multidimensional scaling*. Sage Publications.

Lakatos, S. and Marks, L. (1999). Haptic form perception: relative salience of local and global features. *Percept Psychophys*, 61(5):895–908.

Laws, K., Gale, T., and Leeson, V. (2003). The influence of surface and edge-based visual similarity on object recognition. *Brain Cogn*, 53:232–234.

Lederman, S. and Klatzky, R. (1987). Hand movements: a window into haptic object recognition. *Cognit Psychol*, 19(3):342–368.

Lederman, S. and Klatzky, R. (1993). Extracting object properties through haptic exploration. *Acta Psychol*, 84:29–40.

Lederman, S., Summers, C., and Klatzky, R. (1996). Cognitive salience of haptic object properties: role of modality-encoding bias. *Perception*, 25:983–998.

Lee, M. (2001). Determining the Dimensionality of Multidimensional Scaling Representations for Cognitive Modeling. *J Math Psychol*, 45(1):149–166.

Leskovsky, P., Cooke, T., Ernst, M., and Harders, M. (2006). Using multidimensional scaling to quantify the fidelity of haptic rendering of deformable objects. In *Proceedings of Eurohaptics 2006*.

Lim, J. and Lawless, H. (2005). Qualitative Differences of Divalent Salts: Multidimensional Scaling and Cluster Analysis. *Chem. Senses*, 30(9):719–726.

Newell, F., Ernst, M., Tjan, B., and Bülthoff, H. (2001). Viewpoint dependence in visual and haptic object recognition. *Psychol Sci*, 12(1):37–42.

Norman, J., Norman, H., Clayton, A., Lianekhammy, J., and Zielke, G. (2004). The visual and haptic perception of natural object shape. *Percept Psychophys*, 66(2):342–351.

Nosofsky, R. (1991). Tests of an exemplar model for relating perceptual classification and recognition memory. *J Exp Psychol Hum Percept Perform*, 17(1):3–27.

O'Regan, J. and Noe, A. (2001). A sensorimotor account of vision and visual consciousness. *Behav Brain Sci*, 24(5):939–973.

Oviatt, S. (1999). Ten myths of multimodal interaction. *Communications of the ACM*, 42:74–81.

Richardson, M. (1938). Multidimensional psychophysics. *Psychol Bull*, 35:659–660. (Abstract cited by Torgerson, 1952).

Shepard, R. (1962). The analysis of proximities: multidimensional scaling with an unknown distance function. i. *Psychometrika*, 27:125–140.

Shepard, R. (1963). Analysis of proximities as a technique for the study of information processing in man. *Human Factors*, 5:33–48.

Simons, D. (1996). In sight, out of mind: When object representations fail. *Psychol Sci*, 7(5):301–305.

Torgerson, W. (1952). Multidimensional scaling: I. theory and method. *Psychometrika*, 17:401–419.

Vermeij, G. (1996). *Privileged Hands: A Scientific Life*. W.H. Freeman & Co.

# Statement of Personal Contribution

**Paper 1: Cooke, T., Jäkel, F., Wallraven, C., and Bülthoff, H. (2007). Multimodal similarity and categorization of novel, three-dimensional objects. Neuropsychologia, 45(3), 484-495.**

I designed the experiments together with C. Wallraven, F. Jäkel and H.H. Bülthoff. I carried out or supervised all experiments, performed most data analysis, and wrote the paper. F. Jäkel also assisted with data analysis.

**Paper 2: Cooke, T., Wallraven, C., and Bülthoff, H. (2006). Multidimensional scaling analysis of haptic exploratory procedures. *(submitted)***

I designed the experiments together with C. Wallraven and H.H. Bülthoff. I carried out or supervised all experiments, performed all data analysis, and wrote the paper.

**Paper 3: Cooke, T., Kannengiesser, S., Wallraven, C., and Bülthoff, H. (2006). Object feature validation using visual and haptic similarity ratings. ACM Transactions on Applied Perception, 3(3), 239-261.**

I designed the experiments together with C. Wallraven and H.H. Bülthoff. I carried out or supervised all human experiments, implemented some of the computational algorithms, performed all data analysis, and wrote the paper. S. Kannengiesser implemented computational algorithms and performed the bulk of computational stimulus processing.

ELSEVIER

NEUROPSYCHOLOGIA

Neuropsychologia xxx (2006) xxx–xxx

www.elsevier.com/locate/neuropsychologia

# Multimodal similarity and categorization of novel, three-dimensional objects

Theresa Cooke *, Frank Jäkel, Christian Wallraven, Heinrich H. Bülthoff

*Max Planck Institute for Biological Cybernetics, Spemannstr. 38, 72076 Tübingen, Germany*

Received 25 August 2005; received in revised form 26 January 2006; accepted 20 February 2006

**Abstract**

Similarity has been proposed as a fundamental principle underlying mental object representations and capable of supporting cognitive-level tasks such as categorization. However, much of the research has considered connections between similarity and categorization for tasks performed using a single perceptual modality. Considering similarity and categorization within a multimodal context opens up a number of important questions: Are the similarities between objects the same when they are perceived using different modalities or using more than one modality at a time? Is similarity still able to explain categorization performance when objects are experienced multimodally? In this study, we addressed these questions by having subjects explore novel, 3D objects which varied parametrically in shape and texture using vision alone, touch alone, or touch and vision together. Subjects then performed a pair-wise similarity rating task and a free sorting categorization task. Multidimensional scaling (MDS) analysis of similarity data revealed that a single underlying perceptual map whose dimensions corresponded to shape and texture could explain visual, haptic, and bimodal similarity ratings. However, the relative dimension weights varied according to modality: shape dominated texture when objects were seen, whereas shape and texture were roughly equally important in the haptic and bimodal conditions. Some evidence was found for a multimodal connection between similarity and categorization: the probability of category membership increased with similarity while the probability of a category boundary being placed between two stimuli decreased with similarity. In addition, dimension weights varied according to modality in the same way for both tasks. The study also demonstrates the usefulness of 3D printing technology and MDS techniques in the study of visuohaptic object processing.
© 2006 Elsevier Ltd. All rights reserved.

*Keywords:* Similarity; Categorization; Multisensory; Haptics; MDS

The question of whether similarity can provide a theoretical basis for general categorization behaviour has been a source of heated debate in the field of cognitive psychology (Goldstone, 1994; Hahn & Ramscar, 2001). Critics of this idea have argued that the notion of similarity is vague and context-dependent, that it cannot explain category coherence, and that it does not account for the important role of theoretical knowledge in categorization decisions (Murphy & Medin, 1985). Nonetheless, similarity has served as the basis for a number of influential models of categorization (Medin & Schaffer, 1978; Nosofsky, 1992; Rosch & Mervis, 1975), which have been particularly successful in explaining classification of perceptual stimuli, including novel, 3D objects (Edelman, 1999). However, much of this work has been carried out within the context of perception involving a single modality, usually vision. Considering similarity and categorization within a multimodal context opens up a number of important questions: Are the similarities between objects the same when they are perceived using different modalities or by more than one modality at a time? Is similarity still able to explain categorization performance when objects are experienced multimodally?

In a preliminary study (Cooke, Steinke, Wallraven, & Bülthoff, 2005), we showed how multidimensional scaling (MDS) techniques can be used to quantify differences in perceptual similarities when objects are perceived using touch and vision. In that study, subjects saw or touched novel, 3D objects which varied parametrically in shape and texture and then rated the similarity between object pairs. Using similarity as a psychological distance measure, MDS was used to visualize stimuli as points in multidimensional perceptual spaces, as, for example, in Shepard and Cermak (1973), Garbin (1988), and Hollins, Faldowski, Rao, and Young (1993). We found that the relative importance of shape and texture in these perceptual spaces differed according to modality: shape alone sufficed to represent

* Corresponding author. Tel.: +49 7071 601 729; fax: +49 7071 601 616.
*E-mail address:* theresa.cooke@tuebingen.mpg.de (T. Cooke).

0028-3932/$ – see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neuropsychologia.2006.02.009

the stimuli when perceived visually, while shape and texture were both required when the stimuli were perceived haptically.

In the present study, we extend this line of research by adding a second task, free sorting categorization, and including a condition in which objects are simultaneously both seen and touched. The categorization task was included in order to test whether a connection between similarity and categorization could be established within this multimodal context. The bimodal condition was added in order to assess whether multimodal similarity and categorization would be dominated by one specific modality. At first glance, vision appears to be the most likely candidate. Vision is traditionally considered to be the "dominant" modality (Rock & Victor, 1964). Furthermore, object shape has been shown to play a special role in category formation (Landau & Leyton, 1999; Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976) and shape is thought to be a particularly salient feature for vision (Klatzky, Lederman, & Reed, 1987). On the other hand, recent studies have challenged the notion of ubiquitous visual capture and have argued in favour of weighted averaging models (Ernst & Bülthoff, 2004; Guest & Spence, 2003).

The results of this study show an effect of modality on the relative importance of object properties for both similarity and categorization tasks. In the bimodal condition, shape and texture were weighted roughly evenly for both tasks, rejecting the visual capture hypothesis. The probability of objects being grouped together in a category increased with similarity, while the probability of a category boundary being placed between two stimuli decreased with similarity. In addition, the relative importance of dimension weights for similarity and categorization tasks varied in the same way as a function of modality. The connection between similarity and categorization in the context of visuo-haptic object processing is discussed in light of these findings.

## 1. Methods

This section describes the stimulus set, the psychophysical tasks, and the analysis techniques used in this study.

### 1.1. Stimuli

A family of 25 novel, 3D objects (Fig. 1) was designed using the graphics package 3D Studio Max (3DS). The "base object" in the family (Fig. 1, object 1) consists of three parts connected to a centre sphere, specifying its macrogeometrical structure ("shape") and a displacement map applied to this 3D mesh, specifying its microgeometrical structure ("texture"). The remaining family members were created by parametrically varying the macrogeometrical and/or microgeometrical smoothness of the base object. Macrogeometrical smoothing was accomplished by applying a mesh relaxation operator which locally averages angles in the mesh in five linearly increasing steps (before application of the texture displacement). Microgeometrical smoothing was performed by linearly decreasing the amount of mesh displacement allowed by the application of the texture map in five steps. It is important to understand that the specific values of these parameters are only meaningful within 3DS. In addition, one cannot assume that equidistant changes in a software parameter yield perceptually equidistant changes in object properties.

Once an object is created in 3DS, it can either be rendered into a 2D image or printed into a solid 3D model. Printing is performed by a rapid prototyping machine (Dimension 3D Printer, Stratasys, Minneapolis, USA). The manufacturing process involves a head which deposits filaments of heated plastic such that the model is built up layer by layer. The final result is a hard, white, opaque, plastic model. In our case, models weighed about 40 g each and measured $9.0 \pm 0.1$ cm wide, $8.3 \pm 0.2$ cm high, and $3.7 \pm 0.1$ cm deep. It took 2–4 h to print each object. The same set of 3D models was used in all experiments described below.

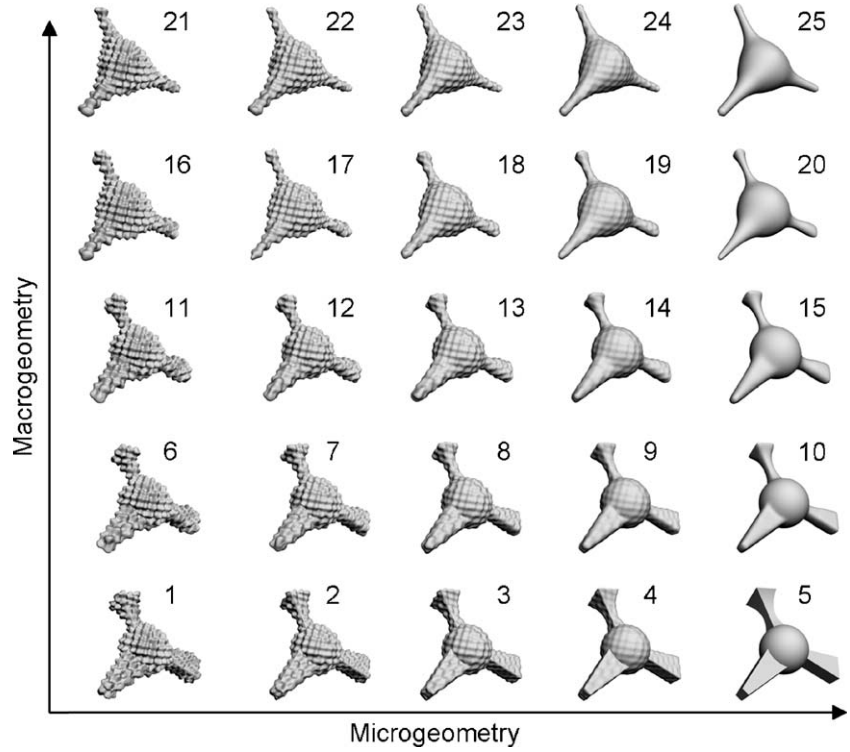


Fig. 1. Stimuli: novel, 3D objects ordered according to shape (macrogeometry) and texture (microgeometry).

Fig. 2. Experimental setup: top view of subject participating in the visuohaptic exploration condition. She follows the object's contour with her right index finger, repeats this for a second object, and then rates their similarity. The experimenter (shown on the right) records her response.

### 1.2. Similarity rating task

Thirty naive subjects (16 men and 14 women) were paid €8 per hour to participate in the experiments. The first task was to rate the similarities between pairs of objects on a scale between 1 (low similarity) and 7 (high similarity). Ten subjects explored the objects visually only, 10 subjects explored the objects haptically only, and 10 subjects explored the objects using both modalities simultaneously. All subjects in the haptic and bimodal conditions were right-handed. The same experimental setup was used in all conditions (Fig. 2). Subjects used a chin rest placed 30 cm away from the stand on which the objects were presented; the height of the chin rest was set such that the centre of the object was aligned with the line of sight. An opaque curtain hung between the subjects and the stand and could be slid back and forth along a rod to hide or reveal the objects. A piece of black sheet metal (30 cm × 30 cm) was mounted on the back of the stand so that when the curtain was open, subjects saw the object on a black background. A set of grooves and a section of rubber tubing on the mount piece ensured that the objects were securely held in place in exactly the same upright position on every trial.

In the visual condition, the experimenter placed the first object on the stand, slid the curtain over to reveal the object, waited for 3–5 s, slid the curtain back to cover the object, replaced the first object by a second object, uncovered the second object, and waited for the subject's response. In the haptic condition, the curtain was left in place while subjects explored the objects. Subjects were instructed to follow the contour of the objects with their right index finger. The contour-following procedure was selected because it has been shown to allow haptic extraction of both global shape and local texture properties (Lederman & Klatzky, 1993). Before the experiment, subjects practiced the procedure until they could trace the contour comfortably in 5 s or less. Subjects in the bimodal condition performed the same contour-following procedure while viewing the objects. In the visual condition, the presentation time of the first stimulus was set by the experimenter, whereas in the haptic and bimodal conditions it depended to a certain degree upon the subject's exploratory movements. The presentation time of the second stimulus was determined by the timing of the subject's response in all three conditions. In general, we strove to maintain presentation times of 3–5 s for all conditions.

The experiment consisted of 3 blocks of 325 randomized trials (each object was compared once with itself and once with every other object resulting in $25 + (25 \times 24)/2 = 325$ trials) and the order of appearance of stimuli was randomized over blocks. Each trial took about 20–30 s and the similarity ratings experiment ran for approximately 2 h per day for five consecutive days. The experiment began with a number of practice trials to help subjects become accustomed to the task. After the practice session, subjects were asked to write down their criteria for each value on the rating scale (e.g., "I say 7 whenever the objects are exactly the same"). On each of the subsequent days of the experiment, subjects were asked to read what they had written to ensure consistency over the course of the experiment.

### 1.3. Debriefing questionnaire

Immediately after the similarity ratings experiment, subjects filled out a form in which they were asked to describe the objects ("How do the objects look?" in the visual condition, "How do the objects feel?" in the haptic condition, and both questions in the bimodal condition), to explain how they had performed the similarity judgments, and to describe how they would group the stimuli into categories.

### 1.4. Free sorting categorization task

After having filled out the questionnaire, subjects performed a free sorting categorization task. We chose a free sorting task because of its relative simplicity and the ecological relevance of spontaneous categorization, also referred to as category construction (Milton & Wills, 2004). We designed the task to make it as close to the similarity task as possible. Using the same setup, stimuli were shown one at a time in random order and subjects explored the stimuli using the same exploratory procedure which they had used before. They were asked to assign a category number to each object, using the groups they had described in their questionnaire responses. The stimuli were repeatedly cycled through until the subject assigned the same category number to each object twice in a row.

### 1.5. Analysis techniques

#### 1.5.1. Analysis of similarity data

A multidimensional scaling technique was used to analyze the similarity data. MDS techniques take pair-wise proximities data for a set of objects (human similarity ratings in this case) and return the coordinates of the objects in a multidimensional space which best explains the proximity data. We used the individual differences weighted Euclidean distance model implemented as part of the ALSCAL MDS package in SPSS (Carroll & Chang, 1970; Young & Harris, 2003), with proximity data considered to be ordinal measurements (i.e., non-metric) and untying of tied proximities allowed. This particular technique was chosen because it allows for comparison of individual subject data and because the dimensions of the resulting spaces are uniquely specified, allowing for clearer interpretation (Borg & Groenen, 1997; Cox & Cox, 2001). The algorithm takes as input a set of individual subject similarity data and, for a specified dimensionality, returns a single underlying stimulus configuration together with a set of subject-specific weights. The weights specify how the underlying configuration should be scaled along each dimension to best fit each subject's similarity data. In addition, the SPSS implementation provides a goodness-of-fit measure, Young's S1 Stress, which is the normalized difference between the fitted distances and the observed proximities.

It is important to note that the weighted individual differences MDS model we used carries with it the following assumptions: (1) that the appropriate metric for the psychological similarity space is Euclidean[1] and (2) that each set of individual subject data included in the analysis can be modeled by linear stretching of the centroid configuration, as specified by the individual subject weights. If these assumptions hold true, one expects low stress values for the overall MDS solution. Although establishing a threshold for acceptable values of stress is notoriously controversial, Monte Carlo studies suggest that stress values below 0.2 are indicative of an output configuration which provides a good fit to the similarity data (Cox & Cox, 2001).

[1] There is no general consensus on the most appropriate general psychological similarity metric for haptically perceived stimuli. We were aware of the possibility that subjects' psychological metric could be non-Euclidean and tried to fit the similarity data using a city-block metric (Garner, 1974). We did not find a significant decrease in fit error compared to using a Euclidean metric and thus felt that the more intuitive Euclidean approach was preferable, especially given that our stimulus dimensions may not be strictly separable.

*1.5.2. Analysis of categorization data*

In the free sorting task, our observations consisted of: (1) the category membership which the subject assigned to each stimulus, (2) the total number of categories created by the subject, and (3) the number of repetitions of the free sorting task that was required before the subject provided the same categorization twice in a row. In addition to these raw measures, we calculated a measure of the relative importance of texture as compared to shape in the categorization task. The measure we chose relies on the assumption that subjects perceived the shape-based and texture-based adjacencies in the stimulus map, which indeed turned out to be the case (see Section 3). Subjects' free sorting categories were superimposed on the stimulus map (Fig. 1). When this is done, the boundaries between categories cross a certain set of adjacencies in the map. Each boundary separates two neighbours either on the basis of a difference in shape or on the basis of a difference in texture. We used the proportion of separations based on differences in texture as our measure of the relative importance of texture compared to shape for categorization.

*1.5.3. Analysis of the relationship between similarity and categorization*

We computed two basic measures of correlation between similarity and categorization data. First, for each value on the similarity scale (1–7), we created groups of stimulus pairs which had received that particular rating. We then divided each group into two subgroups according to whether subjects had placed the two objects into the same category or into different categories. For the second measure, we selected those pairs of stimuli which are neighbours along either the shape or texture dimension and computed the probability of subjects setting a category boundary between these neighbours as a function of their perceptual similarity.

*1.5.4. Analysis of subject questionnaires*

Questionnaires were read and scored independently by three judges (two authors and one external judge). We evaluated which object properties subjects mentioned when (1) describing the object, (2) describing how they made similarity judgments, and (3) describing how they would categorize the objects. For each of these points, we evaluated whether the subject had made at least one reference to the attributes shape, texture, material properties, colour, or temperature. We additionally distinguished between references to global shape and part shape. A reference to part shape was defined as the explicit mention of one of the objects' parts ("leg", "centre ball"), while a reference to global shape was defined as the use of holistic shape term such as "star-like" or a reference to part configuration, such as "three ends which extend from a ball-shaped center". When subjects mentioned a configuration of parts, we counted this as both a reference to global and part shape.

## 2. Results and discussion

This section presents the results of the similarity rating and free sorting categorization tasks, the debriefing questionnaires, as well as results obtained by comparing data from the similarity and categorization tasks.

### *2.1. Results of similarity rating task*

*2.1.1. Number of underlying perceptual representations*

A critical question for this study is whether allowing for separate, *modality-specific* stimulus representations provides a better explanation of the data than a single, multimodal representation which combines information from touch and vision. Recall that the weighted individual difference MDS model makes the following assumption: the data under consideration can be explained by a single underlying map and a set of dimension weights which are individually adjusted for each subject (see Section 2). The better this assumption holds true for a given data set, the lower the MDS stress will be. Here, we use this

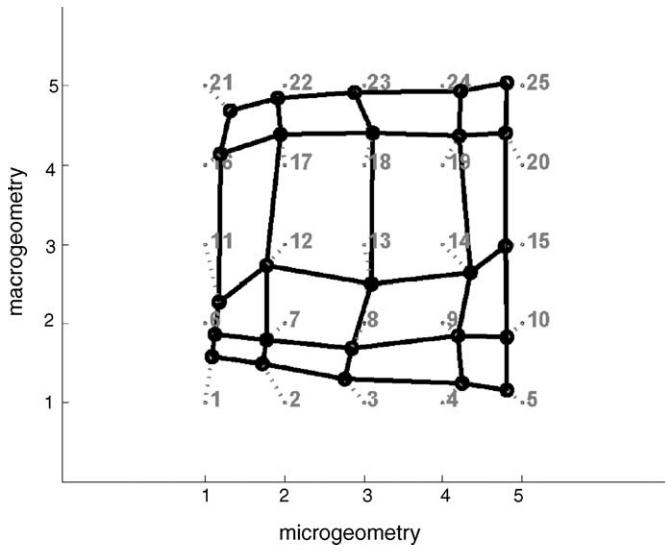


Fig. 3. Perceptual stimulus map: map derived by MDS using similarity data from all modality conditions.

idea to evaluate whether (1) a single, multimodal representation with individual weights or (2) three, modality-specific representations provide better fits to our data by comparing stress values from (1) a global MDS computed over all similarity data and (2) three separate MDS solutions, one for each set of modality-specific similarity data. The stress for a two-dimensional MDS solution using grouped similarity data from all modality conditions was 0.167. Using modality-specific similarity data sets, stress values were 0.157 (visual), 0.168 (haptic), and 0.160 (bimodal). Since stress values were below 0.2 in all cases, all two-dimensional solutions provided good fits to the respective sets of similarity data (see Section 2). The similarity amongst these stress values indicates that positing a single, multimodal representation of the stimuli provides an equally good explanation for our data as positing three separate representations.

*2.1.2. Dimensionality of the underlying perceptual representation*

*Two* perceptual dimensions were recovered in all MDS analyses, i.e., increasing the dimensionality of the spaces did not produce substantial decreases in stress. This shows that subjects recovered a two-dimensional representation using visual, haptic, and visuohaptic exploration. Subjects' descriptions of how they made similarity judgments (Fig. 7) confirmed that they perceived the two dimensions as "shape" and "texture". Although several additional object properties (such as material, colour, and temperature) were mentioned when describing the objects, only shape and texture properties were mentioned when subjects explained how they made similarity judgments.

*2.1.3. Topology of the underlying perceptual representation*

The stimulus configuration resulting from the two-dimensional MDS solution computed over all similarity data is shown in Fig. 3. This topology was qualitatively very similar to the topologies recovered with modality-specific MDS analyses. Ordinal adjacency relationships between the stimuli were

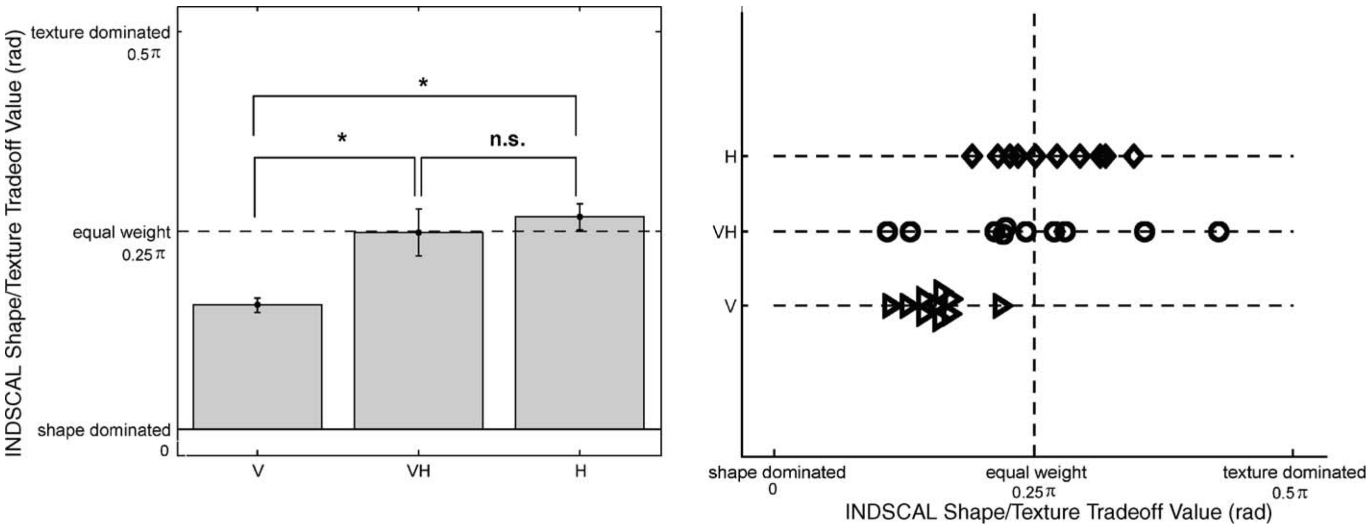


Fig. 4. Relative weight of shape and texture in similarity judgments. Population mean (left) and individual subject data (right). Error bars represent standard error. The relative weight is calculated as the angle that the weight vector in MDS subject space makes relative to the shape axis. Overlapping individual data points are vertically shifted for better visualization. V, visual; VH, visuohaptic; H, haptic; *, significant difference; n.s., not significant.

preserved, demonstrating that subjects were able to recover the *ordering* of shape and texture variations in the stimuli, i.e., the adjacency relationships between neighbours in the map. This was also true of maps derived from separate analyses of individual subject similarities: there was perfect recovery of the ordinal relationships in 22/30 cases and recovery with one exception in 5/30 cases. As we have previously discussed (Cooke et al., 2005), this is a non-trivial task given the high dimensionality of the measurement spaces involved. To fully appreciate this, one need only consider the difficulty of recovering these relationships using computational methods, which we demonstrated in the aforementioned study.

It is also of note that the perceptual distances between stimuli, reflected in the MDS map, deviate from the distances between stimuli in the space defined by the manipulation of software parameters, in which stimuli lie on a rectangular grid (see Section 2). This provides an important reminder that the *perceptual* distances between stimuli cannot be assumed to vary linearly with the distances defined by manipulating parameters in a software program, an issue often neglected in the growing number of studies involving parametrically controlled stimuli (e.g., stimuli created using morphing techniques).

### *2.1.4. Modality-based analysis of dimension weights*

The subject weights provided by the individually weighted MDS can be visualized as vectors connecting each subject to the origin of a two-dimensional weight space. Because the sum of squared weights is constrained to a constant value, we simply calculated the angle between each subject vector and the shape axis of this weight space as a single variable which represents the relative weighting of shape and texture dimensions (Fig. 4). The mean weights for the 10 subjects in the visual condition ($M = 0.16\pi$ rad, S.E. $= 0.01\pi$ rad) and for the 10 subjects in the haptic condition ($M = 0.27\pi$ rad, S.E. $= 0.02\pi$ rad) were found to be significantly different ($t[14] = 6.0$, $p < 0.001$). The mean weights for the 10 subjects in the visual condition and the 10 subjects in the bimodal condition ($M = 0.25\pi$ rad, S.E. $= 0.03\pi$ rad) were also found to be significantly different ($t[10.6] = 2.9$, $p = 0.01$) using a two-sample, two-tailed *t*-test for independent samples with unequal variances. No significant difference in mean weights was found for the haptic and bimodal conditions ($t[13.8] = 0.57$, $p > 0.1$). These results show a clear effect of modality on the relative weighting of stimulus dimensions for similarity judgments. Note that the mean tradeoff value in the bimodal condition ($M = 0.25\pi$ rad) lay between the values obtained in the unimodal conditions ($0.16\pi$ rad for vision and $0.27\pi$ rad for touch), although statistically speaking, there was no difference between the bimodal and haptic weights.

Next, we tested the hypothesis that the weights came from a distribution with a mean of $0.25\pi$ rad, representing equal importance for shape and texture properties in the subjects' similarity judgments. A single-sample *t*-test rejected the hypothesis for the visual condition ($t[9] = 10$, $p < 0.001$), but not in the haptic ($t[9] = 1.0$, $p > 0.1$) or bimodal conditions ($t[9] = 0.1$, $p > 0.1$). Taken together, these results show that shape dominated texture when similarity judgments were performed visually, while shape and texture were equally important when similarity ratings were performed either haptically or bimodally.

The individual data (Fig. 4, right) shows that all visual subjects were indeed shape-dominated, while haptic subjects were quite evenly distributed around $0.25\pi$ rad (equal shape and texture weight). In the bimodal condition, 6/10 subjects weighted shape and texture quite evenly. However, two subjects weighted shape much more heavily than texture, and two subjects weighted texture much more heavily than shape. One explanation for the wide range of weights observed overall in the bimodal condition is that the involvement of two modalities requires an integration of information from touch and vision to occur and that this integration process varies across subjects. For example, the reliability of shape/texture estimates may vary from subject to subject (e.g., as a function of relative expertise or familiarity with a given modality). Another possibility is that

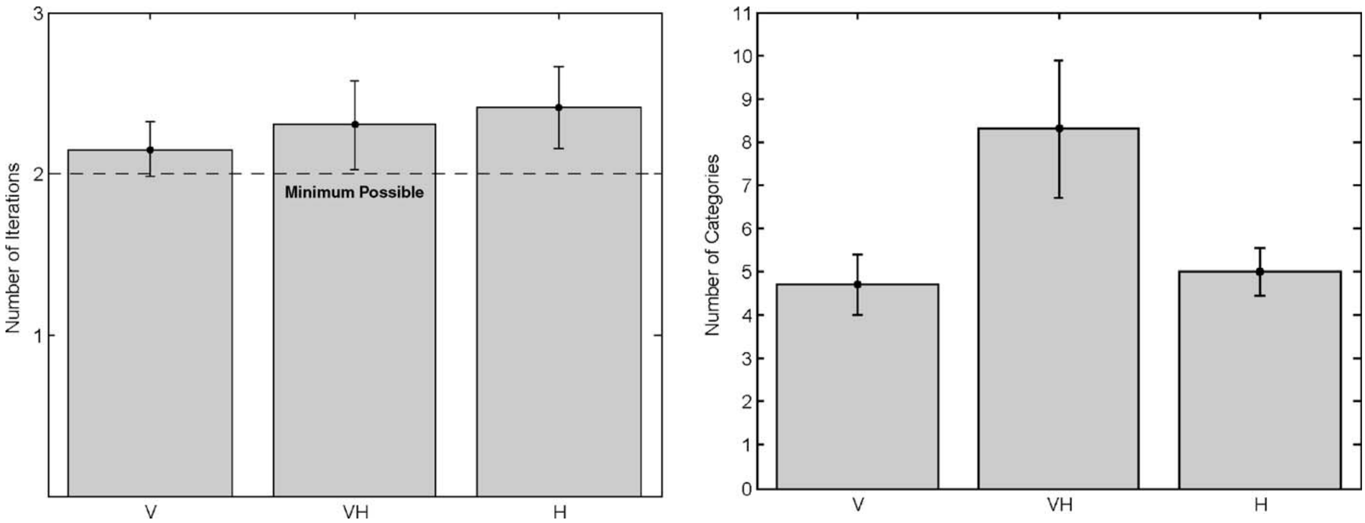


Fig. 5. Performance of free sorting categorization. Number of repetitions of free sorting task needed until all stimuli were categorized the same way twice in a row (left) and number of categories created by subjects in each modality condition (right). Error bars represent standard error.

the integration process leads to a conflict between shape/texture weights dictated by the haptic and visual systems and that subjects attempt to resolve the conflict by making a conscious decision about the relative feature weights.

### 2.2. *Results of free sorting categorization task*

#### 2.2.1. *General task measures*

Fig. 5 (left) shows that subjects in all modality conditions performed the free sorting task with a very high degree of consistency, rarely requiring more than the minimum number of two iterations through the stimuli in order to provide the same categorization twice (visual condition: $M = 2.1$ iterations, S.E. = 0.2 iterations; bimodal condition: $M = 2.3$ iterations, S.E. = 0.3 iterations; haptic condition: $M = 2.4$ iterations, S.E. = 0.3 iterations). As shown in Fig. 5 (right), there was a noticeable effect of modality on the number of categories created by subjects: subjects in the bimodal condition created more categories ($M = 8.3$ groups, S.E. = 1.6 groups) than subjects in unimodal conditions (visual: $M = 4.7$ groups, S.E. = 0.7 groups; haptic: $M = 5$ groups, S.E. = 0.5 groups). This could be due to a combinatorial effect of having redundant or conflicting information available from the two modalities. We also observed a tendency for subjects to use dimension-based rules to construct their categories: half of our 30 subjects appeared to use rules along a single dimension, while 8 subjects combined rules along both shape and texture axes; categories constructed by the remaining 7 subjects could not be well-described using combinations of unidimensional rules.

#### 2.2.2. *Modality-based analysis of dimension weights*

As a measure of the relative importance of texture compared to shape, we calculated the proportion of category boundaries

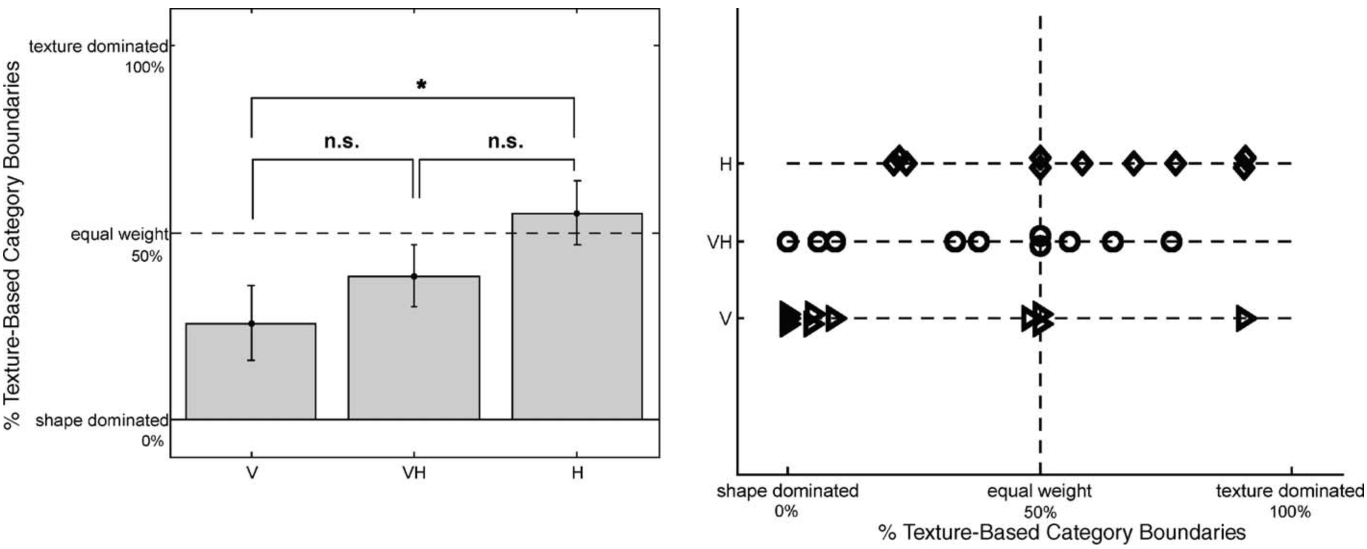


Fig. 6. Relative weight of shape and texture in categorization. Population mean (left) and individual subject data (right). Error bars represent standard error. The relative weight is calculated as the percentage of the total number of category boundaries which separated stimuli on the basis of texture differences. Overlapping individual data points are vertically shifted for better visualization. V, visual; VH, visuohaptic; H, haptic; *, significant difference; n.s., not significant.

based on texture differences between stimuli, shown in Fig. 6 (see Section 2). The mean proportion of texture-based boundaries for the 10 subjects in the visual condition ($M = 26\%$, S.E. = 10%) and for the 10 subjects in the haptic condition ($M = 55\%$, S.E. = 8%) was found to be significantly different ($t[17.5] = 2.2$, $p = 0.04$; two-sample, two-tailed $t$-test for independent samples with unequal variances), indicating that subjects in the visual condition relied more on shape in their categorization decisions than subjects in the haptic condition. No significant difference was found between the mean proportion of texture-based boundaries used in the visual condition and in the bimodal condition ($M = 38\%$, S.E. = 8%) nor was a significant difference found between the bimodal and haptic conditions (visual–bimodal: $t[17.3] = 0.98$, $p > 0.1$; bimodal–haptic: $t[18] = 1.4$, $p > 0.1$). Thus, there was no statistical evidence in favour of more "visual-like" or more "haptic-like" use of stimulus dimensions in bimodal categorization, however the mean weight for bimodal categorization (38% texture-based boundaries) lay between the values obtained in the unimodal conditions (26% for vision and 55% for touch).

Next, we tested the hypothesis that weights came from a distribution with a mean of 50%, i.e., that subjects based category boundaries equally often on shape and texture differences. A single-sample $t$-test rejected this hypothesis for the visual condition ($t[9] = 2$, $p = 0.04$), but not for the haptic ($t[9] = 0.6$, $p > 0.1$) or bimodal conditions ($t[9] = 1.4$, $p > 0.1$). Strikingly, this is the same pattern which we observed for the similarity tradeoff values: shape dominated texture for visual categorization while shape and texture were roughly evenly weighted for both haptic and bimodal categorization. There was a large amount of individual variation in the relative importance of shape/texture for categorization across all modalities (Fig. 6). However, 7/10 haptic subjects weighted texture as heavily or more heavily than shape, 9/10 visual subjects weighted shape as or more heavily than texture, and 6/10 bimodal subjects exhibited fairly equal weighting of the two properties, a pattern which supports the outcome of the $t$-test. The remaining variation could be due to the measure we computed, to the design of the free sorting task, or to intrinsic modality effects. Further studies involving different categorization tasks are needed to disentangle these factors.

### 2.3. Results of subject questionnaires

Fig. 7 shows the frequency with which subjects mentioned various object features when describing the objects (left column) and when explaining how they performed similarity ratings (centre column) and free sorting (right column). When subjects were asked to *describe* the objects, the frequency with which they mentioned various object features depended on modality. Shape was mentioned by 10/10 subjects in the visual condition (V), 9/10 subjects in the visuohaptic condition (VH), and 3/10 subjects in the haptic condition (H); texture was mentioned by 5/10 subjects (V), 10/10 subjects (VH), and 9/10 subjects (H); material properties were mentioned by 1/10 subjects (V), 2/10 subjects (VH), and 5/10 subjects (H); colour was mentioned by 4/10 subjects (V), 2/10 subjects (VH), and 0/10 subjects (H).

Interestingly, although subjects described the objects using a variety of features, they only mentioned shape and texture when asked to explain how they had performed the similarity and categorization tasks. For haptic subjects, there was a particularly striking difference in that 10/10 subjects mentioned shape for the similarity task and 9/10 subjects mentioned shape

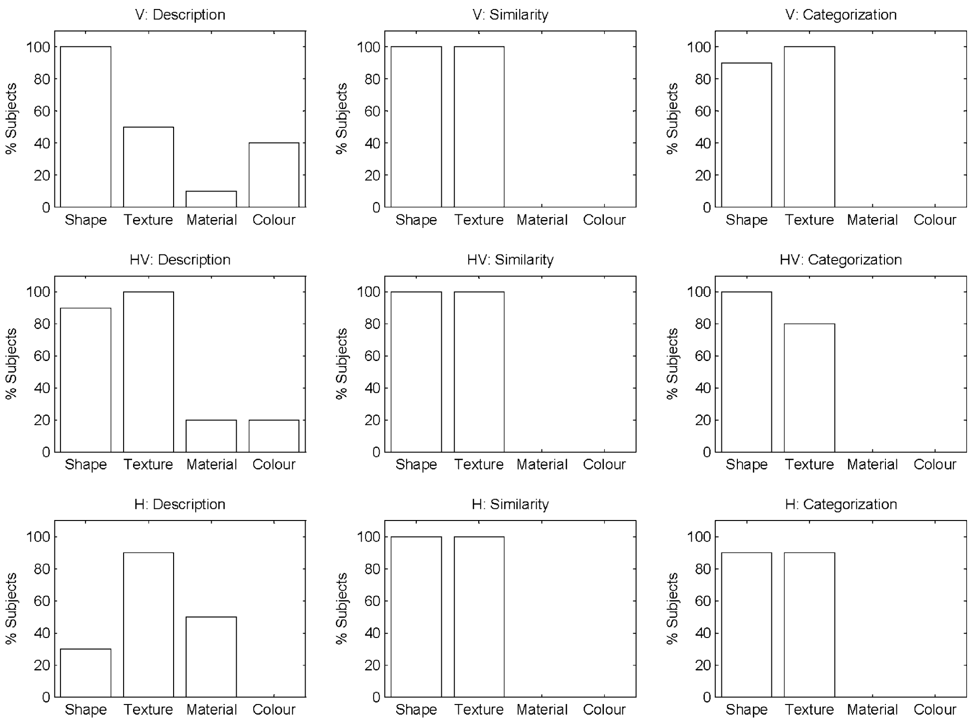


Fig. 7. Verbal mention of object properties for object descriptions, similarity judgments, and categorization. Shape responses include references to both global and part shape properties. Two haptic subjects also mentioned temperature when describing the objects. V, visual; VH, visuohaptic; H, haptic.

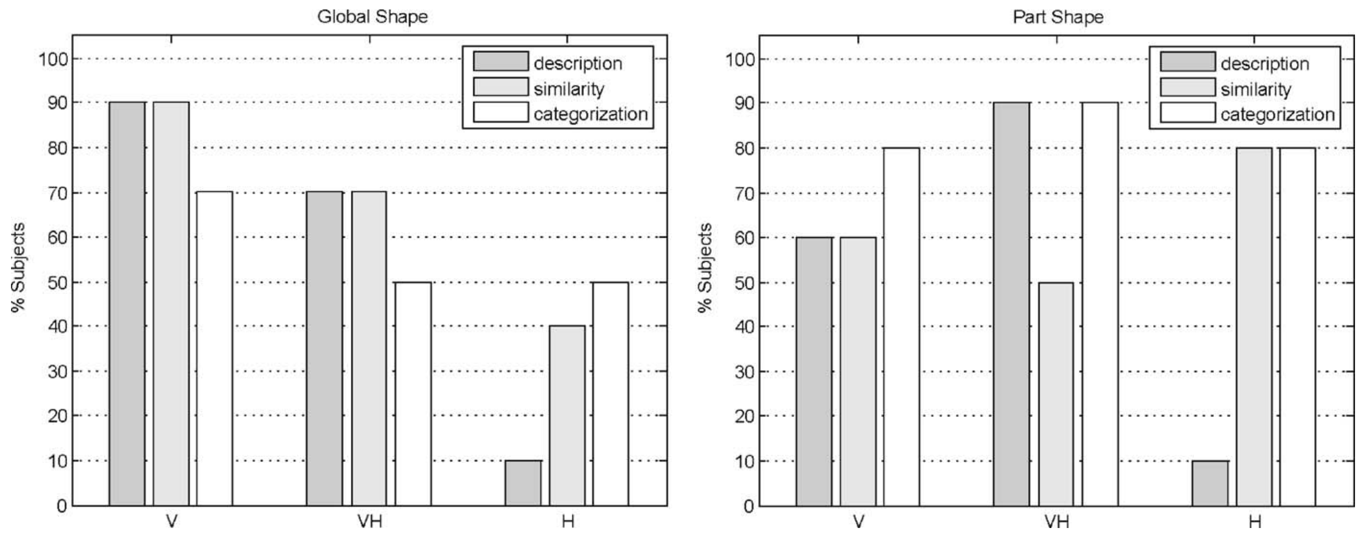


Fig. 8. Verbal mention of global shape (left) and part shape (right) for object descriptions, similarity judgments, and categorization. V, visual; VH, visuohaptic; H, haptic.

for the categorization task, even though only 3/10 mentioned it when describing the objects. One explanation could be that spontaneous description better reflects intrinsic modality feature biases (i.e., texture and material properties for haptics), whereas descriptions of features for similarity and categorization are more strongly influenced by the experimental task and stimulus set.

To help identify which aspects of the stimulus geometry play a role in the perceptual dimension "shape", we separated the subjects' references to shape into two categories: part shape and global shape (see Section 2). Global shape (Fig. 8, left) was mentioned more often by visual than by haptic subjects for all tasks (description: 9/10 (V) and 1/10 (H); similarity: 9/10 (V) and 4/10 (H); categorization: 7/10 (V) and 5/10 (H)). This could be explained by differing amounts of effort involved in extracting global shape in the two conditions: in particular, the contour-following procedure causes haptic extraction of global shape to be slow and memory-intensive. Allowing subjects to enclose the objects, for example, would have provided a quicker, albeit cruder estimate of global shape (Lederman & Klatzky, 1993) and might have increased the mention of global form. Interestingly, bimodal subjects' mention of global shape falls between the values for visual and haptic conditions (description: 7/10 (VH); similarity: 7/10 (VH); categorization: 5/10 (VH)). Part shape (Fig. 8, right) was consistently mentioned by subjects in all modality conditions when describing how categorization was performed (8/10 (V), 9/10 (VH), and 8/10 (H)), which is consistent with the recognized importance of part information in basic level categorization (Tversky & Hemenway, 1984) and with the importance of part information in haptic categorization by blind and sighted children (Morrongiello, Humphrey, Timney, Choi, & Rocca, 1994).

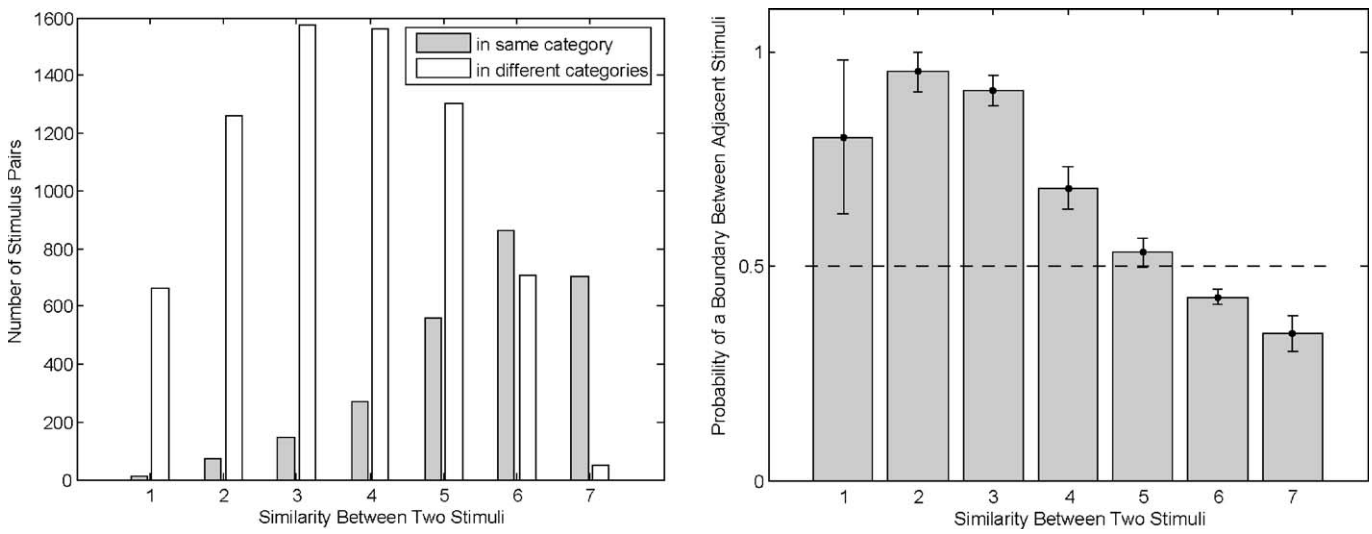


Fig. 9. Similarity and categorization. Left: histogram showing the number of stimulus pairs placed in the same category (grey) and the number of stimulus pairs placed in different categories (white) as a function of pair-wise similarity. Right: mean probability of subjects (all modality conditions, $N = 30$) placing a category boundary between *adjacent* stimulus pairs as a function of pair-wise similarity. Error bars represent standard error.

### 2.4. Results on the connection between similarity and categorization

Fig. 9 (left) shows the number of stimulus pairs which were placed into either the same or different categories, sorted by similarity. The highest frequency of stimulus pairs being placed in different categories occurs for similarity ratings of 3 and 4, while the highest frequency of stimulus pairs being placed in the same category occurs for similarity ratings of 6. Note that the decrease in same-category occurrences for a similarity value of 7 and the decrease in different-category occurrences for similarity values of 1 and 2 are due to the relative infrequency of these similarity values; the *relative proportion* of different-category to same-category occurrences is indeed a monotonically decreasing function of similarity.

Fig. 9 (right) shows the mean probability of subjects setting a category boundary between shape or texture neighbours in the stimulus map as a function of their perceptual similarity. Note that the fact that subjects consistently recover ordinal shape and texture structure in the stimulus set justifies the use of these adjacency relationships in the calculation of this measure. Here, we observe that the probability of subjects violating such an adjacency relationship decreases monotonically as a function of similarity. Again, the large amount of variance in the measure at a similarity value of 1 is due to the fact that adjacent stimulus pairs were rarely rated with a similarity of 1.

A further connection between similarity and categorization is that modality had the same effect on the relative weight of stimulus dimensions in both tasks. Shape was weighted more heavily than texture when similarity and categorization were performed visually, whereas shape and texture were weighted roughly evenly when the tasks were performed either haptically or bimodally (Figs. 4 and 6).

## 3. General discussion

### 3.1. Modality-dependent weighting of stimulus dimensions

We found that the relative importance of shape and texture for judging similarities between objects and for creating categories of objects varied systematically according to modality. For both similarity and categorization, shape dominated texture in the visual condition, while texture and shape were evenly weighted in the haptic condition. In the bimodal condition, we found texture and shape to be weighted evenly on average for both our similarity and categorization tasks. This finding is in agreement with Klatzky et al. (1987), in which subjects were instructed to sort wafers varying in shape, texture, size, and hardness based on their similarity. Substance dimensions (hardness and texture) were most salient after haptic exploration while saliency was evenly distributed across dimensions when subjects used both touch and vision. When subjects were explicitly instructed to use visual imagery to compare the objects after haptic exploration, shape became overwhelmingly dominant.

The results of this study replicate our previous results on unimodal shape and texture weights in similarity judgments (Cooke et al., 2005); together, these studies provide clear evidence that the perceptual modality used to interact with objects has an effect on object representations. The fact that we obtained the same pattern of similarity-based weights as in our previous study despite differences in the experimental conditions (e.g., stimuli in the previous visual condition were 2D images presented on a computer monitor, with shorter presentation times; haptic stimuli lay flat on a table instead of upright) indicates that the weight pattern we obtained is robust against these variations. Interestingly, Lakatos and Marks (1999) reported that local shape initially played an important role relative to global shape in haptic similarity ratings of 3D objects but that the importance of local shape decreased when exploration time was increased. Variables such as exploratory procedure, exploration time, and viewpoint need to be systematically manipulated in order to characterize the sensitivity of modality weights to such factors.

### 3.2. Convergence of stimulus representations

Positing a single, multimodal stimulus representation with modality-dependent weights provided the same goodness-of-fit to our similarity data as three, modality-specific representations. This suggests that the modalities make use of similar (or even perhaps common) object representations for the purposes of judging similarity. The idea that object information coming from touch and vision converges or at least overlaps in a multimodal object representation agrees with evidence from a number of visuohaptic processing studies using brain imaging techniques (e.g., Amedi, Jacobson, Hendler, Malach, & Zohary, 2002; Amedi, Malach, Hendler, Peled, & Zohary, 2001; Forti & Humphreys, 2005; James et al., 2002; Pietrini et al., 2004), and psychophysics (e.g., Easton, Greene, & Srinivas, 1997; Easton, Srinivas, & Greene, 1997; Norman, Norman, Clayton, Lianekhammy, & Zielke, 2004; Reales & Ballesteros, 1999). Elucidating the computational principles which govern multimodal integration is an important area of current research (Ernst & Bülthoff, 2004). Early studies of visuohaptic integration proposed that vision simply dominated touch when both modalities were available (Rock & Victor, 1964). In this study, we did not find evidence for visual capture for similarity and categorization tasks. Instead, our results in the bimodal condition appear to be more compatible with weighted averaging models of multisensory integration. In one such model (Ernst & Banks, 2002), the bimodal estimates of stimulus properties are weighted by the reliability of the unimodal estimates. A variant of our experiment in which the reliability of unimodal estimates is manipulated in the bimodal condition (e.g., by having subjects wear gloves, blurring the visual stimulus, or showing different stimuli in haptic and visual conditions) could be used to test whether this model is capable of predicting integration effects for similarity judgments and category construction.

### 3.3. Connection between similarity and categorization in a multimodal setting

In this study, we were able to establish a connection between similarity and categorization: similarity was lower for pairs which subjects placed in different categories and higher for pairs which subjects placed in the same category. In addition, when we made use of the fact that subjects perceived nearest-neighbour adjacencies in the stimulus set (e.g., between two objects which differ only in terms of one step along either the shape or texture dimensions), the probability of crossing such an adjacency with a category boundary decreased as a function of the perceptual similarity between the objects. We also found that the relative weight of shape and texture varied with modality in the same way for our similarity and categorization tasks. One explanation for this could be that modality-specific biases affect both tasks (or the representations upon which they operate) in a uniform fashion. Modality-specific biases towards features can arise due to a number of factors, including the relative discriminability of features, the relative reliability of feature value estimates, directing of attention, past experience, and ecological validity (Ernst & Bülthoff, 2004; Guest & Spence, 2003; Lederman, Summers, & Klatzky, 1996). The effects of modality bias in determining the relative weights of features in object representations may co-exist or compete with the effects of top–down category learning (Goldstone & Steyvers, 2001; Nosofsky, 1986; Sigala & Logothetis, 2002).

Despite the connection we found between performance in similarity and categorization tasks, we were not able to demonstrate a strong relationship between the two. We hypothesized that a strong connection between similarity and categorization might enable us to predict subjects' free sorting categories using the clusters of stimuli in their individual similarity-based stimulus spaces. However, this proved to be more difficult than expected. In several cases, subjects categorized the stimuli based on rules which corresponded to unidimensional decision boundaries along shape or texture levels. Although these decision boundaries were compatible with the configurations recovered from the subjects' similarity data in a certain number of cases, they *clearly contradicted* subjects' similarity-based stimulus representations in several other cases. This result was surprising to us given the large amount of evidence that perceptual categorization is intrinsically related to perceptual similarity (Goldstone, 1994; Hahn & Ramscar, 2001).

The tendency for subjects to sort according to unidimensional rules as opposed to similarity could have been an artifact of the free sorting task, as pointed out by one of our reviewers. A number of studies in the cognitive psychology literature (e.g., Ahn & Medin, 1992; Imai & Garner, 1965; Medin, Wattenmaker, & Hampson, 1987), have reported the use of single-feature rules in free sorting tasks. Interestingly, Regehr and Brooks (1995) found that when stimuli were presented all at once (the traditional "array procedure" for free sorting), subjects used unidimensional rules, but when stimuli were presented sequentially and matched to standards of Categories A and B (present at all times), subjects suddenly began sorting according to similarity. A recent study showed that the match-to-standards procedure only led to similarity-based sorting for a perceptually simple stimulus set (a sequence of line drawings of basic geometrical shapes) whose dimensions were spatially separated (Milton & Wills, 2004). For a more perceptually complex stimulus set (schematic butterflies) with spatially co-located features, subjects again resorted to unidimensional rules. These results are consistent with our findings considering that our 3D object stimuli are "perceptually complex" and vary in terms of shape and texture, two features which are spatially co-located. However, it is important to note that our task was neither a match-to-sample nor a classical array procedure; rather, subjects had to construct their categories and then sequentially assign stimuli to them. This may have imposed a significant working memory load which is not present in the other tasks. One study of memory-based category construction found that although sensitivity to similarity relationship was observed in perceptual sorting, subjects preferred to sort according to single dimensions in memory-based tasks (Wattenmaker, 1992). Thus, the memory requirements of our task may have been another factor which encouraged the use of unidimensional rules. A final factor could be that our stimuli only varied along two dimensions; it has been shown that subjects tend to classify using similarity when objects vary simultaneously along many dimensions, but prefer unidimensional rules when objects vary along fewer dimensions (Smith, 1981). Further studies are needed to disentangle the effects of stimulus dimensions and task design on category construction and to determine whether, under certain circumstances, categorization can be predicted from similarity in a multimodal setting.

### 3.4. A new approach to the study of visuohaptic processing

This study makes use of a novel combination of computer graphics, 3D printing technology, and MDS techniques presented in Cooke et al. (2005). In recent years, studies of visual perception have profited from advances in computer graphics and virtual reality, but studies of haptic perception have been hampered by the lack of adequate haptic presentation devices and the paucity of techniques available to easily create artificial, controlled three-dimensional stimuli. For example, stimuli have been made by precision-cutting (Klatzky et al., 1987; Lakatos & Marks, 1999), casting (Norman et al., 2004), moulding by hand (Forti & Humphreys, 2005; James et al., 2002), or manually assembling toy bricks (Forti & Humphreys, 2005; Newell, Ernst, Tjan, & Biilthoff, 2001). The technique used here facilitates the production and reproduction of novel, 3D objects and allows for a high degree of control over object properties. In addition, the combination of parametrically-varying stimuli and MDS techniques allows for intuitive visualizations and quantification of relative differences in feature weights. Another advantage of this approach is that it can be used to generate stimulus maps and dimension weights using any kind of proximity data gathered on parametrically varying stimuli. For example, maps and dimension weights generated by computational models of the visual system can be tested against those provided by human viewers, as we have demonstrated in Cooke et al. (2005), and the same could conceivably be done to evaluate computational models of the haptic system and/or models of visuohaptic perception. Given its broad potential applicability, the method offers a valuable tool for research in multisensory processing.

## 4. Summary and outlook

This study provides clear evidence that the perceptual modality used to interact with objects affects the representations used

+ Model
ARTICLE IN PRESS



for similarity judgments and categorization. The relative importance of shape and texture varied systematically according to modality for both our similarity and categorization tasks: shape was more important than texture when tasks were performed using vision only, whereas texture and shape were roughly equally important when tasks were performed either haptically or bimodally. We were able to model these differences as a modality-dependent rescaling of a single map, suggesting similar or perhaps even common multimodal representations.

The study also demonstrates a connection between similarity and categorization within a multisensory context. The same basic modality effects on dimension weights were observed for similarity and categorization tasks; in addition, the probability of within-category membership increased with perceptual similarity, while the probability of a category boundary being placed between neighbouring stimuli decreased with similarity. The lack of a stronger connection between similarity space and category structure was discussed in relation to the free sorting task which may have encouraged the use of unidimensional rules; additional studies involving are required to test this hypothesis.

Further work is needed to generalize these results by applying the same methodology to new stimulus sets which vary, for example, in terms of part–whole configuration or in terms of the scales at which macrogeometrical and microgeometrical manipulations are applied. Systematic variation of exploratory procedures and viewpoint will also be important steps towards the goal of understanding the cognitive consequences of multisensory object perception.

## Acknowledgements

The authors wish to thank Fiona Newell and Douglas Cunningham for helpful conversations, Michael Renner and Martin Breidt for technical assistance with stimulus production, as well as Karin Bierig and Bianca Arsene for their role in running the experiments. Informed consent was obtained from all subjects who participated in the experiments.

## References

Ahn, W., & Medin, D. (1992). A two-stage model of category construction. *Cognitive Science*, *16*(1), 81–121.

Amedi, A., Jacobson, G., Hendler, T., Malach, R., & Zohary, E. (2002). Convergence of visual and tactile shape processing in the human lateral occipital complex. *Cerebral Cortex*, *12*, 1202–1212.

Amedi, A., Malach, R., Hendler, T., Peled, S., & Zohary, E. (2001). Visuo-haptic object-related activation in the ventral visual pathway. *Nature Neuroscience*, *4*(3), 324–330.

Borg, I., & Groenen, P. (1997). *Modern multidimensional scaling*. Springer.

Carroll, J., & Chang, J. (1970). Analysis of individual differences in multidimensional scaling via an *n*-way generalization of "Eckart–Young" decomposition. *Psychometrika*, *35*, 283–319.

Cooke, T., Steinke, F., Wallraven, C., & Bülthoff, H. (2005). A similarity-based approach to perceptual feature validation. In *APGV 2005—Symposium on applied perception in graphics and visualization*.

Cox, T., & Cox, M. (2001). *Multidimensional scaling* (2nd ed.). Chapman & Hall.

Easton, R., Greene, A., & Srinivas, K. (1997). Transfer between vision and haptics: Memory for 2-d patterns and 3-d objects. *Psychonomic Bulletin & Review*, *4*(3), 403–410.

Easton, R., Srinivas, K., & Greene, A. (1997). Do vision and haptics share common representations? Implicit and explicit memory within and between modalities. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *23*(1), 153–163.

Edelman, S. (1999). *Representation and recognition in vision*. MIT Press.

Ernst, M., & Banks, M. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*, 429–433.

Ernst, M., & Bülthoff, H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Science*, *8*(4), 162–169.

Forti, S., & Humphreys, G. (2005). Cross-modal visuo-tactile matching in a patient with a semantic disorder. *Neuropsychologia*, *43*(11), 1568–1579.

Garbin, C. (1988). Visual-haptic perceptual nonequivalence for shape information and its impact upon cross-modal performance. *Journal of Experimental Psychology. Human Perception and Performance*, *14*(4), 547–553.

Garner, W. (1974). *The processing of information and structure*. Maryland: Lawrence Erlbaum Associates.

Goldstone, R. (1994). The role of similarity in categorization: Providing a groundwork. *Cognition*, *123*, 125–157.

Goldstone, R., & Steyvers, M. (2001). The sensitization and differentiation of dimensions during category learning. *Journal of Experimental Psychology. General*, *130*(1), 116–139.

Guest, S., & Spence, C. (2003). Tactile dominance in speeded discrimination of textures. *Experimental Brain Research*, *150*(2), 201–207.

Hahn, U., & Ramscar, M. (Eds.). (2001). *Similarity and categorization*. Oxford University Press.

Hollins, M., Faldowski, R., Rao, S., & Young, F. (1993). Perceptual dimensions of tactile surface texture: A multidimensional scaling analysis. *Perception Psychophysics*, *54*(6), 687–705.

Imai, S., & Garner, W. (1965). Discriminability and preference for attributes in free and constrained classification. *Journal of Experimental Psychology*, *69*, 596–608.

James, T., Humphrey, G., Gati, J., Servos, P., Menon, R., & Goodale, M. (2002). Haptic study of three-dimensional objects activates extrastriate visual areas. *Neuropsychologia*, *40*, 1706–1714.

Klatzky, R., Lederman, S., & Reed, C. (1987). There's more to touch than meets the eye: The salience of object attributes for haptics with and without vision. *Journal of Experimental Psychology. General*, *116*(4), 356–369.

Lakatos, S., & Marks, L. (1999). Haptic form perception: Relative salience of local and global features. *Perception & Psychophysics*, *61*(5), 895–908.

Landau, B., & Leyton, M. (1999). Perception, object kind, and object naming. *Spatial Cognition and Computation*, *1*, 1–29.

Lederman, S., & Klatzky, R. (1993). Extracting object properties through haptic exploration. *Acta Psychology*, *84*, 29–40.

Lederman, S., Summers, C., & Klatzky, R. (1996). Cognitive salience of haptic object properties: Role of modality-encoding bias. *Perception*, *25*, 983–998.

Medin, D., & Schaffer, M. (1978). Context theory of classification learning. *Psychological Review*, *85*(3), 207–238.

Medin, D., Wattenmaker, W., & Hampson, S. (1987). Family resemblance, conceptual cohesiveness, and category construction. *Cognitive Psychology*, *19*(2), 242–279.

Milton, F., & Wills, A. (2004). The influence of stimulus properties on category construction. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *30*(2), 407–415.

Morrongiello, B., Humphrey, G., Timney, B., Choi, J., & Rocca, P. (1994). Tactual object exploration and recognition in blind and sighted children. *Perception*, *23*(7), 833–848.

Murphy, G., & Medin, D. (1985). The role of theories in conceptual coherence. *Psychological Review*, *92*(3), 289–316.

Newell, F., Ernst, M., Tjan, B., & Biilthoff, H. (2001). Viewpoint dependence in visual and haptic object recognition. *Psychological Science*, *12*(1), 37–42.

Norman, J., Norman, H., Clayton, A., Lianekhammy, J., & Zielke, G. (2004). The visual and haptic perception of natural object shape. *Perception & Psychophysics*, *66*(2), 342–351.

Nosofsky, R. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology. General*, *115*(1), 39–57.

Nosofsky, R. (1992). Similarity scaling and cognitive process models. *Annual Review of Psychology*, *43*, 25–53.

Pietrini, P., Furey, M., Ricciardi, E., Gobbini, M., Wu, W., Cohen, L., et al. (2004). Beyond sensory images: Object-based representations in the human ventral pathway. *Proceedings of the National Academy of Sciences of the United States of America*, *101*, 5658–5663.

Reales, J., & Ballesteros, S. (1999). Implicit and explicit memory for visual and haptic objects: Cross-modal priming depends on structural descriptions. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *25*(3), 644–663.

Regehr, G., & Brooks, L. (1995). Category organization in free classification: The organizing effect of an array of stimuli. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *21*(1), 347–363.

Rock, I., & Victor, J. (1964). Vision and touch: An experimentally created conflict between the two senses. *Science*, *143*, 594–596.

Rosch, E., & Mervis, C. (1975). Family resemblances: Studies in the internal structures of categories. *Cognitive Psychology*, *7*, 573–605.

Rosch, E., Mervis, C., Gray, W., Johnson, D., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*, 382–439.

Shepard, R., & Cermak, G. (1973). Perceptual-cognitive explorations of a toroidal set of free-form stimuli. *Cognitive Psychology*, *4*, 351–377.

Sigala, N., & Logothetis, N. (2002). Visual categorization shapes feature selectivity in the primate temporal cortex. *Nature*, *415*, 318–320.

Smith, L. (1981). Importance of the overall similarity of objects of adults' and children's classifications. *Journal of Experimental Psychology. Human Perception and Performance*, *7*(4), 811–824.

Tversky, B., & Hemenway, K. (1984). Objects, parts, and categories. *Journal of Experimental Psychology. General*, *113*(2), 169–197.

Wattenmaker, W. (1992). Relational properties and memory-based category construction. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *18*(5), 1125–1138.

Young, F., & Harris, D. (2003). ALSCAL. In *SPSS 12.0 command syntax reference* (pp. 100–116). Chicago, IL: SPSS.

# Multidimensional Scaling Analysis of Haptic Exploratory Procedures

THERESA COOKE
and
CHRISTIAN WALLRAVEN
and
HEINRICH H. BÜLTHOFF
Max Planck Institute for Biological Cybernetics

Previous work in real and virtual settings has shown that the way in which we interact with objects plays a fundamental role in the way we perceive them. This paper uses multidimensional scaling (MDS) analysis to further characterize and quantify the effects of using different haptic exploratory procedures (EPs) on perceptual representations. In Experiment 1, twenty subjects rated similarity on a set of nine novel, 3D objects varying in shape and texture after either following their contours, laterally rubbing their centers, gripping them, or sequentially touching their tips. MDS analysis was used to recover perceptual maps of the objects and relative weights of perceptual dimensions from similarity data. Both the maps and relative weights of shape/texture properties were found to vary as a function of the EP used. In addition, large individual differences in the relative weight of shape/texture were observed. In Experiment 2, 17 of the previous participants repeated Experiment 1 after an average of 105 days. The same patterns of raw similarity ratings, perceptual maps, dimension weights, and individual differences were observed, indicating that perceptual similarities remained stable over time. The findings underscore the role of hand movements and individual biases in shaping haptic perceptual similarity. A framework for validating multimodal virtual displays based on the approach used in the study is also presented.

Categories and Subject Descriptors: H.5.1 [**Information Interfaces and Presentation**]: Multimedia Information Systems—Artificial, augmented and virtual systems, evaluation/methodology; H.5.2 [**Information Interfaces and Presentation**]: User Interfaces—Haptic I/O, evaluation/methodology

General Terms: Measurement, Design, Reliability

Additional Key Words and Phrases: haptic, exploratory procedure, shape, texture, similarity, multidimensional scaling

## 1. INTRODUCTION

This study investigates how the way in which we interact with objects changes the way we perceive them. More specifically, it addresses how haptic exploratory procedures (EPs) affect the perception of object similarities. A number of validation studies have been carried out in the haptic virtual reality community, but many have focused on optimizing *device*

Permission to make digital/hard copy of all or part of this material without fee for personal or classroom use provided that the copies are not made or distributed for profit or commercial advantage, the ACM copyright/server notice, the title of the publication, and its date appear, and notice is given that copying is by permission of the ACM, Inc. To copy otherwise, to republish, to post on servers, or to redistribute to lists requires prior specific permission and/or a fee.
© 20YY ACM 0000-0000/20YY/0000-0001 $5.00

parameters. However, given the tight coupling between perception and action in haptics, it is critical to understand how the *action* parameters of a system's user affect the perceptual outcome. There is a large body of work by Klatzky & Lederman concerning the role of EPs in *real-world* object recognition and classification (see [Lederman and Klatzky 1993] for a review). In virtual environments, Klatzky & Lederman have also studied the effects of exploratory factors and tool parameters [Klatzky et al. 2003]. The present paper represents an extension of Klatzky & Lederman's work on exploratory procedures in which the perceptual effects of changing EPs are visualized and quantified using a multidimensional scaling (MDS) framework, recently developed for studies of crossmodal human perception and validation of computer vision algorithms [Cooke et al. 2005; Cooke et al. 2006]. In this paper, we also discuss how the MDS framework can be used as a tool for comparing and benchmarking perception in real and virtual environments.

In Experiment 1, twenty participants haptically explored pairs of novel, 3D objects which varied in shape and texture (Figure 1) and rated the similarity between pairs of objects. Each participant explored the objects using four different EPs. Three of the EPs provided access to both shape and texture information (contour-following (CF), sequential exploration of object tips (TP) and enclosure or gripping (GR)), while the fourth EP, lateral rubbing motion along the surface (LM), provided access to texture information only. Similarity data were analyzed using MDS, resulting in perceptual maps of stimuli and individual-specific dimension weights. Maps resulting from the use of different EPs were compared to test whether a change in haptic exploratory procedure affected participants' perceptual representation of the objects. The results show that using different EPs to explore a set of objects does indeed affect representations, but also that there are large individual differences, specifically in the relative weights of object properties. In Experiment 2, we sought to test the stability of these results over time. Participants in the first experiment returned after several months to repeat the experiment. Consistent patterns of raw similarity, perceptual maps, and individual weights were observed, indicating that perceptual similarities remained stable over time.

The paper is organized as follows: we first review related work on exploratory procedures, then describe the methods used to carry out and analyze results of the psychophysical experiments. Results and their relevance for haptic interface design are discussed. Finally, we present a framework for applying similarity-based methods to study human perception in real and virtual environments and to validate haptic devices.

## 2. RELATED WORK ON EXPLORATORY PROCEDURES

[Lederman and Klatzky 1987; 1993] carried out seminal work on exploratory procedures, classifying typical hand movements into six types (lateral motion, pressure, static contact, holding, enclosure, and contour following) and characterizing each type based on factors such as compatibility with other EPs and execution speed. Of particular interest for this study is their demonstration that the choice of EP determines the nature of the information which can be extracted about an object. For each EP, they estimated *EP-to-property weightings* which represent the extent to which an object property can be extracted using a given EP. For instance, lateral motion, a back-and-forth rubbing motion of the fingers over a surface, is best-suited for extracting texture, but provides little or no shape information. Enclosing objects in the hand provides information about global shape and texture, but little exact shape information. Contour-following provides access to texture and global

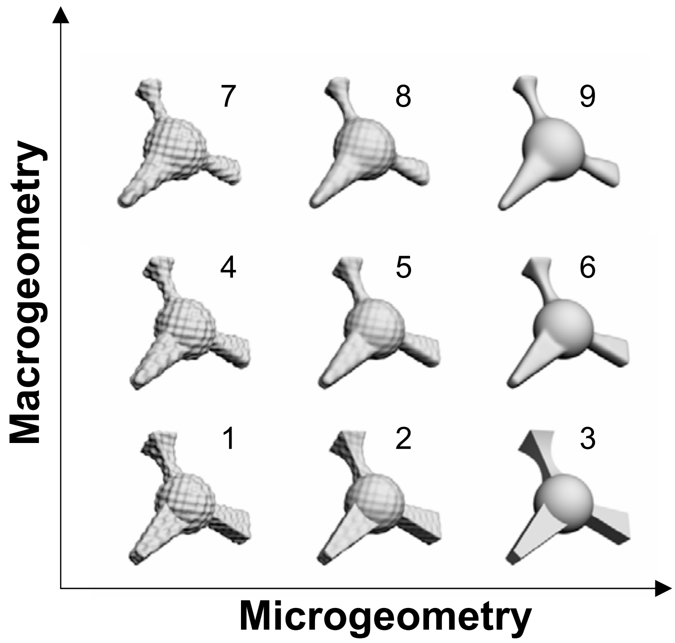


Fig. 1. Stimuli: The stimuli consisted of 9 novel, 3D objects varying in terms of microgeometry (”texture”) and macrogeometry (”shape”). Objects were created with 3D modelling software and manufactured in plastic using a 3D printer. Along the microgeometry axis, the objects’ bumpy texture gradually becomes smoother. Along the macrogeometry axis, sharp angles in objects’ meshes are relaxed.

shape, while also providing the most information about an object’s exact shape.

The effect of using two different exploratory procedures to judge similarities amongst a set of objects was examined in [Lakatos and Marks 1999]. Subjects explored 16 geometric objects varying in local and global geometry, using either a contour-following or an enclosure procedure and then rated the similarity between pairs of objects. Ratings were comparable in both conditions and the authors concluded that neither EP (contour-following or enclosure) was exclusively associated with a differential emphasis on local versus global shape. However, they found an effect of exploration *time* on the weighting of local vs. global shape: subjects were biased towards local shape for exploration times of 1s and 4s, but this effect decreased significantly for exploration times of 8s and 16s, i.e., global shape became more important for judging similarity when more time was provided for exploration.

The role of exploratory procedures (or modes of interaction) has also been addressed in haptic perception involving haptic devices. In [Dostmohamed and Hayward 2005], users interacted with a virtual fingerpad display using one of four modes (one or two finger, active or semi-active exploration). They found that curvature discrimination varied as a function of interaction mode: active, two-finger exploration offered higher sensitivity than active one-finger exploration and one/two-finger semi-active exploration.

In this study, we demonstrate how multidimensional scaling (MDS) techniques together with a parametrically-varying stimulus set can be used to gain further insight into the perceptual consequences of changes in exploratory procedures. MDS refers to a family of algorithms which operate on proximity data taken between pairs of objects. The output is a configuration of objects embedded in a multidimensional space. Psychologists have used MDS to explore perceptual representations of visually and haptically explored object sets e.g., [Shepard and Cermak 1973; Garbin and Bernstein 1984; Hollins et al. 1993; Bergmann et al. 2006; Cooke et al. 2006]. The technique has also found a large following in domains such as marketing [Carroll and Green 1997] and knowledge mapping [Chen 2003] because it allows for the identification of important psychological dimensions of stimulus variation (e.g., dimensions along which buyers differentiate amongst competing products) and quantification of perceptual distances between stimuli (e.g., how "closely-related" one field of research is to another). In cognitive psychology, the inputs usually consist of human similarity ratings taken over a set of objects; the output configuration is then interpreted as a *map of the objects in a psychological space* which explains the similarity data [Borg and Groenen 2005].

MDS analysis provides the following types of information about the psychological representation of stimuli:

(1) how many dimensions of variation in the stimuli are apparent to the participants;
(2) whether these dimensions correspond to properties which were deliberately being manipulated;
(3) whether one or more unexpected perceptual dimensions were also apparent to the participants;
(4) the relative weights of the psychological dimensions;
(5) interstimulus distances in the psychological space.

The goals of this paper are to use MDS to gain new, quantitative insight into the specific question of how exploratory procedures shape the perceptual representation of objects and the more general question of how MDS approaches can be applied to study human perception in both real and virtual environments.

## 3. METHODS

This section describes the stimuli used in the experiments, the experimental procedure, and MDS analysis.

### 3.1 Stimuli: Novel 3D Objects

A family of nine novel, 3D objects (Figure 1) was used in the experiments. The objects were designed in the 3D graphics software package 3D Studio Max (3DS) and manufactured using a 3D printer (Dimension 3D Printer, Stratasys, Minneapolis, USA). The printed 3D models were made of hard, white, opaque plastic (acrylnitrile butadene styrene). The objects were 9.0 +/- 0.1 cm wide, 8.3 +/- 0.2 cm high, and 3.7 +/- 0.1 cm deep and weighed about 40 g.

Each object consisted of three parts connected to a center sphere, defining the object's macrogeometry ("shape"), plus a displacement map applied to the 3D mesh, defining the object's microgeometry ("texture"). The displacement map was made up of repeated conical elements with base widths of 3mm, peak widths of 2mm, maximum height of 2mm from the surface of the object, and inter-element spacing of 3-5mm. As such, the texture can be considered as a "macrotexture" encoded by SAI mechanoreceptors [Klatzky and Lederman 2003]. Variations amongst the objects were generated by smoothing object geometry at two scales, microgeometrical and macrogeometrical. Microgeometrical smoothing was achieved by decreasing the amount of displacement of 3D vertices caused by the displacement map. Macrogeometrical smoothing was accomplished via the 3DS "Relax" operator, which moves vertices towards a local average 3D position, thereby gradually reducing sharp angles in the macrogeometry. This resulted in changes in macrogeometry at a scale of about 1 $cm^2$. Macrogeometrical changes at this scale can be extracted via SAI mechanoreceptors [Srinivasan and LaMotte 1991], by integrating local changes in curvature across the duration of finger motion [Pont et al. 1999], and by using kinesthetic cues in muscle spindles generated during exploration [Klatzky and Lederman 2003].

Note that manipulations created *input* dimensions corresponding to parameters in the 3D software package, but this does not imply that these dimensions will necessarily be recovered in the perceptual *output* space. Revealing the dimensions which are important for human perception is precisely one of the reasons for performing MDS analysis of human similarity ratings.

### 3.2 Experiment 1: Haptic Similarity Ratings with Naive Participants

Twenty naive, right-handed subjects (10 men, 10 women) were paid 8 Euros per hour to participate in the experiment. Their task was to rate the similarities between pairs of objects on a scale between 1 (low similarity) and 7 (high similarity) after exploring them haptically. The same experimental setup was used in all conditions (Figure 2). Subjects used a chin rest placed 40 cm away from a stand on which the objects were mounted. The height of the chin rest was set such that the centre of the object was aligned with the line of sight. A set of grooves and a section of rubber tubing on the mount piece ensured that the objects were securely held in place in exactly the same upright position on every trial. A black metal sheet stood between the subjects and the stand to obstruct their view of the objects.

On each trial, the experimenter placed the first object on the stand, verbally instructed the subject to start the exploration, counted to three using a stopwatch as a metronome, and removed the object after 3s. The presentation time of 3s was chosen because it was the minimum amount of time that subjects needed to perform the longest of the procedures (contour-following) in a pilot experiment. The experimenter then replaced the first object by a second object, instructed the subject to begin exploration, and removed the object after 3s. The experimenter then waited for the subject's response. Before the experiment, subjects were sequentially presented with the two pairs of objects in the outermost corners of the space, allowed to palpate each one in their hand for about 5s and told that these were the largest differences they would encounter in the experiment.

The experiment consisted of four blocks of 45 randomized trials (each object was compared once with itself and once with every other object resulting in $9 + (9 \cdot 8)/2 = 45$ trials) and the order of appearance of stimuli was randomized over blocks. In each of the four blocks, subjects explored the objects using a different procedure. The order of procedures was randomly selected for each subject. The following procedures were used:

—*contour-following (CF)*, which provides information about a broad range of object properties (texture, hardness, temperature, weight, volume, global shape, and exact shape), while being specialized for the extraction of exact shape [Lederman and Klatzky 1993];

—*enclosure/gripping (GR)*, which also provides information about a wide range of object properties except exact shape, is relatively quick to perform, and is compatible with almost all other EPs (e.g., you can apply pressure to test an object's hardness while gripping);

—*lateral motion (LM)*, a side-to-side rubbing restricted to the objects' centres; lateral motion is known to be particularly well-suited to extraction of an object's texture - by restricting exploration to the objects' centers, we expected to remove the ability to detect changes in local shape;
—*tip-touching (TP)*, a brief (1s) contact of each of the three object tips which we expected to focus subjects' attention on changes in shape.

Each exploration trial took about 20-30 seconds and the experiment ran for approximately two hours. At the end of the experiment, subjects were asked to write a short description of the objects, to describe the properties they had used to judge similarity using each EP, and to comment on whether they found any particular EP easier or harder to perform than the others. Just before leaving the experimental room, they viewed the objects for a few minutes.

### 3.3 Experiment 2: Haptic Similarity Ratings with Experienced Participants

Seventeen of the original participants in Experiment 1 returned on average 105 days after their initial visit (SE=4 days, MIN=69 days, MAX=130 days). Of the three remaining participants in Experiment 1, two could not be reached to repeat the experiment and one was excluded because he had not viewed the objects at the end of Experiment 1. Experiment 2 was conducted using the same protocol as Experiment 1, with three notable differences: 1) in Experiment 2, participants had had prior experience with the similarity rating task and the EPs; 2) in Experiment 2, participants had had prior haptic experience with the objects; 3) in Experiment 2, participants had had brief visual exposure to the objects (at the end of Experiment 1). After Experiment 2, subjects were asked to rate the similarity of their experience in Experiment 2 as compared to Experiment 1 on a scale between 1 (very different) and 5 (exactly the same) and explain their rating.

### 3.4 MDS Analysis of Similarity Data

Human similarity matrices were analyzed using two variants of MDS, both implemented in the SPSS ALSCAL MDS package [Carroll and Chang 1970; Young and Harris 2003].[1] The first variant we used was replicated MDS (RMDS), which allows for the simultaneous analysis of several similarity matrices simultaneously; for a given dimensionality, it returns a single map and a goodness-of-fit measure, Kruskal's STRESS, a normalized difference between the fitted distances and the observed proximities. Although establishing a threshold for acceptable values of stress is often debated, Monte Carlo studies have demonstrated that stress values below 0.2 indicate that the output configuration fits the similarity data well [Cox and Cox 2001, p.79]; another common heuristic is to look for a sharp drop in stress values, the so-called "statistical elbow" in the plot[Cox and Cox 2001, p.88]. RMDS was used to analyze all the similarity matrices gathered using a given EP; maps and stress values were computed for 1-4 dimensional solutions. The second MDS variant that was used is called weighted MDS (WMDS), often referred to as individual differences scaling (INDSCAL). WMDS takes in one input matrix of similarities per subject and returns a "base" map of stimuli together with a set of weights for each subject, which specify how the dimensions of the base map should be stretched in order to best fit the individual data.

[1] For all analyses, the distance metric was taken as Euclidean, proximities were taken as ordinal measurements (i.e., non-metric), and untying of tied proximities was allowed.

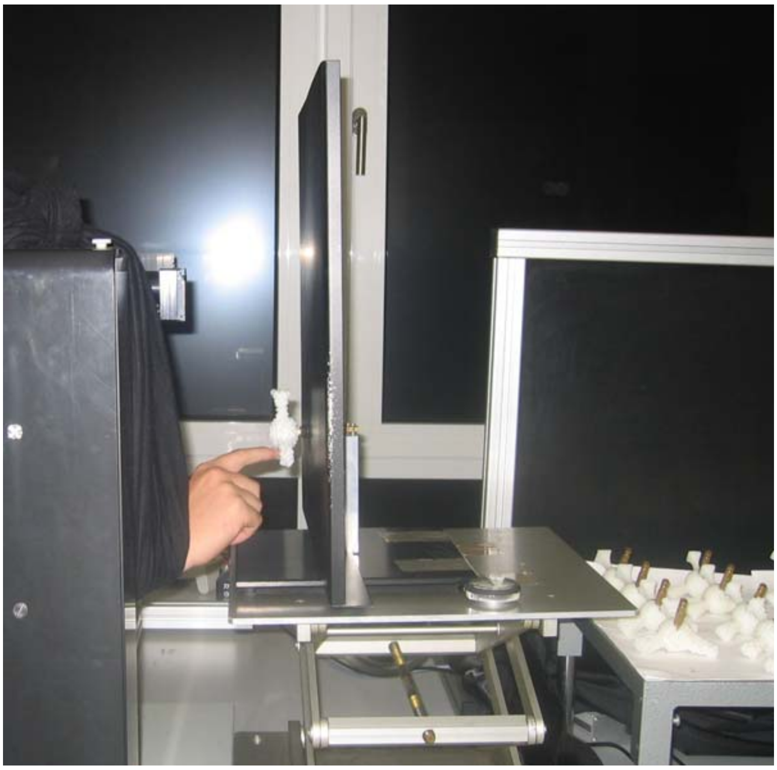

Fig. 2. Experimental setup for haptic similarity ratings. The experimenter places the objects on a mount placed behind an opaque curtain. The participant haptically explores the object using one of four exploratory procedures.

Low stress for a WMDS model indicates that the input matrices can be modelled by linearly rescaling a single underlying map. As input to WMDS, we used similarity matrices gathered from all EP conditions and all subjects. This allowed us to test whether the model of a base map and individual weights could account for our data and, if so, to compare weights across participants and EPs.

### 3.5 Analysis of Debriefing Questionnaires

For Experiment 1, we computed statistics from debriefing questionnaires concerning 1) salient object properties and 2) the ease of judging similarity using each EP.

The mention of object properties was tallied for two questions: 1) a question asking subjects to describe the objects overall and 2) questions asking subjects to describe how they made similarity judgments using each EP. For each question, we tallied the number of times each subject had made at least one reference to object shape, texture, or other property (e.g., material, colour, temperature). We also distinguished between references to global shape and references to local shape. References to part width, shape, or corner sharpness were counted as references to local shape, while references to holistic shape terms such as "star-like" were counted as references to global shape. When subjects mentioned a configuration of parts, such as "three ends which extend from a ball-shaped center", we counted this as both a reference to global and part shape. When subjects simply referred to "shape", this was counted as an "unspecified shape property".

Statistics on the ease of judging similarity for a given EP were computed by counting the number of subjects who said that judging similarity was "easiest" or "easier" with a given EP, as well as by counting the number of subjects who said that the task was "hardest" or "harder" with a given EP. Some subjects provided an overall ranking of the four EPs from easiest to hardest; in this case, the score for easier/easiest was increased by one for the first two EPs and the score for harder/hardest was increased by one for the second two EPs.[2]

## 4. EXPERIMENT 1: RESULTS AND DISCUSSION

We first discuss the results of performing RMDS analysis of EP-specific similarity data, then turn to the results of WMDS analysis performed using all similarity data.

### 4.1 Experiment 1: EP-specific RMDS Analysis

RMDS solutions were computed separately over the set of 20 subject-specific similarity matrices gathered using each EP. This was done for output dimensionalities of one through four. Stress values are shown in Table I and corresponding maps are shown in Figure 3 (top row).

For the LM procedure, stress fell below 0.2 for a one-dimensional solution. From visual inspection of the corresponding map, this single dimension corresponded to texture variation (all stimuli with the same texture level project to the same point in space and stimuli are ordered according to texture level). The fact that texture alone sufficed to explain similarity data in this condition was expected since lateral motion was restricted to the objects' centres, which were extended areas of low curvature offering little object-specific shape information. The dominance of texture was confirmed by the fact that this was the only property mentioned by subjects when describing what they had felt while performing this EP (Figure 4, left). In addition, subjects said that judging similarity was easiest using this EP (Figure 4, right), with the reasons being that similarity only had to be judged on the basis of a single dimension (particularly given the 3s time limit) and that the EP itself was easy to perform.

For the other EPs (TP, CF, and GR), stress only fell below 0.2 for *two-dimensional* RMDS solutions. From visual inspection of the corresponding maps, it can be seen that the dimensions corresponded to the texture and shape dimensions manipulated in the input stimulus space. Thus, subjects were able to perceive these manipulations and to recover ordinal relationships amongst the objects in that space, as was also shown for an extended set of 25 objects [Cooke et al. 2006]. Note that this ability of the human haptic perceptual system is quite remarkable given the high dimensionality of the measurement space. How did subjects themselves label these two dimensions of stimulus variation? In the questionnaires, they mentioned using variations in both shape and texture to judge similarity using TP, CF, and GR (Figure 4, left). Most references to shape concerned *local* shape properties such as the sharpness of corners, the part width, or tip surface area, although when using GR and CF, a few subjects did make use of global shape terms such as *"Gesamtform"* (shape of the whole). There was an interesting difference between the 2D maps obtained using these EPs: in the contour-following and tip-touching maps, the three shape rows were relatively equidistantly spaced, whereas in the gripping map, there was a larger separation between the bottom row (objects with sharp edges in the macrogeometry) and the middle row (objects with slightly rounded edges) than between the middle row and the upper row (objects with even more rounded edges). This could be related to decreased saliency of exact shape properties when gripping as opposed to contour-following or tip-touching (pulling the top two rows together) [Lederman and Klatzky 1993]. Alternatively,

[2]Some subjects focused their answer on the *comfort* of performing the EP as opposed to the ease with which *similarity* could be judged; these answers were not included here.

| EP | 1D Stress | 2D Stress | 3D Stress | 4D Stress |
|---|---|---|---|---|
| CF | 0.33 | **0.17** | 0.14 | 0.11 |
| GR | 0.33 | **0.19** | 0.15 | 0.12 |
| LM | **0.15** | 0.10 | 0.10 | 0.07 |
| TP | 0.40 | **0.18** | 0.13 | 0.11 |

Table I. RMDS stress: Experiment 1. For each EP, the first stress value to fall below 0.2 has been bolded.

it could reflect a cognitive, categorical separation between sharp and smooth objects which becomes more salient through gripping (pushing the bottom row away from the top two rows).

In summary, the RMDS analysis showed that subjects were able to recover the full dimensionality as well as ordinal relationships in the input space when using CF, GR, and TP; when using LM, subjects still recovered ordinal relationships in the input space, but only along the texture dimension. Questionnaire data confirmed that the dimensions of the perceptual output space could be labelled as shape and texture. Next, a WMDS analysis was used to compute comparable relative weights of these properties for similarity judgments using the four different EPs.

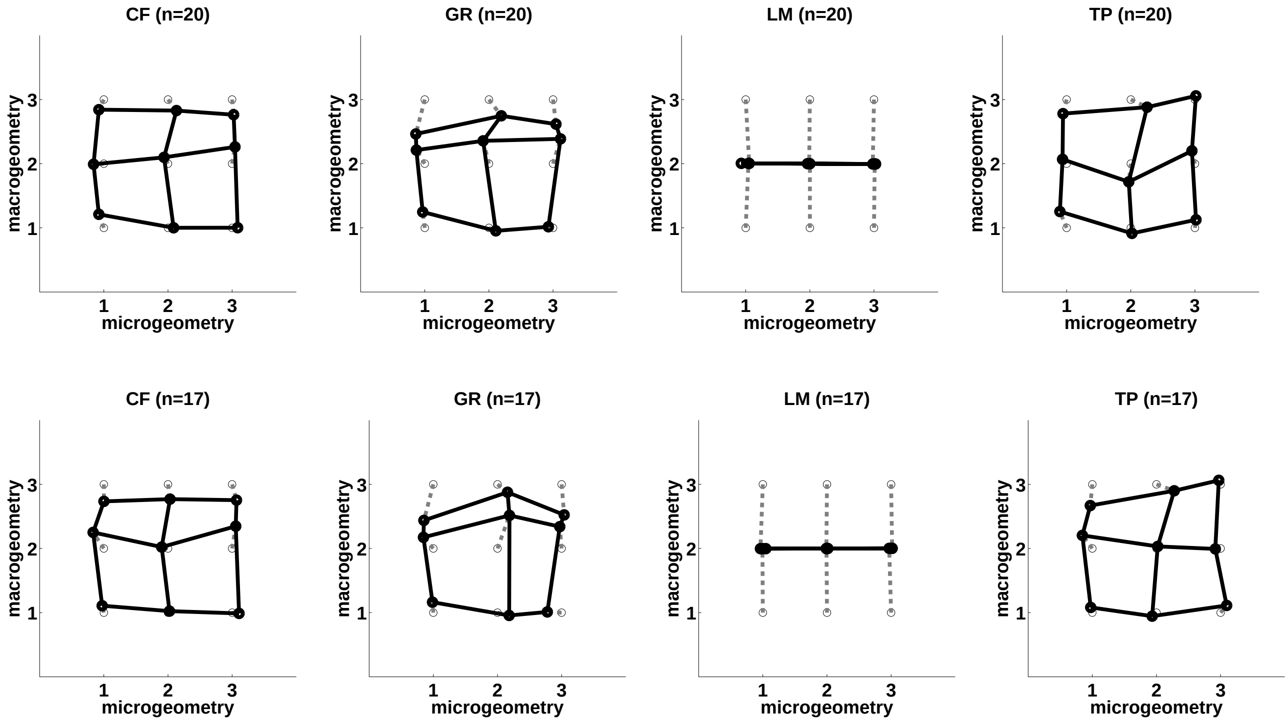


Fig. 3. Configurations recovered using RMDS on EP-specific similarity data. Top row: Experiment 1. Bottom row: Experiment 2. Note that for LM, the 1D MDS solution is shown. 2D solutions are shown for all other EPs.
.

## 4.2 Experiment 1: Global WMDS Analysis

A 2D WMDS solution was computed using all 80 similarity matrices (4 EPs x 20 subjects). This yielded a stress value of 0.16, indicating that the model of a single underlying 2D map with individually-adjusted weights provides a good fit to the data.

| EP | 1D Stress | 2D Stress | 3D Stress | 4D Stress |
|---|---|---|---|---|
| CF | 0.31 | **0.16** | 0.12 | 0.10 |
| GR | 0.35 | **0.19** | 0.15 | 0.10 |
| LM | **0.09** | 0.04 | 0.04 | 0.03 |
| TP | 0.34 | **0.17** | 0.13 | 0.11 |

Table II. RMDS stress: Experiment 2. For each EP, the first stress value to fall below 0.2 has been bolded.

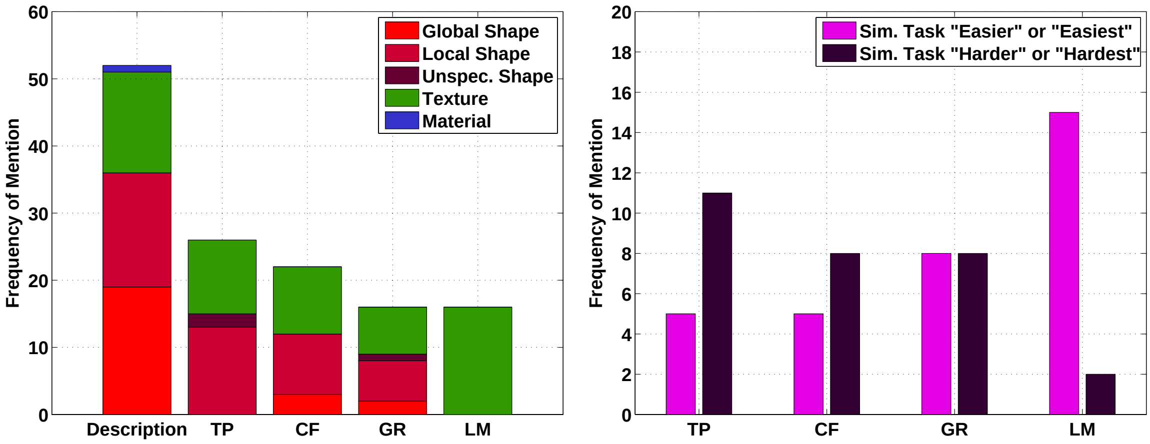


Fig. 4. Debriefing questionnaire. Left: Object properties mentioned by subjects when describing the objects in general and when describing what they felt using each EP. Right: Frequency with which subjects said that the similarity task was either "easier/easiest" and "harder/hardest" using a given EP. ("Unspecified shape": subject used the word shape without specifying whether they were referring to local or global shape.)

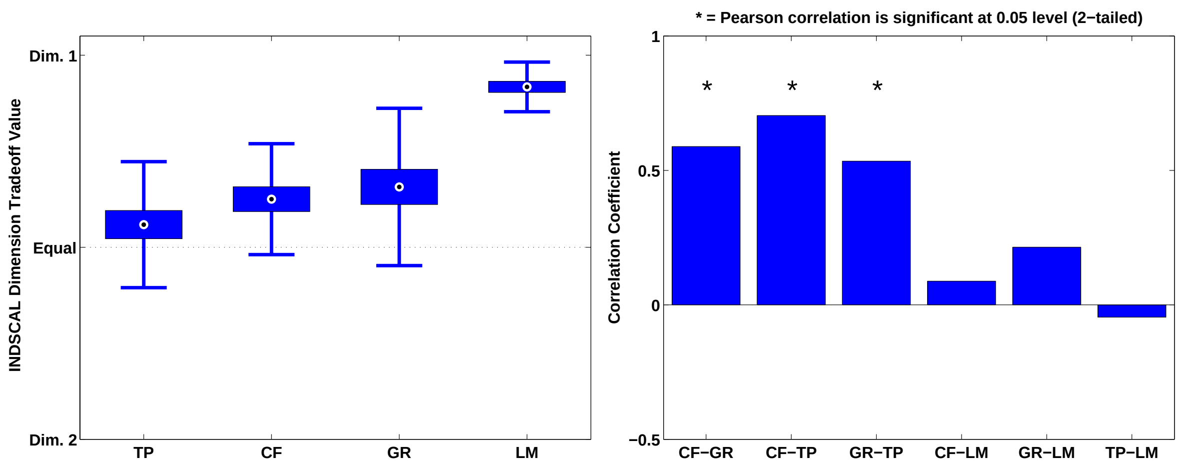


Fig. 5. Experiment 1: Weights per EP (left) and correlation between weights (right). For weight plot, total box height is twice the standard error (n=20); total whisker height is twice the standard deviation.

Figure 5 (left) shows the relative weight of the two dimensions for all subjects as a function of EP[3] (TP: M=0.56, SE=0.04, CF: M=0.63, SE=0.03; GR: M=0.66, SE=0.05; LM: M=0.92, SE=0.01).

[3]WMDS returns one weight per dimension, however these weights are constrained to lie on a circle in the 2D case. Therefore, we report a single value representing the relative weight of the two dimensions. It is calculated as the arc tangent of each point in the weight space, with 1 representing maximum weight of the first dimension and 0 representing maximum weight of the second dimension. For TP, GR, and CF, the first dimension corresponds

*Mean Values of WMDS Weights* For TP, CF, and GR, mean relative weights ranged from 0.56 to 0.66. These values agree with subjects' reports of having used both shape and texture to judge similarity (Figure 4, left) and with the finding from RMDS analysis that two dimensions are needed to account for similarity data in these three conditions. A two-sided T-test was carried out to test the hypothesis that relative weights in the GR, CF, and TP conditions came from distributions with equal means; the null hypothesis was not rejected (GR-CF: $p>0.5$, CF-TP: $p=0.18$, GR-TP: $p=0.1$). Therefore, when using any one of the three EPs which enabled extraction of both shape and texture, subjects weighted shape and texture in the same way on average. For the LM condition, a mean weight value of 0.92 was found, which corresponds with the finding that a single dimension corresponding to texture sufficed to explain similarity data.

One surprising aspect of this analysis was that there was no significant difference between mean weights in the CF and TP conditions. We had assumed that shape was not optimally extracted in the CF condition and designed the TP procedure to facilitate the extraction of differences in the objects' shape, expecting that this would result in a greater reliance on shape in judging similarity. There may be two explanations for the ineffectiveness of this manipulation:

(1) We were correct to assume that CF extracts shape information sub-optimally in this context, but the new procedure we designed did not further facilitate shape extraction. One reason could be that it was more difficult to perform, as evidenced by high difficulty ratings given to the TP procedure (Figure 4). The difficulty may have been related to the EP's discontinuous nature, making the physical movement more difficult to execute and perhaps also causing spatial disorientation (one subject reported getting "an incomplete, superficial impression" and "missing the centre of the object" while using this EP).

(2) Alternatively, shape information may already have been optimally extracted by contour-following and was not affected by the modifications made to create the TP procedure. This could be tested by increasing the relative amount of shape variation in the stimulus set and checking whether the use of shape increases by the same amount for both EPs.

*Variability in WMDS Weights* Interestingly, the WMDS analysis revealed a large amount of variability in the weights used by different subjects. The maximum absolute difference in weights was 0.56 for TP, 0.54 for CF, and 0.73 for GR. We first investigated whether variability could be linked to gender: although men did show a trend towards stronger shape bias than women (men: $M=0.59$, $SE=0.07$; women: $M=0.72$, $SE=0.05$), the difference was not statistically significant (two-sided, two sample T-test at 95% confidence level, $p=0.14$). To test whether biases could be related to an individual-specific but EP-general factor, we tested the Pearson correlation between tradeoff values for each pair of EPs. As shown in Figure 5 (right), significant correlations were found between all pairs of EPs for which both shape and texture could be extracted. Thus, it appears that when EPs allow for both properties to be extracted, individuals tend to weight the two properties similarly across EPs. If this is indeed the case, an interesting question is whether these

to texture and the second to shape. For LM, the first dimension corresponds to texture; no label can be attributed to the second dimension, which likely corresponds to noise.

subject-specific weights vary over time. To test this, we carried out Experiment 2, in which subjects repeated Experiment 1 after an interval of several months.

## 5. EXPERIMENT 2: RESULTS AND DISCUSSION

We proceed by first presenting the RMDS analysis of EP-specific similarity data and then discussing the results of the WMDS analysis performed over data gathered using all EPs.

### 5.1 Experiment 2: EP-specific RMDS Analysis

Stimulus maps (Figure 3, bottom row) as well as stress values (Table II) were computed using RMDS analysis of EP-specific similarity data. For LM, stress values dropped below 0.2 for a one-dimensional solution; using the output map, this dimension was labelled as texture. For the other EPs, stress values showed that two dimensions were required to explain similarity data. The maps show that these two dimensions correspond to texture and shape. An additional degree of perceptual separation between objects with sharp-edged macrogeometry (bottom row) and smooth-edged macrogeometry (top two rows) appeared when subjects gripped the objects.

### 5.2 Experiment 2: Global WMDS Analysis

A 2D WMDS solution was computed using all 68 similarity matrices (4 EPs x 17 subjects). This yielded a stress value of 0.15, indicating that the 2D WMDS model provides a good fit to the data. Figure 6 (left) shows the relative weights of the two dimensions for all subjects as a function of EP (TP: M=0.55, SE=0.03, CF: M=0.63, SE=0.04; GR: M=0.62, SE=0.04; LM: M=0.93, SE=0.01). Large variability in weights was observed for TP (MAX-MIN=0.33), CF (MAX-MIN=0.52), and GR (MAX-MIN=0.61). Significant correlations were found between individual weights for CF-GR, CF-TP, and GR-TP, but not for CF-LM, GR-LM or TP-LM (Figure 6, right), indicating that subjects weighted shape and texture similarly across EPs which allowed for extraction of both properties.

## 6. EXPERIMENTS 1 VS. 2: RESULTS AND DISCUSSION

*Subject Reports* When subjects were asked to rate the similarity between their experiences during the two experiments on a scale between 1 (very different) and 5 (exactly the same), the mean rating was 3.3 (SE=0.25). In justifying their ratings, subjects stated that the experimental protocol was identical, but that their experiences differed somewhat in that they were already familiar with the protocol at the outset of Experiment 2 and had prior knowledge about the objects before beginning Experiment 2.

*Raw Similarity Ratings* The mean signed difference in raw similarity ratings taken over all trials was -0.014 with a standard deviation of 1.4. A paired T-test did not reject the hypothesis that the two samples of similarity ratings (17 subjects x 4 EPs x 45 judgments per EP = 3060 judgments) came from distributions with equal means (p=0.65, t(3059)=0.45). Taken together, these results show that subjects did not systematically shift their ratings upwards or downwards.

*RMDS Maps* For each EP, similar stimulus maps were generated in both experiments (Figure 3, top vs. bottom rows). In both experiments, stress values showed that a single dimension corresponding to texture sufficed to explain similarity ratings performed based on LM, whereas two dimensions were required for the other EPs. All EPs led to relatively equidistant spacing along the texture dimension. CF and TP led to equidistantly-spaced objects along the shape dimension, whereas GR led to increased separation between the bottom and top two rows. The stability of these maps indicates that on an EP-by-EP basis, similar perceptual representations of the objects resulted from both experiments.

*Population WMDS Weights* To test whether the population mean of weights had changed between experiments, we carried out paired T-tests on the two matched samples of 17 weights calculated for each EP. No significant differences in the means were found for any EP. Furthermore, the large variability in weight distributions observed for TP, CF, and GR in Experiment 1 was also observed in Experiment 2. These results suggest that the overall distributions of weights remained stable over time.

*Individual WMDS Weights* To assess the amount of change in subject weights across the experiments, we first calculated *signed* differences in relative weights for each EP. For each subject, we tested whether the mean difference taken over the four EPs was different from 0. We found no significant differences for any subject, showing that no individual systematically changed their weights in one direction or another; in other words, no subject became more shape or texture-biased overall. Second, we computed *absolute* differences in relative weights for each subject and each EP and plotted the values in a histogram (Figure 7). Absolute differences of less than 0.2 were observed in 64 of 68 cases (17 subjects x 4 EPs), indicating that individual, EP-specific weights were quite consistent across the two experiments. Furthermore, the same correlations amongst weight pairs were found in the two experiments, i.e., in both experiments, individuals tended to use the same relative weight across EPs (Figure 6, right). Taken together, these results point to a notable degree of stability in individual weights over time.

*Grip Position as a Source of Variability in Weights* As noted above, individual variability was particularly marked in the GR condition. By photographing subjects' grip position in Experiment 2, we were able to correlate subjects' weights in the GR condition with grip position. Figure 8 shows examples of subjects whose selected grip contacted 0, 1, 2, or 3 of the objects' tips. The number of object tips contacted was found to be significantly correlated with relative weight in the GR condition (Figure 9). Although we did not photograph grip positions in Experiment 1, subjects reported using the same grip in both experiments; thus grip position may account for both inter-subject variability and consistency in weights over time in the GR condition.

## 7. GENERAL DISCUSSION

Experiment 1 showed that the choice of exploratory procedure led to differences in the subjects' haptic similarity judgments; using MDS analysis, perceptual maps of stimuli were visualized and differences in relative property weights were quantified. Both interstimulus distances (as seen in the maps) and relative shape/texture weights varied as a function of EP and a large degree of individual variability in weights was observed. Experiment 2 replicated these results and showed that both population and individual weights were stable after several months.

In this section, we discuss potential sources of intersubject variability and intrasubject stability in haptic similarity ratings and present a framework for applying MDS techniques to the study of human perception in real and virtual environments.

### 7.1 Intersubject Variability in Object Property Weights

In both experiments, we observed high variability in the way different subjects weighted shape and texture properties when both properties were made available by the selected

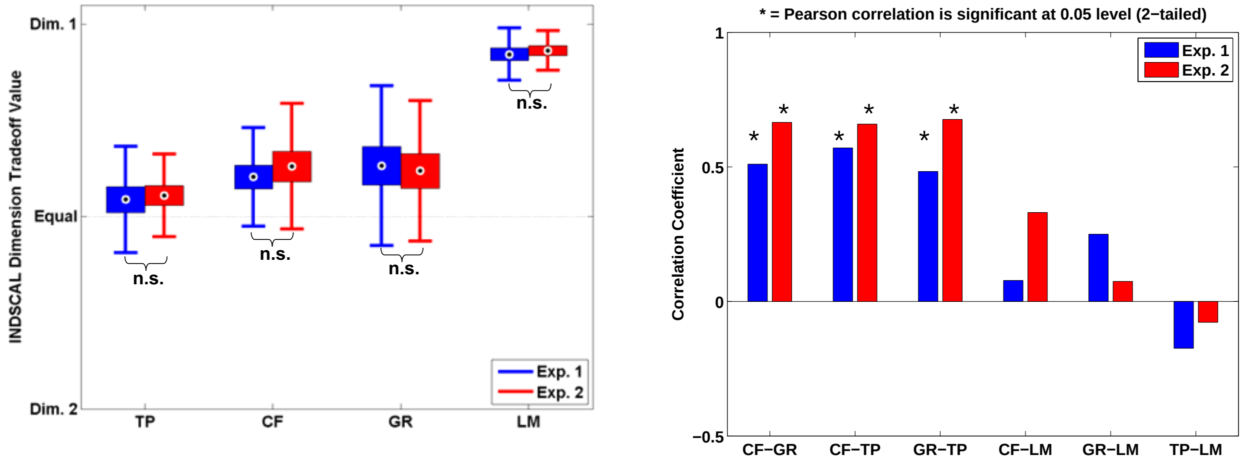


Fig. 6. Experiment 1 vs. 2: Weights per EP (left) and correlation between weights (right). For Experiment 1, only data from the 17 subjects who also participated in Experiment 2 is plotted for comparison. In the weight plot (left), the total box height is twice the standard error (n=17); total whisker height is twice the standard deviation.

.

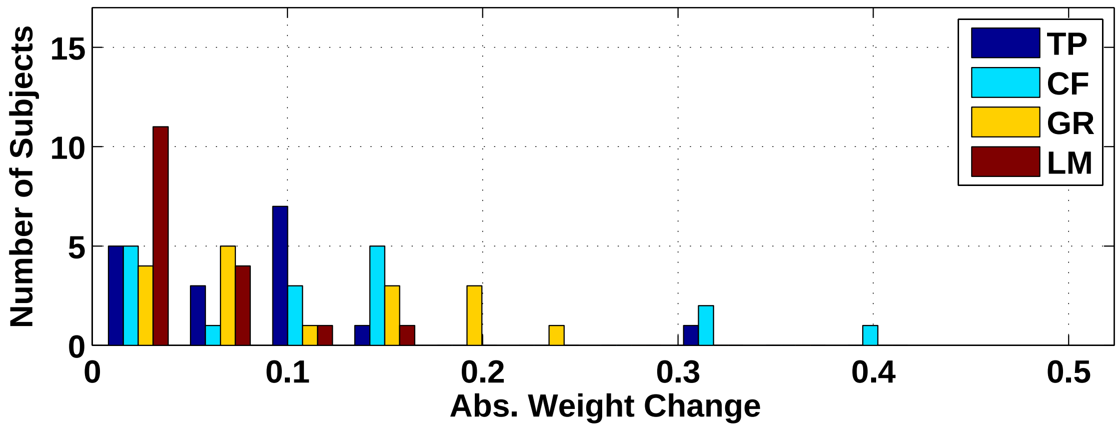


Fig. 7. Experiment 1 vs. 2: Histogram of absolute change in individual subject weights

.

EP. By photographing the way subjects gripped the objects, we were able to correlate variability in the GR condition with the number of tips contacted while gripping. But what could explain variability in the tip-touching and contour-following conditions? We hypothesize two possible sources:

(1) Variability could be due to an individual *execution* bias arising from subtle, but systematic differences in the way subjects perform the EPs (e.g., differences in pressure and velocity which were not measured in the current experiments). Differences in EP execution could then lead to differences in the relative amount of shape/texture information being extracted from the objects. Hand forces and dynamics need to be recorded in order to test this hypothesis.

(2) Variability could be due to an individual *cognitive* bias arising from individual preferences and/or experience, such as an internal prior on the relative reliability of property estimates ([Ernst and Bülthoff 2004]) or a difference in *a priori* cognitive saliencies of the two features [Lederman et al. 1996]. This cognitive bias could then cause an

Fig. 8. Experiment 2: Variation in number of tips contacted while gripping. From left to right: 0, 1, 2, and 3 object tips contacted. Top and bottom rows show different viewpoints of same subject's grip.
.

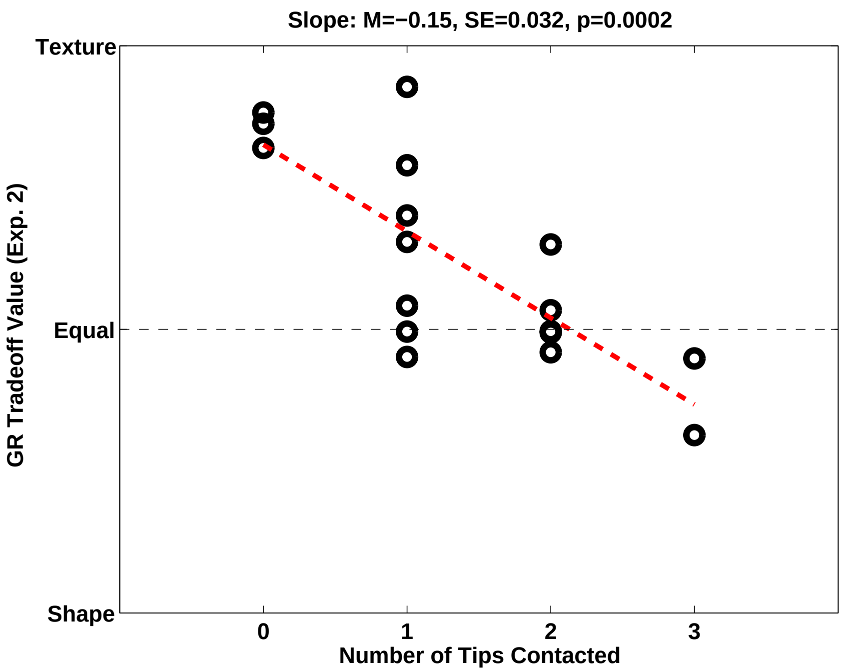


Fig. 9. Experiment 2: Weights in GR condition versus number of tips contacted during gripping.

internal reweighting of raw shape/texture information used to judge similarity. In our experiments, the use of a cognitive bias may have been favoured by the relatively high level of reported difficulty for TP and CF (Figure 4, right), especially given the 3s limitation on exploration time. In addition, the presence of a cognitive bias may also have led to systematic differences in EP execution. For example, [Riley et al. 2002] found that different perceptual intentions led to subtle differences in movement dynamics in an exploratory wielding task. The existence of such a bias could be tested by investigating whether individual subjects retain personal shape/texture biases when judging similarity on different sets of objects or over longer exploration times. It would also be interesting to test whether individual biases converge when two participants are allowed to communicate, as is the case with discrepant definitions of perceptual concepts [Barsalou 1999].

### 7.2 Intrasubject Stability in Object Property Weights

Even though different subjects used different property weights to judge similarity, individual subject weights remained quite consistent over the two experiments. One possible explanation for this is that during the first experiment, subjects built up and stored information about the objects and task in long-term memory, and that this information strongly influenced their performance in the second experiment. This information could pertain to the objects themselves (encoding structural models [Biederman 1987], view-based models [Bülthoff and Edelman 1992; Newell et al. 2001], or similarity relationships amongst objects [Edelman 1999]), to the experimental context, and/or to the dynamics of exploratory actions [Wippich et al. 1994]. On the other hand, if prior experience played a negligible role, stability could be explained by the fact that both experiments involved the same inputs being processed by the same perceptual system, and thus yielded the same output. This explanation is in line with the idea that detailed stimulus information is not stored in representations, but rather that "the world serves as an external memory store" [Simons 1996; O'Regan and Noe 2001] and would explain the fact that we obtained consistent results even after several months. If this hypothesis holds, subjects' performance should remain consistent even when performance on implicit and explicit tests for memory of objects, hand dynamics, and experimental context is poor.

### 7.3 An MDS Framework for Comparative and Validation Studies

Here, we place the study within a more general framework for MDS-based studies of human perception in real and virtual environments, illustrated in Figure 10. The perceptual process begins with the extraction of features from a set of real or virtual objects. Proximity data (e.g., similarity ratings) are then derived from these features. Note that in this paper, we have used human similarity ratings as proximity data, but proximities can also be computed directly based on physical object properties or derived from interaction parameters such as hand or tool dynamics. MDS is then used to 1) construct maps of the objects in perceptual spaces and 2) to compute relative dimension weights. Comparing these data provides an opportunity to visualize and quantify differences in:

(1) perception under different real-world conditions (e.g., using different EPs, as done in this study);
(2) perception under different virtual reality conditions (e.g., using two different rendering algorithms);
(3) perception in a real-world vs. a virtual environment (e.g., to assess haptic fidelity).

The results of all three types of studies can be used to optimize the parameters of virtual environments, as indicated by the dotted lines in Figure 10. Note that the framework is not only applicable to haptic interfaces, but also to interfaces for other modalities, as well as multimodal interfaces.

*MDS Compared to Other Approaches* The MDS approach presented here addresses a growing need for tools which allow for 1) validation of haptic/multimodal displays relative to real-world perception and 2) comparison and benchmarking of different displays, algorithms, and usage patterns. A number of studies have already applied paradigms developed in the field of experimental psychology to the problem of interface validation. Several studies have measured the speed, accuracy, or forces exerted by a human user during a task and tested how similar these are under real-world and virtual conditions, e.g., [Tan et al. 1994; Greenish et al. 2002; Unger et al. 2003]. Magnitude estimation tasks have also been used to characterize the perception of virtual object properties (e.g., roughness), as a function of environment parameters (e.g., probe type [Klatzky et al. 2003; Jansson and Pieraccioli 2004]). Other groups have measured the discriminability of object properties (e.g., curvature) and used this as a metric [Lawrence et al. 1996; Pao and Lawrence 1998; Srinivasan et al. 1999; Webster et al. 2005]. Finally, some studies have begun to use metrics based on performance of more cognitive tasks such as object recognition, categorization, and similarity judgments [Tan et al. 2000; Greenish et al. 2002; Salada et al. 2002]. Nevertheless, validation methods for haptic technologies are in the early stages of development and there is still a need for robust measures which provide insight into complex, cognitive human experience of virtual environments, while at the same time being easy and quick to use.

We suggest that an ideal validation paradigm includes the following attributes:

(1) The paradigm provides *robust statistics*, i.e., a set of statistics and corresponding measures of confidence, which allow differences between real-world and virtual experiences to be quantified and allow for benchmarking of different virtual experiences/systems.

(2) The paradigm offers insight into *cognitive* aspects of virtual experiences, i.e., metrics and/or visualizations that reveal how cognitive-level processes such as learning and coping strategies, meanings, and representations are affected by changes in the haptic environment. This may involve a shift towards higher-level similarity, recognition, categorization, semantic, and memory tasks [Swindells et al. 2005].[4] Cognitive metrics also need to be flexible enough to extend to multimodal interactions.

(3) The paradigm is *easy to use* in a general sense, i.e., the method used to gather data is easy to understand, straightforward to implement on a wide range of systems, and can be carried out quickly. The analysis procedure required to transform raw measures into the desired metrics should also be easy to implement or acquire and quick to carry out.

MDS approaches provide a partial answer: they offer quantitative measures (interstimulus distances, stress as a measure of dimensionality, and dimension weights) and they offer insight into higher-level stimulus representations. One drawback, however, is that pairwise similarity data are time-consuming to gather; different proximity measures such

[4]As noted by [Garbin 1990], attributes that enable objects to be discriminated may not be those which play the most important role in their perceptual representation, although poorly discriminable properties likely do not play an important perceptual role.

as same/different judgements or confusion errors, as well as techniques for analyzing incomplete data sets may offer a solution to this (for an overview of various approaches, see [Borg and Groenen 2005, Chapter 6]). Secondly, data analysis requires several fitting and optimization steps, however standard implementations are available in packages such as MATLAB and SPSS. Third, output dimensions are not labelled; if dimensions do not match hypotheses, further studies must be performed to label them. Despite these drawbacks, the flexibility and generalizability of MDS make it a powerful tool for investigating human perception in both real and virtual environments.

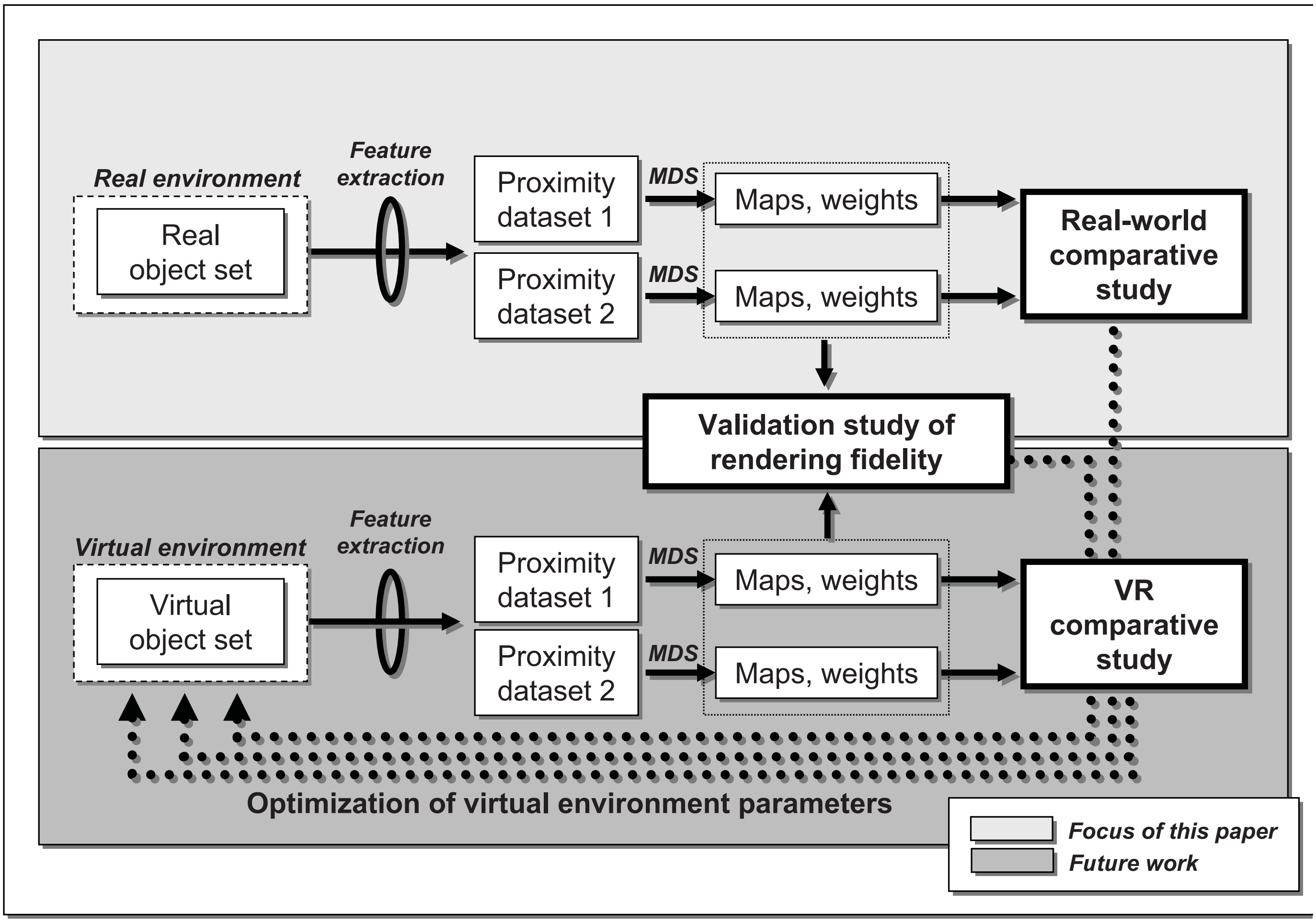


Fig. 10. A framework for validation and comparative studies using MDS. Features are extracted from interactions in real or virtual environments and proximity data are derived from operations on these features. MDS is used to construct perceptual maps and compute relative dimension weights. Results are compared to evaluate perception 1) under different real-world scenarios (e.g., to characterize human perception), 2) under different virtual reality scenarios (e.g., to benchmark different technical systems), and 3) in real-world vs. virtual scenarios (e.g., to validate a particular technical system).

## 8. SUMMARY AND OUTLOOK

### 8.1 Summary of Psychophysical Findings

This paper shows how MDS techniques can be used to obtain rich, quantitative characterization of haptic EPs and their effects on how objects are perceived. The results of

Experiments 1 and 2 showed that subjects were able to extract both shape and texture variations in the stimulus set, but that the perceptual distances between the objects, as well as the relative perceptual weight of shape and texture dimensions, varied systematically according to the EP used. Restricting exploration to lateral motion on the objects' centers led to similarity judgments based solely on texture, as expected. When objects were explored using contour-following, tip-touching, or gripping, subjects used both shape and texture properties to judge similarity. The relative importance of the two properties varied substantially from subject to subject. When the same subjects repeated the experiment several months later in Experiment 2, similarities, maps, and weights remained stable over time. Intersubject variability in the gripping condition was found to correlate with differences in grip style, but further work is needed to identify the source of variability when subjects use tip-touching or contour-following.

## 8.2 Implications for Haptic Interface Design

We conclude by highlighting four aspects of this work which are relevant for haptic interface design.

(1) *Variability Due to Exploratory Procedure* Our results provide a clear demonstration that the way in which we interact with objects can change how we perceive them: changing the EP affected the number of dimensions used to judge similarity as well as interstimulus distances in perceptual space (Figure 3). For haptic engineers, this finding underscores the importance of carefully specifying the way users are expected to interact with the device to ensure that the mode of interaction provides access to task-relevant dimensions and that, when needed, it leads to the desired set of similarity relationships amongst the objects being manipulated by the user.

(2) *High Intersubject Variability* We found high variability in property weights used by different subjects when they used exploratory procedures which allowed for the extraction of more than one object property. Even though subjects were presented with exactly the same objects and interacted with them under careful human supervision, perceptual similarities differed. In this study, lateral motion yielded the most consistent stimulus representation across subjects - one dominated by texture differences - and thus it could be regarded as the "safest" EP to allow in a haptic environment. Of course, it was also highly limited in that it did not allow for shape information to be extracted. On the other hand, the EPs that did allow for shape to be extracted were associated with high between-subject variability in the relative perceptual weight of shape versus texture. For haptic engineers, this finding underscores the need to select modes of interaction with the lowest intersubject variability possible (e.g., by minimizing the number of perceptual dimensions being extracted) while still providing access to all necessary stimulus dimensions. Furthermore, if a single perceptual representation is required, user-specific calibration may be required to compensate for individual differences.

(3) *Low Intrasubject Variability* Property weights for single subjects changed little across the two experiments, even though the tests were separated by several months on average. Thus, although individual subjects weight properties very differently from one another, their weighting appears to remain stable over time. For haptic engineers, this finding suggests that calibration to compensate for individual differences may only need to be repeated at intervals of several months or longer.

(4) *MDS as a Tool for Haptic Interface Design* We have outlined several ways in which haptic engineers might use MDS techniques to test how variations in device parameters or human-device interaction affect high-level, cognitive representations of objects being displayed. MDS can also be used as a validation technique by comparing data gathered in real versus virtual environments, or by having similarity judged on a mixed set of real and virtual objects as done in [Leskovsky et al. 2006]. Given the flexibility of MDS approaches, we believe they offer a valuable tool for haptic researchers investigating human perception in both real and virtual environments.

## Acknowledgements

The authors wish to thank Karin Bierig and Bianca Arsene for technical assistance in carrying out the experiments, Steffen Baier and the staff of the workshop at the Max Planck Institute for Biological Cybernetics for helping to build the haptic setup, Martin Breidt and Mirko Thiesen for assistance in stimulus design and production, and Marc Ernst for helpful feedback on the preliminary study for this paper, which was presented at Eurohaptics 2006.

## REFERENCES

BARSALOU, L. 1999. Perceptual symbol systems. *Behav Brain Sci 22*, 577–660.

BERGMANN, T., WOUTER, M., AND KAPPERS, A. 2006. Analysis of haptic perception of materials by multidimensional scaling and physical measurements of roughness and compressibility. *Acta Psychol (Amst) 121,* 1, 1–20.

BIEDERMAN, I. 1987. Recognition-by-components: a theory of human image understanding. *Psychol Rev 94,* 2 (Apr), 115–147.

BORG, I. AND GROENEN, P. 2005. *Modern multidimensional scaling*, 2nd ed. Springer.

BÜLTHOFF, H. AND EDELMAN, S. 1992. Psychophysical support for a 2-d view interpolation theory of object recognition. *Proceedings of the National Academy of Science 89*, 60–64.

CARROLL, J. AND CHANG, J. 1970. Analysis of individual differences in multidimensional scaling via an n-way generalization of "Eckart-Young" decomposition. *Psychometrika 35*, 283–319.

CARROLL, J. AND GREEN, P. 1997. Psychometric methods in marketing research: Part ii, multidimensional scaling. *Journal of Marketing Research 34,* 2, 193–204.

CHEN, C. 2003. *Mapping Scientific Frontiers: The Quest for Knowledge Visualization*. Springer Verlag.

COOKE, T., JÄKEL, F., WALLRAVEN, C., AND BÜLTHOFF, H. 2006. Multimodal similarity and categorization of novel, three-dimensional objects. *Neuropsychologia in press*, –.

COOKE, T., STEINKE, F., WALLRAVEN, C., AND BÜLTHOFF, H. 2005. A similarity-based approach to perceptual feature validation. In *APGV 2005 - Symposium on Applied Perception in Graphics and Visualization*. ACM SIGGRAPH.

COOKE, T., WALLRAVEN, C., AND BLTHOFF, H. 2005. A comparison of visual and haptic object representations based on similarity. In Proceedings of the Second International Symposium on Non-Visual & Multimodal Visualization. *Proceedings of the Second International Symposium on Non-Visual & Multimodal Visualization*.

COX, T. AND COX, M. 2001. *Multidimensional Scaling*, 2nd ed. Chapman & Hall.

DOSTMOHAMED, H. AND HAYWARD, V. 2005. Trajectory of contact region on the fingerpad gives the illusion of haptic shape. *Exp Brain Res 164,* 3, 387–394.

EDELMAN, S. 1999. *Representation and recognition in vision*. MIT Press.

ERNST, M. AND BÜLTHOFF, H. 2004. Merging the senses into a robust percept. *Trends Cogn Sci 8,* 4, 162–169.

GARBIN, C. 1990. Visual-touch perceptual equivalence for shape information in children and adults. *Percept Psychophys 48,* 3, 271–279.

GARBIN, C. AND BERNSTEIN, I. 1984. Visual and haptic perception of three-dimensional solid forms. *Percept Psychophys 36,* 2, 104–110.

GREENISH, S., HAYWARD, V., CHIAL, V., OKAMURA, A., AND STEFFEN, T. 2002. Measurement, analysis and display of haptic signals during surgical cutting. *Presence 11,* 6, 626–651.

HOLLINS, M., FALDOWSKI, R., RAO, S., AND YOUNG, F. 1993. Perceptual dimensions of tactile surface texture: A multidimensional scaling analysis. *Percept Psychophys 54,* 6, 687–705.

JANSSON, G. AND PIERACCIOLI, C. 2004. Effects of surface properties on haptic perception of the form of virtual objects. In *Proceedings of the 4th International Conference Eurohaptics 2004*, M. Buss and M. Fritschi, Eds. München, Germany, 211–216.

KLATZKY, R. AND LEDERMAN, S. 2003. *Experimental psychology*. Handbook of Psychology, vol. 4. Healy, AF and Proctor, RW and Weiner, IB (eds.), Wiley, Chapter Touch, 147–176.

KLATZKY, R., LEDERMAN, S., HAMILTON, C., GRINDLEY, M., AND SWENDSEN, R. 2003. Feeling textures through a probe: effects of probe and surface geometry and exploratory factors. *Percept Psychophys 65,* 4, 613–631.

KLATZKY, R., LEDERMAN, S., AND MATULA, D. 1993. Haptic exploration in the presence of vision. *J Exp Psychol Hum Percept Perform 19,* 4, 726–743.

KLATZKY, R., LEDERMAN, S., AND REED, C. 1987. There's more to touch than meets the eye: The salience of object attributes for haptics with and without vision. *Experimental Psychology: General 116,* 4, 356–369.

LAKATOS, S. AND MARKS, L. 1999. Haptic form perception: relative salience of local and global features. *Percept Psychophys 61,* 5, 895–908.

LAWRENCE, D., PAO, L., SALADA, M., AND DOUGHERTY, A. 1996. Quantitative experimental analysis of transparency and stability in haptic interfaces. In *Proc. Fifth Annual Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*. Atlanta, GA.

LEDERMAN, S. AND KLATZKY, R. 1987. Hand movements: a window into haptic object recognition. *Cognit Psychol 19,* 3, 342–368.

LEDERMAN, S. AND KLATZKY, R. 1993. Extracting object properties through haptic exploration. *Acta Psychol 84*, 29–40.

LEDERMAN, S., SUMMERS, C., AND KLATZKY, R. 1996. Cognitive salience of haptic object properties: role of modality-encoding bias. *Perception 25,* 8, 983–998.

LESKOVSKY, P., COOKE, T., ERNST, M., AND HARDERS, M. 2006. Using multidimensional scaling to quantify the fidelity of haptic rendering of deformable objects. *Proceedings of EUROHAPTICS 2006*, 7.

MENG, X., ROSENTHAL, R., AND DB, R. 1992. Comparing correlated correlation coefficients. *Psychol Bull 111,* 1, 172–175.

NEWELL, F., ERNST, M., TJAN, B., AND BÜLTHOFF, H. 2001. Viewpoint dependence in visual and haptic object recognition. *Psychol Sci 12,* 1, 37–42.

O'REGAN, J. AND NOE, A. 2001. A sensorimotor account of vision and visual consciousness. *Behav Brain Sci 24,* 5, 939–973.

PAO, L. AND LAWRENCE, D. 1998. Synergistic visual/haptic computer interfaces. In *Proc. Japan/USA/Vietnam Workshop on Research and Education in Systems, Computation, and Control Engineering*. Hanoi, Vietnam, 155–162.

PONT, S., KAPPERS, A., AND KOENDERINK, J. 1999. Similar mechanisms underlie curvature comparison by static and dynamic touch. *Percept Psychophys 61,* 5, 874–894.

RILEY, M., WAGMAN, J., SANTANA, M., CARELLO, C., AND TURVEY, M. 2002. Perceptual behaviour: recurrence analysis of an exploratory procedure. *Perception 31*, 481–510.

SALADA, M., COLGATE, J., LEE, M., AND VISHTON, P. 2002. Validating a novel approach to rendering fingertip contact sensations. In *Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*. 217–224.

SHEPARD, R. AND CERMAK, G. 1973. Perceptual-cognitive explorations of a toroidal set of free-form stimuli. *Cognit Psychol 4*, 351–377.

SIMONS, D. 1996. In sight, out of mind: When object representations fail. *Psychol Sci 7,* 5, 301–305.

SRINIVASAN, M., BASDOGAN, C., AND WU, W. 1999. Visual, haptic, and bimodal perception of size and stiffness in virtual environments. In *Proceedings of the ASME Dynamic Systems and Control Division*.

SRINIVASAN, M. AND LAMOTTE, R. 1991. *Encoding of shape in the responses of cutaneous mechanoreceptors*. Wenner-Gren International Symposium Series. MacMillan Press, Chapter 5, 59–61.

SWINDELLS, C., SMITH, J., AND MACLEAN, K. 2005. An exploration of representations to aid design of haptic behaviours. In *Proceedings of Conference on Human Factors in Computing Systems (CHI) Workshop – Hands-on Haptics: Exploring Non-Visual Visualization Using the Sense of Touch.* ACM Press, 5–8.

TAN, H., LIM, A., AND TRAYLOR, R. 2000. A psychophysical study of sensory saltation with an open response paradigm. In *Proceedings of the Ninth (9th) International Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*. American Society of Mechanical Engineers Dynamic Systems and Control Divisio, Orlando, FL, 1109–1115.

TAN, H., SRINIVASAN, M., EBERMAN, B., AND CHENG, B. 1994. Human factors for the design of force-reflecting haptic interfaces. In *Proceedings of the ASME Dynamic Systems and Control Division*. 353–359.

UNGER, B., KLATZKY, R., AND HOLLIS, R. 2003. A telemanipulation system for psychophysical investigation of haptic interaction. In *ICRA*. 1253–1258.

WEBSTER, R., MURPHY, T., VERNER, L., AND OKAMURA, A. 2005. A novel two-dimensional tactile slip display: design, kinematics and perceptual experiments. *TAP 2,* 2, 150–165.

WIPPICH, W., MECKLENBRAUKER, S., AND NORBERT-WURM, J. 1994. Motor and sensory effects of haptic experiences in implicit and explicit memory tests. *Z Exp Angew Psychol 41,* 3, 500–522.

YOUNG, F. AND HARRIS, D. 2003. ALSCAL. In *SPSS 12.0 Command Syntax Reference*. SPSS, Chicago, IL, 100–116.

# Object feature validation using visual and haptic similarity ratings

Theresa Cooke, Sebastian Kannengiesser, Christian Wallraven, and Heinrich H. Bülthoff
Max Planck Institute for Biological Cybernetics, Tübingen, Germany

The perceived similarity between objects may well vary according to the sensory modality/modalities in which they are experienced, an important consideration for the design of multimodal interfaces. In this study, we present a similarity-based method for comparing the perceptual importance of object properties in touch and in vision and show how the method can also be used to validate computational measures of object properties. Using either vision or touch, human subjects judged the similarity between novel, 3D objects which varied parametrically in shape and texture. Similarities were also computed using a set of state-of-the art 2D and 3D computational measures. Two resolutions of 2D and 3D object data were used for these computations in order to test for scale dependencies. Multidimensional scaling (MDS) was then performed on all similarity data, yielding maps of the stimuli in both perceptual and computational spaces, as well as the relative weight of shape and texture dimensions. For this object set, we found that visual subjects accorded more importance to shape than texture, while haptic subjects weighted them roughly evenly. Fit errors between human and computational maps were then calculated to assess each feature's perceptual validity. Shape-biased features provided good overall fits to the human visual data; however, no single feature yielded a good overall fit to the haptic data, in which we observed large individual differences. This work demonstrates how MDS techniques can be used to evaluate computational object features using the criterion of perceptual similarity. It also demonstrates a way of assessing how the perceptual validity of a feature varies as a function of parameters such as the target modality and the resolution of object data. Potential applications of this method for the design of unimodal and multimodal human-machine interfaces are discussed.

Categories and Subject Descriptors: I.4.7 [**Image Processing and Computer Vision**]: Feature Measurement—Feature representation, size and shape, texture; H.5.1 [**Information Interfaces and Presentation**]: Multimedia Information Systems—Artificial, augmented and virtual systems, evaluation/methodology; H.5.2 [**Information Interfaces and Presentation**]: User Interfaces—Haptic I/O, evaluation/methodology

General Terms: Experimentation, Human Factors, Measurement

Additional Key Words and Phrases: similarity, multidimensional scaling, perception, vision, touch, haptic, features, validation, shape, texture

## 1. INTRODUCTION

The design of effective and efficient multimodal displays requires an understanding of how humans make use of their different senses to build up representations of their surroundings. Models of human visual object processing have proposed that the visual system extracts object features or properties from images projected onto the retina [Marr 1982]. These fea-

e-mail: theresa.cooke@tuebingen.mpg.de

Permission to make digital/hard copy of all or part of this material without fee for personal or classroom use provided that the copies are not made or distributed for profit or commercial advantage, the ACM copyright/server notice, the title of the publication, and its date appear, and notice is given that copying is by permission of the ACM, Inc. To copy otherwise, to republish, to post on servers, or to redistribute to lists requires prior specific permission and/or a fee.
© 20YY ACM 0000-0000/20YY/0000-0001 $5.00

tures are then used as the basis for representing objects in the brain [Bülthoff and Edelman 1992; Ullman 1996]. Inspired by this, a similar approach has been taken in the field of computer vision: a set of computational measures are extracted from 2D images of objects (or scenes) and used to create artificial representations of objects for automated reconstruction, recognition, or categorization tasks [Riesenhuber and Poggio 1999; Ullman et al. 2002]. Work in this field has given rise to a large number of feature extraction algorithms, including biologically-inspired filters (Gabor jets) which mimic the response of cells in visual cortex [Jones and Palmer 1987] and algorithms derived from statistical optimization procedures [Ankerst et al. 1999]. These computational measures have been evaluated in a variety of ways, e.g., based on their performance in machine vision tasks. Biological plausibility has mainly been assessed at a relatively low level (e.g., by matching receptive field properties). In this paper, we propose a new method for validating computational measures based on the *high-level, cognitive criterion of object similarity*. Similarity is thought to underlie a number of cognitive processes, including both categorization [Rosch et al. 1976] and recognition [Edelman 1999]. The similarity-based criterion we propose is as follows: a good feature is one which, for a set of parametrically-defined objects, generates similarity-based stimulus configurations akin (in one or more respects) to those derived from human similarity ratings.

Most perceptual validation of computational object features has been carried out in relation to *visual* perception. However, a feature's perceptual validity may well vary as a function of sensory modality used to perceive the objects, e.g., [Klatzky et al. 1987]. The method presented in this paper provides a solution to this problem by enabling validation to be performed relative to any sensory modality. For the haptic modality, measures computed on 3D objects are particularly interesting, e.g., [Nefs and Kappers 2003], and a large number of such measures have been proposed in the 3D graphics literature [Funkhouser et al. 2003]. However, there have been few studies which have assessed these measures relative to the haptic modality using a high-level, cognitive criterion such as similarity. Knowledge of which 3D computational measures correlate with high-level human stimulus representations derived from haptic perception would not only help in the design of more realistic artificial haptic systems (for example, [Acosta et al. 2002]) and reduce the heavy demands of haptic rendering [Salisbury et al. 2004], but could also play an important role in elucidating the computational mechanisms of the human haptic system.

Our method can be situated in the context of a framework connecting the development of artificial representational systems and advances in our understanding of human representational systems (Figure 1). Physical objects constitute the input to both types of systems, which use various sensors to measure object properties (photoreceptors, mechanoreceptors, etc.). For artificial systems, the way these properties are extracted depends on the sensor and the computational algorithm applied to the measured quantities. For humans, it is a function of sensory modality. In both human and artificial systems, the extracted properties can then be used (either directly or indirectly) to embed objects in a *representational space* or "map." With the appropriate tools, these representational spaces can be compared at either the *unimodal* or at the *multimodal* level. Comparing a map derived from human unimodal perception (e.g., from pure visual exposure to the objects) to a map derived using a computational measure (e.g., pixel-wise differences between images) allows for *unimodal validation* of computational measures. Two human unimodal maps can also be compared to identify modality-dependent differences in human object processing. The

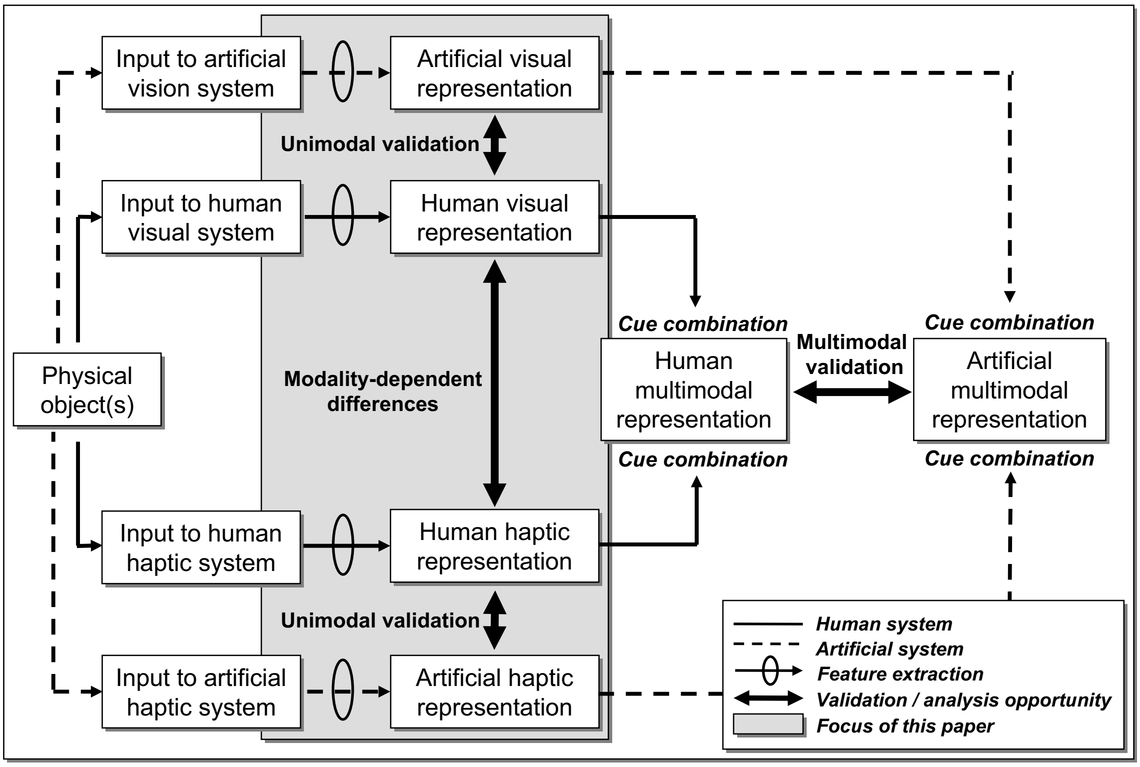


Fig. 1. A framework for studying human and artificial uni/multimodal object representations.

same approach can be applied at the *multimodal level* to test hypotheses about human cue combination and to validate approaches to artificial cue combination (e.g., in the design of visuotactile interfaces for telemedecine).

The method presented in this paper connects perceptual and artificial systems at the level of *unimodal* representations. We derive maps of our stimuli based on human visual and haptic similarity measures, and from similarity measures using a set of computational methods which we wish to perceptually validate. We first show how our method can be used to compare *human* haptic and visual stimulus maps. Then, we demonstrate how the method can be used to evaluate the perceptual validity of the computational measures by comparing the human maps against those derived from the computational measures.

## 2. METHODS

### 2.1 Stimuli

The stimuli consist of a family of novel, 3D objects (Figure 2), created in the graphics package 3D Studio Max 6.0. This software provides full control of object properties such as size, shape, and texture, allowing them to be varied in defined steps. The family begins with a family "prototype" (see Figure 2, object 1), which consists of: 1) three parts connected to a center sphere, defining the object's macrogeometry and 2) a displacement map applied to the 3D mesh, specifying the object's microgeometry. The other family members are generated by two manipulations. The first manipulation increases the smoothness of the object's microgeometry (or "texture") by decreasing the amount of mesh displacement caused by the displacement map. The second manipulation increases the smoothness of the object's macrogeometry (or "shape") by moving mesh vertices towards a local average, removing sharp angles in the global shape. Note that from a haptic rendering perspective, two distinct sets of force-rendering algorithms would be needed to convey these two kinds of variations: geometric-dependent rendering algorithms for shape variations and surface property-dependent rendering algorithms for texture variations [Salisbury et al. 2004].

The displacement map applied to the objects consisted of triangular-shaped elements with a base width of 3mm, a peak width of 2mm, and a maximum height of 3mm from the surface of the object; texture elements were spaced 3-5 mm apart. The scale of this pattern qualifies it as a macrotexture, the properties of which are known to be encoded by SAI mechanoreceptors [Klatzky and Lederman 2003]. The displacement manipulation simultaneously reduced element height and increased the peak width of elements; inter-element spacing remained constant. Increasing element width has been shown to decrease perceived roughness [Sathian et al. 1989]. In previous work with this stimulus set [Cooke et al. 2005; Cooke et al. 2006], subjects consistently described the objects as varying in "texture," and often referred to their "roughness" or "bumpiness."

The macrogeometrical manipulation applied to the objects averages out sharp angles in the mesh. In previous studies, subjects consistently referred to this manipulation as a change of "shape." Since the sharpest angles are located at the objects' extremities (maximum surface area of roughly 1 $cm^2$), the mesh relaxation affects these areas much more than the rest of the object (and thus has a more localized character than the evenly-distributed texture manipulation). It has been shown that SAI mechanoreceptors from a single finger are capable of curvature estimates for angles which fall onto the same region of the skin [Srinivasan and LaMotte 1991]. In addition to static cues, changing local curvature gradients created during object exploration can also provide macrogeometrical information [Pont et al. 1999]. Finally, kinesthetic cues to macrogeometry are provided by systematic changes in finger joint and wrist angles during object exploration.

Objects created using these variations can be plotted in a 2D space whose dimensions correspond to microgeometry and macrogeometry (Figure 2). The 3D models were printed out (Dimension 3D Printer, Stratasys, Minneapolis, USA) into hard, white, and opaque objects, measuring 9.0 +/- 0.1 cm wide, 8.3 +/- 0.2 cm high, and 3.7 +/- 0.1 cm deep and weighing about 40 g.

### 2.2 Visual similarity ratings

Ten subjects with normal or corrected-to-normal vision were paid 8 Euros per hour to rate the similarities between photographs of the objects presented at 75 Hz on a Sony Trinitron 21" monitor with a resolution of 1024 x 768 pixels. Photographs of the objects were displayed using the Psychtoolbox extension for MATLAB [Brainard 1997] on a Macintosh G4 computer. The photographs were taken such that the three object parts were aligned with the image plane (referred to as "frontal view"). This viewpoint was chosen in order to provide the best possible match to the viewpoint presented in the haptic task (see below). The image size was 7.6 x 7.6 degrees of visual angle (set to be approximately the same size as if the object were being held at arm's length). Subjects had never seen or touched the objects before. They were seated approximately 60 cm from the monitor in a dimly-lit room. A fixation cross appeared for 500 ms and then each of the objects appeared for 500 ms, separated by a 500 ms interstimulus interval. At the end of each trial, subjects had to rate the similarity of the objects on a scale between one (low similarity) and seven (high similarity). A set of practice trials enabled the subjects to become familiar with the task. Response time was unlimited. There were six experimental blocks of 325 randomized trials (each object was compared once with itself and once with every other object, yielding 25

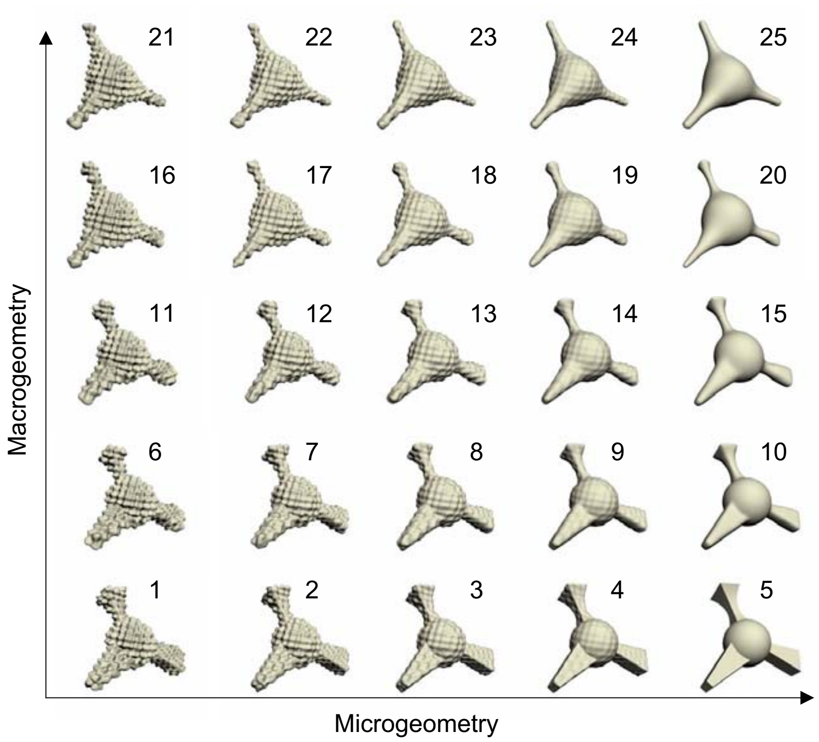


Fig. 2. Stimuli varied parametrically in terms of microgeometry (texture) and macrogeometry (shape).

+ (25·24)/2 = 325 trials.) The order of appearance of stimuli in each pair was randomized over the blocks. The total experiment lasted about two hours. At the end of the experiment, subjects were asked to write a short description of how they had judged similarity amongst the objects.

## 2.3 Haptic similarity ratings

Ten right-handed subjects were paid 8 Euros per hour to rate the similarities between the objects after exploring them haptically. None of the subjects had participated in the visual experiment, or had seen or touched the stimuli before. Subjects sat in front of a table, facing an opaque curtain through which they placed their right hand. They were instructed to keep their eyes closed during the experiment (subjects in a pilot study had reported that they were better able to concentrate on the task with their eyes closed). Behind the curtain, the experimenter presented two objects, one after the other. The objects were always presented in the same fixed position, face up on the table. Subjects were given up to ten seconds to trace the contour of each object, after which they rated the similarity between the objects on a scale from one (low similarity) to seven (high similarity). The contour-following procedure was chosen because it has been shown to allow for haptic extraction of a wide range of object properties, including local texture and global shape [Lederman and Klatzky 1993]. In the ten seconds provided, even untrained subjects had sufficient time to trace the object's contour twice. A set of practice trials allowed the subjects to become familiar with the task. The full experiment consisted of three blocks of 325 randomized trials spread out over five two-hour sessions on consecutive days. The order of appearance of stimuli in each pair was randomized over blocks. At the end of the experiment, subjects were asked to write a short description of how they had judged similarity among the objects.

## 2.4 Computational similarity measures

We implemented nine computational similarity measures: five operating on 2D photographs of the objects and four operating on the objects' 3D mesh geometry. The photographs used to compute similarity values were the same images presented in the human visual similarity experiment.

Given the basic 2D nature of visual input and the 3D nature of haptic input, we hypothesized that 2D measures could provide better fits to maps derived from visual similarities, while 3D measures could provide better fits to maps derived from haptic similarities. Human performance in visual object recognition tasks has been successfully explained by 2D view-based models [Bülthoff and Edelman 1992]. At the same time, however, the human visual system is capable of extracting and using 3D object properties from 2D images, such as in the case of shape-from-shading [Blake and Bülthoff 1991] and shape-from-texture [Todd et al. 2004]; thus we were also compared object maps based on human vision against maps generated from 3D features.

Fewer studies have examined the question of feature dimensionality for haptic object representations. A number of neuroimaging and psychophysical studies have emphasized the role of the visual cortex in the mediation of high-level tactile object representations (e.g., [Deibert et al. 1999]), while other studies have argued for more independent representations [Reed et al. 2004]. Our approach to this issue is to test whether 2D features correlate equally well with vision and haptics, or whether 3D features provide better correlations with haptics than with vision, which would support the idea of greater representational independence. In order to investigate these questions, object representations derived from human visual and haptic similarity ratings were compared against maps derived from the following 2D and 3D computational measures:

(1) *2D image subtraction* A simple pixel-wise subtraction between RGB images of two objects was performed. We took the mean absolute difference over all pixels and RGB channels as the dissimilarity between the two objects.

(2) *2D edge detection* Canny edge detection was performed on each image using MATLAB's edge detection algorithm, resulting in binary edge images (pixel value set to 1 if the pixel location is found to belong to an edge and 0 if not). The mean pixel-wise difference between two edge images was taken as the dissimilarity. Edge detection has been suggested as an important step in models of human vision [Marr 1982].

(3) *2D Gabor jets* The images were filtered with Gabor jets [Nestares et al. 1998] and the pixel-wise difference in the filter response images was computed. The Gabor jet filter has been proposed as a biologically-plausible model for receptive fields in early visual cortex [Jones and Palmer 1987] and has recently been successfully applied in models of object and motion recognition [Giese 2004]. Here, we used a variant which applies Gabor filtering in four orientations using both even and odd channels. A response image was generated for each orientation by summing and squaring the even and odd responses. Pixel-wise difference images were computed at each orientation and the final dissimilarity between two images was computed as the mean over all pixels and all orientations.

(4) *2D Visual Difference Predictor* In addition to the first three relatively straightforward 2D measures, we generated similarity data using two more elaborate measures: the Visual Difference Predictor (VDP) and the Structural Similarity (SSIM) measure. These

are industry standards for computing image differences and thus serve as a benchmarks for comparing the performance of the other 2D measures [Cadik and Slavik 2004]. The VDP [Daly 1993; Mantiuk et al. 2005] incorporates a model of low-level human visual processing, including the visual system's non-linear adaptive response to light, its contrast sensitivity function, and a masking function which models variations in sensitivity related to image content. As our computed measure, we took the total number of pixels which the VDP detected as different in the two images with a probability of at least 95%.[1]

(5) *2D Structural Similarity* Like the VDP, the SSIM [Wang et al. 2004] takes properties of the human visual system into account and computes an index of structural difference between two images after removing differences in average luminance and contrast. [2]

(6) *3D subtraction* Mean Euclidean distance between all 3D vertex coordinates of two object meshes was computed; point-by-point subtraction was possible because the meshes were in correspondence. This measure is the 3D equivalent of our 2D image subtraction measure.

(7) *3D perimeter* Object perimeter was measured along a cross-section taken parallel to the frontal view and the difference was taken for each pair of objects. Although perimeter can be measured in 2D or 3D space, we refer to it here as a 3D measure because of the special role it could play in haptic feature extraction, particularly given that in this experiment, participants explored the objects by following their contours.

(8) *3D curvature* A stable, reliable curvature estimate was obtained by fitting an implicit surface representation to the object and extracting curvatures from it [Steinke et al. 2005]. To estimate an object's "bumpiness," the absolute value of the mean curvature was averaged over the whole surface.

(9) *3D shape* A measure based on 3D shape histograms was implemented, inspired by [Ankerst et al. 1999]. 3D space is partitioned in the radial direction (into shells), which are further subdivided to create bounded sectors (the bound of the outermost shell is determined by the size of the largest object). For each object, we counted the number of vertices populating each sector, thereby creating a 3D shape histogram for that object. As a dissimilarity measure, we took the mean absolute sector-wise difference in vertex count between two object meshes.

*Scale variation* To demonstrate how our method can be used to quantify the effects of changing the parameters used for computing object features, we chose to vary the resolution of object data. Selecting the scale of resolution at which to compute features is an important problem in computer vision [Lindeberg 1994]; the size of stimuli relative to sensory receptors is also a fundamental issue for both visual and haptic perception [Koenderink 1984; Klatzky and Lederman 2003]. For 2D images, data at the finer scale of resolution were 600x600 pixel images (presented in the visual similarity experiment) and data at the coarser scale consisted of the same images downsampled to 38x38. For 3D object data, meshes at the finer scale consisted of the 4728 vertices constituting the original

[1]The VDP can be seen either as a single computational measure or as a combination of several measures; in the latter case, our evaluation of the VDP can be considered to be an evaluation of this particular combination of measures.
[2]A MATLAB implementation of the SSIM was downloaded from http://www.cns.nyu.edu/~lcv/ssim/

models and data at the coarser scale consisted of the same models downsampled to 296 vertices, which remained in correspondence.

### 2.5 MDS analysis of similarity data

Similarity data were analyzed using a multidimensional scaling (MDS) technique. Performing multidimensional scaling analysis of similarity ratings (or other kinds of proximity data) is a standard approach used in cognitive psychology to explore the psychological structure in a data set [Borg and Groenen 1997]. Human and computational similarity data were analyzed using a non-metric MDS algorithm implemented in MATLAB. Non-metric MDS uses the *ranks* of the pair-wise distances as input, as opposed to their precise values. Because of this, the relationship between the similarity data and the distances in the output configuration may be non-linear. The algorithm returns the stress value (Kruskal's stress formula 1), which is used to determine the appropriate dimensionality of the output configuration. Stress values below 0.2 are generally accepted as an indication that the dimensionality of the output space is sufficient to faithfully represent the input distance information [Cox and Cox 2001]. We also calculated the proportion of variance in the similarity data accounted for by the output configuration of a given dimensionality, which we refer to as RSQ. The optimal number of dimensions needed to represent the objects can be determined by looking for either a sharp drop in the stress plot or a plateau in the RSQ plot. Here, we used the RSQ to estimate the perceptual importance of each dimension in the output maps: the RSQ for the 1D solution was taken as the weight for the first dimension and the additional increase in RSQ for the 2D solution was taken as the weight for the second dimension. MDS also returns the coordinates of each object in the output space (though the scaling and rotation of the configuration is not determined). The non-metric MDS technique we used does not provide an interpretation of the dimensions: they must be interpreted by visual inspection of the output configuration.

### 2.6 Validation of computational measures

To assess the perceptual validity of the computational measures, stimulus maps derived from these measures using MDS were fit to the stimulus maps derived from human similarity ratings. Errors in these fits were used to quantify the correspondence between the computational measure and human perception. Map fitting was performed using the Procrustes function in MATLAB. This function determines a linear transformation (translation, reflection, orthogonal rotation, and symmetric scaling) of the points in a matrix Y which minimizes the sum of squared distances to points in a second matrix X, i.e., it computes

$$\min_{b,T,c} \{\|Z - X\| : Z = bYT + c\}$$

where b is a scaling factor, T is an orthogonal rotation and reflection matrix, and c is a translation component. The returned minimum value is normalized by the scale of X which makes it possible to express the fit error as a percentage value and compare it across data sets with different scales.

## 3. RESULTS AND DISCUSSION

### 3.1 Visual similarity ratings

*Similarity data* Mean visual similarity ratings for the twenty-five objects are shown in Figure 3 (left). Similarity between pairs is distinctly higher when both objects come from either the set 1-15 or the set 16-21 than when one object comes from the set 1-15 and one comes from the set 16-21. Within these two sets, similarity varies according to texture level (e.g., decreasing similarity between object 1 and objects 2, 3, 4, and 5).

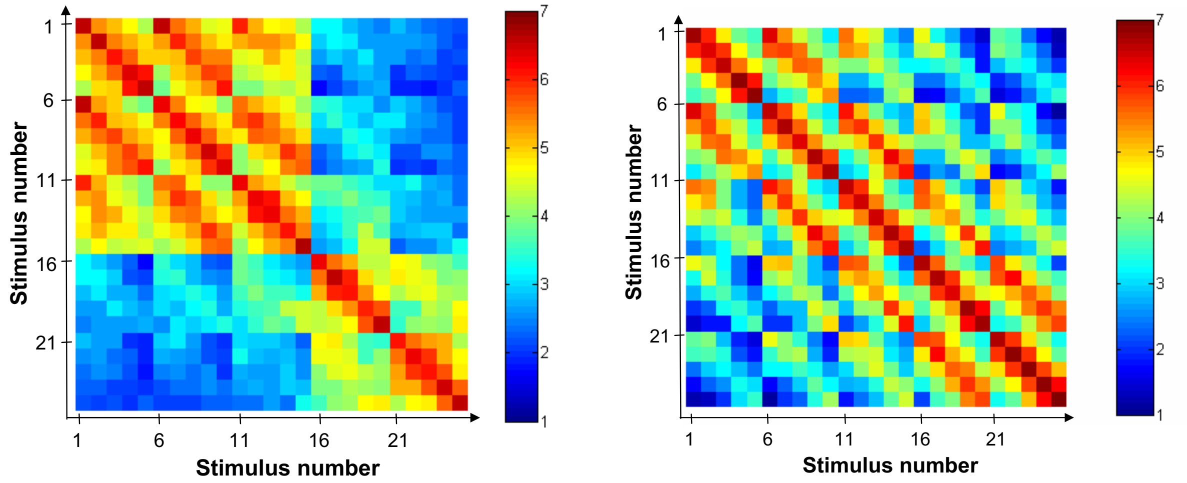


Fig. 3. Mean human visual similarity ratings (left) and mean human haptic similarity ratings (right).

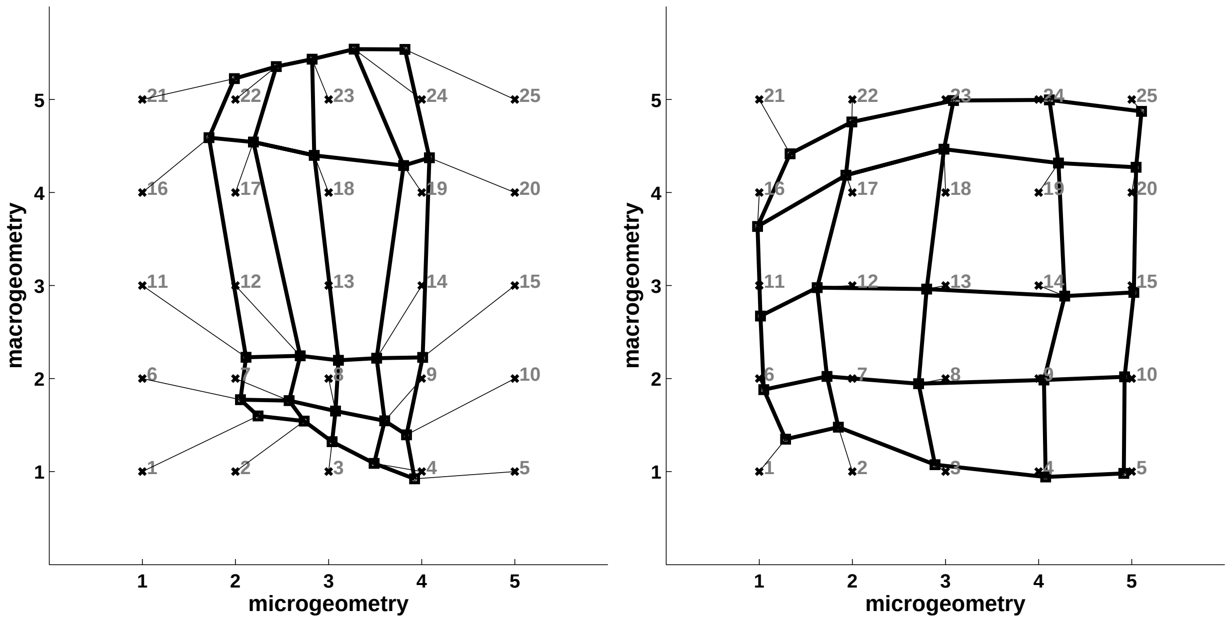


Fig. 4. Perceptual stimulus maps based on mean human visual similarity ratings (left) and mean human haptic similarity ratings (right).

*MDS analysis* Performing MDS allows these patterns in the similarity matrices to be more intuitively visualized as distances in a perceptual space. In order to determine the appropriate dimensionality of the output space, one needs to consider the stress value of the corresponding MDS solution. For mean visual similarity ratings, the stress for a one-dimensional solution was 0.14, indicating that one perceptual dimension is sufficient.

Dimension labels were interpreted by visual inspection of the output configuration (Figure 4, left): the map's most important dimension of variation corresponds to shape, while the second dimension corresponds to texture. Despite the high dimensionality of the visual measurement space, subjects were on average able to recover this low-dimensional variation in the stimulus set (see General Discussion).

In addition to the stress values, the dominance of shape can be seen from the RSQ weights for individual subjects (Figure 5, left). The mean shape weight across subjects was 0.85 (std. err. = 0.03), while the mean weight of the second dimension was 0.06 (std. err. = 0.01). Using a two-tailed t-test for independent samples with equal variances, the mean shape weight was found to be significantly different from the mean texture weight ($t(18)=23.6$, $p<0.01$). The greater importance of shape was also reflected in subjects' descriptions of how they judged similarity: 9 out of 10 subjects mentioned the word "shape" or global shape properties (e.g., geometric descriptions of parts), while 6 out of 10 subjects mentioned the word "texture" or texture-related properties (e.g., bumpiness).

Despite the dominance of shape, most subjects *were* able to recover the structure of the stimulus set along the texture dimension. Another interesting feature of the map is the emergence of two stimulus clusters along the shape dimension, suggesting a connection between similarity judgments and category structure (see General Discussion).

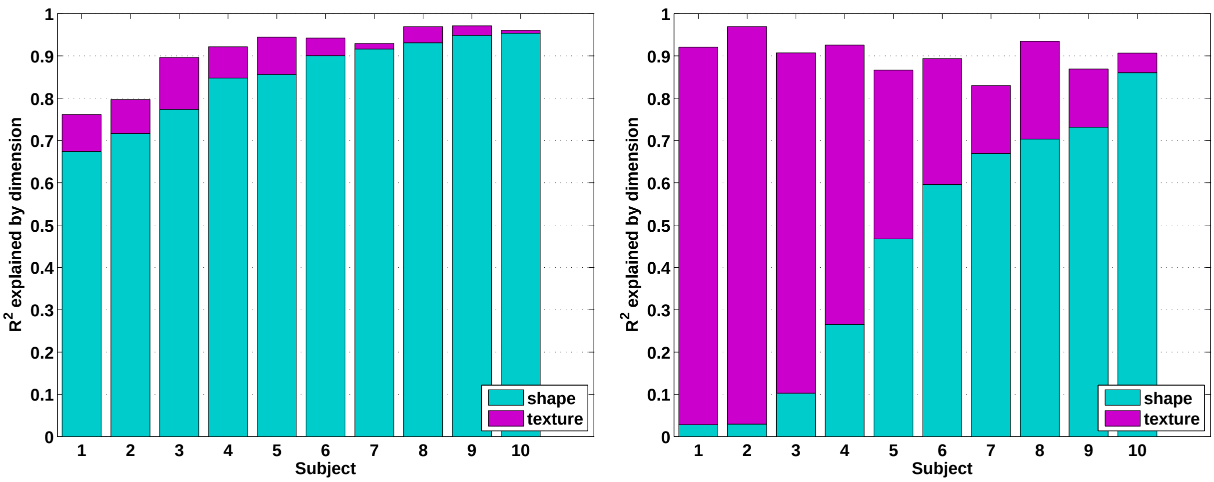


Fig. 5. Dimension weights for subjects in the visual (left) and haptic (right) similarity ratings experiments.

## 3.2 Haptic similarity ratings

*Similarity data* In contrast to the visual data, there is no sharp change in similarity visible in the matrix (3, right). Rather, similarity decreases relatively smoothly with shape change (e.g., stimulus 1 compared to 6, 11, 16, and 21). Similarity also varies smoothly as a function of texture change, even when two very different shapes are compared (e.g., object 1 compared to 21-25).

*MDS analysis* When subjects provided similarity ratings after touching the stimuli, MDS stress computed over mean similarity data was 0.37 for a one-dimensional solution, indicating that a single dimension is *insufficient* to explain the data. Stress dropped to 0.1 for a two-dimensional solution, indicating two dimensions are sufficient. Plotting the output

| Similarity Measure | 1D Stress | 2D Stress |
|---|---|---|
| 2D Subtraction | 0.09 | 0.03 |
| 2D Edge Detection | **0.34** | **0.21** |
| 2D Gabor Jet | 0.17 | 0.10 |
| 2D VDP | 0.12 | 0.03 |
| 2D SSIM | 0.13 | 0.04 |
| 3D Subtraction | 0.10 | 0.05 |
| 3D Shape Histogram | 0.10 | 0.04 |
| 3D Perimeter | 0 | 0 |
| 3D Curvature | 0 | 0 |

Table I. Stress for features computed on finer scale object data for 1D and 2D MDS solutions. Values $> 0.2$ (bolded) indicate that the dimensionality of the output configuration is insufficient to explain similarity data.

| Similarity Measure | 1D Stress | 2D Stress |
|---|---|---|
| 2D Subtraction | 0.09 | 0.03 |
| 2D Edge Detection | 0.15 | 0.08 |
| 2D Gabor Jet | 0.11 | 0.04 |
| 2D VDP | 0.07 | 0.04 |
| 2D SSIM | 0.10 | 0.03 |
| 3D Subtraction | 0.10 | 0.04 |
| 3D Shape Histogram | 0.08 | 0.05 |
| 3D Perimeter | 0 | 0 |
| 3D Curvature | 0 | 0 |

Table II. Stress for features computed on *coarser* scale object data for 1D and 2D MDS solutions.

stimulus configuration (Figure 4, right) enabled us to interpret these perceptual dimensions as texture and shape. Ordinal relationships in the stimulus set were recovered along both dimensions - a remarkable feat given the complexity of the haptic measurement space.

On average, shape and texture played equal roles in haptic similarity judgments. The mean shape weight across subjects was 0.45 (std. err. = 0.1) and the mean texture weight was also 0.45 (std. err. = 0.1). Using a two-tailed t-test for independent samples with equal variances, the mean shape weight was not significantly different from the mean texture weight (t(18)=-0.1, p=0.94). This agrees with the fact that *all* subjects in this experiment mentioned *both* shape-related and texture-related properties when explaining how they had made their similarity judgments. There was, however, a large amount of variation in the way individual subjects weighted shape and texture, from texture dominance, to rough equality between shape and texture, to shape dominance. This finding makes it particularly interesting to try fitting haptic and computational stimulus maps to ascertain whether the variation can be explained by specific computational mechanisms.

### 3.3 Computational similarity measures

In this section, we present the results of the MDS analysis on the similarities computed using the nine computational measures on the two sets of object data (fine and coarse). The data are comprised of stress values for one and two dimensional solutions (Tables I and II), plots of the two dimensional object maps (Figures 6 and 7), and relative shape/texture weights (Figure 8).

*Results using higher-resolution object data* For all measures except 2D edge detection,

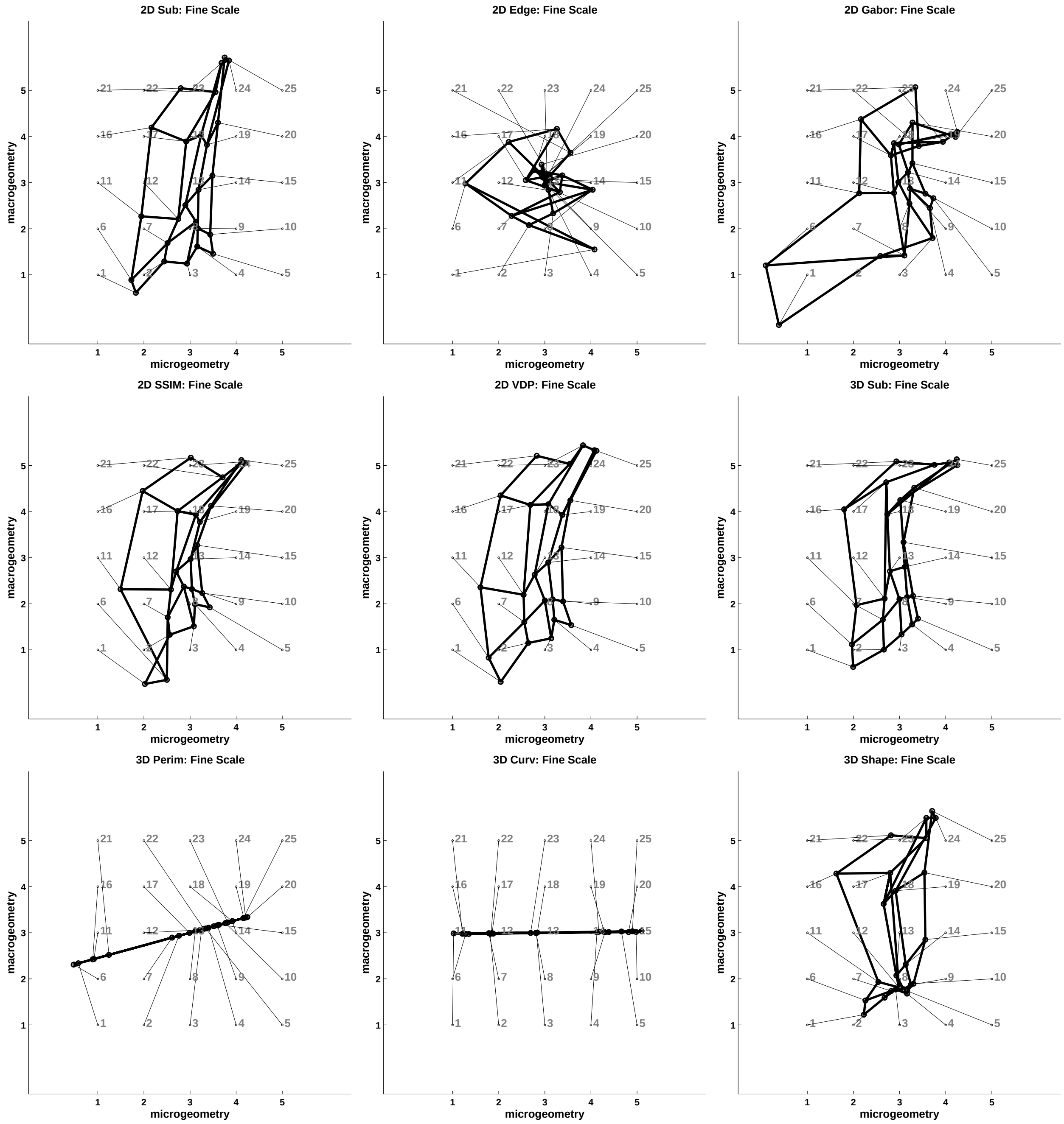


Fig. 6. Stimulus maps derived by MDS analysis of computed similarity data using *finer* resolution object data

MDS stress fell below the threshold of 0.2 for a one-dimensional solution, implying that one dimension sufficed to explain similarity data derived using these measures (Table I). For 2D subtraction, 3D subtraction, VDP, SSIM, and 3D shape histograms, this one dimension corresponded to shape. The dominance of shape over texture for these measures can also be seen from the RSQ weights (Figure 8). For the remainder of the discussion, these measures will be referred to as "shape-dominated measures." For 3D curvature and 3D perimeter, the single dimension required corresponded to texture. In the case of 2D edge detection, two dimensions were required (although texture was weighted much more strongly than shape). One-dimensional stress for the Gabor jet was 0.17 and although shape was technically sufficient to explain computed similarities using Gabor jets, the MDS map shows that the measure is indeed sensitive to texture changes, especially for the bumpiest objects (Figure 6).

MDS maps for the shape-dominated and Gabor jet measures exhibit a larger difference

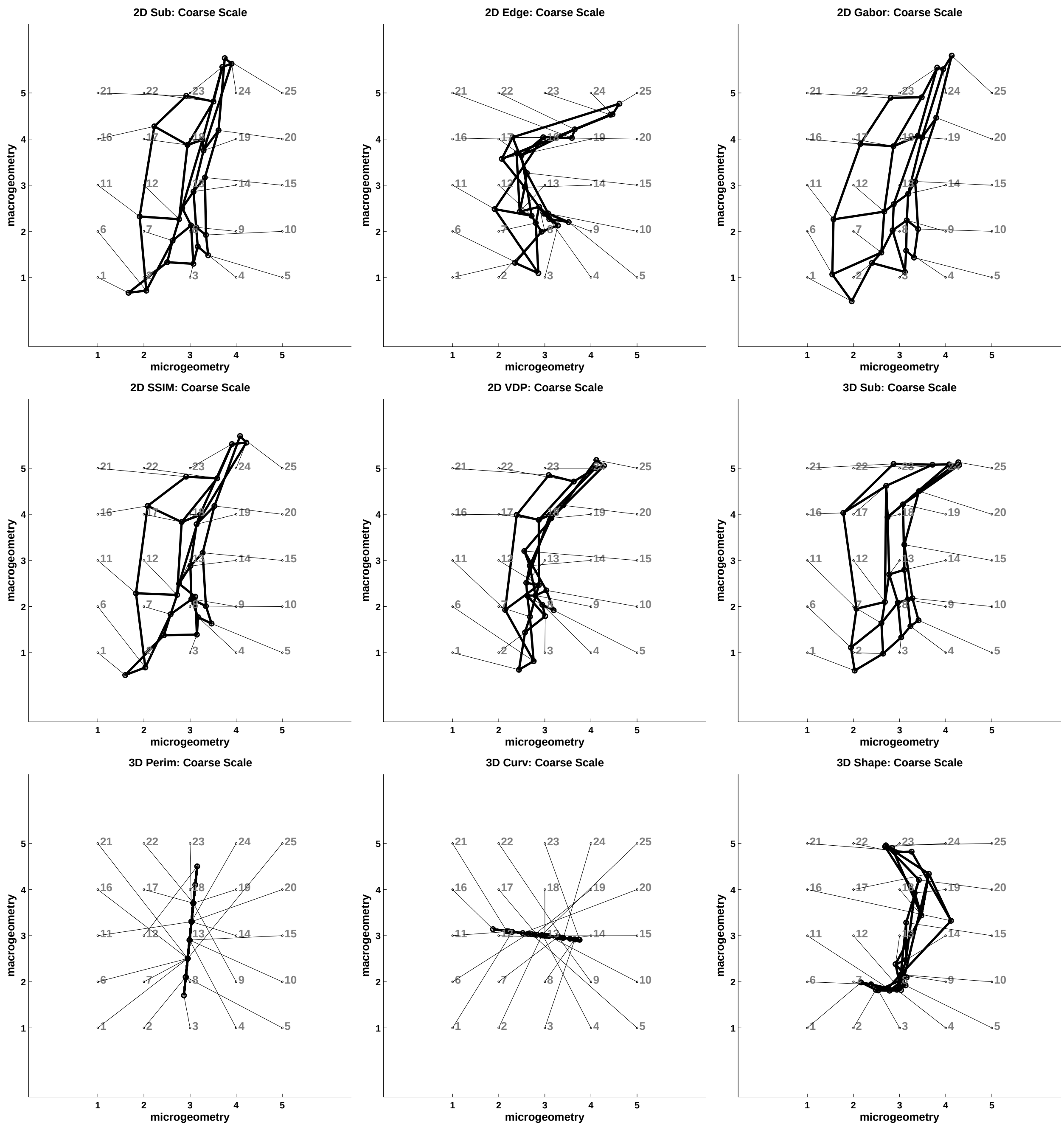


Fig. 7. Stimulus maps derived by MDS analysis of computed similarity data using *coarser* resolution object data

between the bottom three and top two rows of stimuli than, a gap which was also observed in human visual maps. Although they are shape-dominated, these measures are also sensitive to texture differences amongst the objects. However, relative to human maps, they are more sensitive to differences between the bumpiest objects and less sensitive to differences between smooth objects.

The 2D and 3D subtraction maps are quite similar to one another. This is explained by the fact that the 2D images used in this experiment capture most of the variation in the 3D objects. The shape manipulation affects sharp angles in the macrogeometry (such as tips and joints), most of which are visible from the selected 2D view. Had we taken photographs of our stimuli from a viewpoint in which some of these parts were occluded, the maps generated by 2D and 3D subtraction would have differed more. Although texture changes occur over the whole object and are therefore not limited to the frontal view, they have a smaller net effect on 3D vertex positions and pixel values. This also explains the

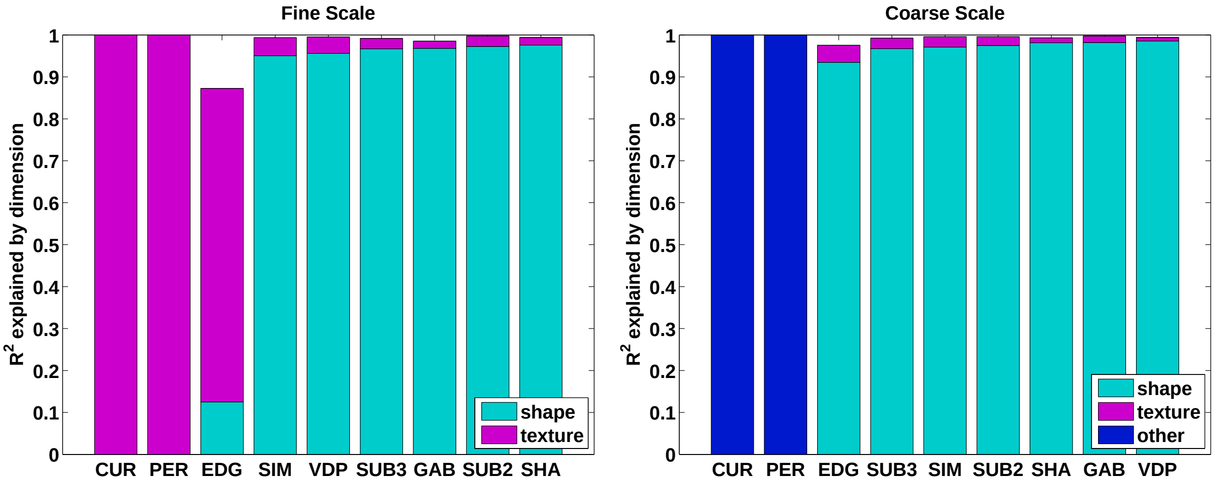


Fig. 8. Dimension weightings based on RSQ for features computed at a finer scale (left) and a coarser scale (right). (CUR: 3D curvature estimate; EDG: 2D edge detection; GAB: 2D Gabor jets; SUB3: 3D subtraction; PER: 3D perimeter; SHA: 3D shape histograms; SIM: 2D Structural Similarity; SUB2: 2D subtraction; VDP: 2D Visual Difference Predictor)

absence of strong texture-related modulation in maps based on global 2D/3D differences.

In contrast, the maps derived from perimeter and curvature show that these measures are not sensitive to changes in object shape for the finer dataset; both features are solely responsive to variations in object texture. The perimeter measure is particularly sensitive to the differences between the most highly-textured objects and the rest. The curvature measure yields a map with more regular spacing between texture levels.

*Results using lower-resolution object data* Using lower resolution object data had a large effect on curvature, perimeter, and edge measures, as can been seen from the maps (Figure 7). At the coarser scale, the perimeter and curvature measures are no longer able to recover texture variation in the stimuli; although their similarity data can be explained using a single dimension (Figure 8), it is not clear how this dimension can be interpreted. The edge measure responds more regularly to shape when computed on the downsampled data, though the map is still quite noisy. Lowering resolution also had a noticeable effect on the Gabor jet map, whose hypersensitivity to the highest texture level was reduced. On the other hand, lowering resolution had lesser effects on the shape-dominated measures. The VDP and shape histogram measures lost some of their ability to separate texture levels. No major effects were observed for 2D subtraction, 3D subtraction, or SSIM.

### 3.4 Perceptual validation of computational measures

In order to assess the perceptual validity of a computational measure, we verified how well the stimulus configuration generated from the measure compared to the configuration generated from human similarity ratings. This was done by fitting the computational maps to the individual visual/haptic human maps, using the fit error as the goodness-of-fit measure, as described in section 2.6.

The fit error enables us to make a *relative* assessment of goodness-of-fit, i.e., we can say that the fit obtained with one measure is better or worse than another; however, it does not provide an *absolute* criterion. To determine such a criterion, we reasoned that a given measure can be deemed to fit the human data well and thus “perceptually valid” to a certain extent (see General Discussion) *if the mean error in fitting the computational map to all individual maps is statistically equivalent to the mean error obtained by fitting each* *individual subject map to all other individual maps.* We refer to the procedure of fitting each individual map to all the other individual maps as "cross-fitting" the individual data and refer to the resulting error as the "cross-fitting error." To test whether a measure met our criterion, we performed a two-tailed t-test between the cross-fitting errors and the fit errors generated by each measure (5% confidence level, assuming independent samples and equal variances).

*Cross-fitting results* Fitting maps derived from individual visual similarity data led to a cross-fitting error of 24% with standard error of 2%. With individual haptic data, we obtained a mean cross-fitting error of 19% with standard error of 2%. Lower cross-fitting error in the haptic condition was related to the fact that maps recovered from individual haptic subject data were more regular than those recovered from individual visual subject data (7/10 haptic maps had 3 or fewer violations of ordinal relationships compared to only 2/10 visual maps). This likely stemmed from differences in experimental settings across conditions; notably, visual stimuli were presented on a computer screen in a darkened room, whereas haptic stimuli were presented by an experimenter in a naturally-lit room. In a follow-up study in which experimental conditions were equated for visual and haptic similarity judgments [Cooke et al. 2006], ordinal relationships were recovered equally well using either modality.

*Fit between computational measures and human visual perception* Several measures met our criterion for perceptual validity: the 2D measures of image subtraction, VDP, and SSIM, as well as the 3D measures of mesh subtraction and shape histograms provided mean fit errors which were not significantly different from mean visual cross-fitting error (all p's $> 0.4$) when computed using finer resolution object data (Figure 9). Gabor jets provided a slightly worse fit. The 3D curvature, edge and perimeter measures provided poor fits to visual data, regardless of the scale of object data. For coarse resolution data, the loss of sensitivity to texture differences resulted in slightly worse fits for VDP and shape histograms, and a noticeable improvement in the fit for the Gabor jet measure. The fit also improved for the edge measure, which responded more regularly to shape when computed at a coarser scale. Although we have used mean fit error over all subjects to define our criterion for perceptual validity, it is also informative to consider how each measure fits individual subjects (Figures 9 and 10, right). For the visual subjects, we see a consistent pattern of fits across subjects, with shape-dominated measures consistently outperforming other measures.

*Fit between computational measures and human haptic perception* All of the computational measures we implemented yielded fit errors which differed significantly from the mean haptic cross-fitting error, i.e., none of the features tested met our criterion for perceptual validity relative to the haptic modality for either data set (Figures 9 and 10, left). However, good fits were found in individual cases. When curvature was computed using higher resolution data, it provided good fits to subjects 1 and 2, who were the most texture-dominated. The VDP and 2D subtraction measures provided good fits to subject 10, who was the most shape-dominated of the haptic subjects. The Gabor jet and subtraction measures also fit this subject well for the coarse data set. The worst fits were obtained for subjects for whom *both* shape and texture were important. One reason for this is that all the measures we tested are either shape or texture dominated. Thus one way of modelling our data would be to use a combination of these measures. On the other hand, a more sophisticated 3D measure which we have not included may be capable of modelling our

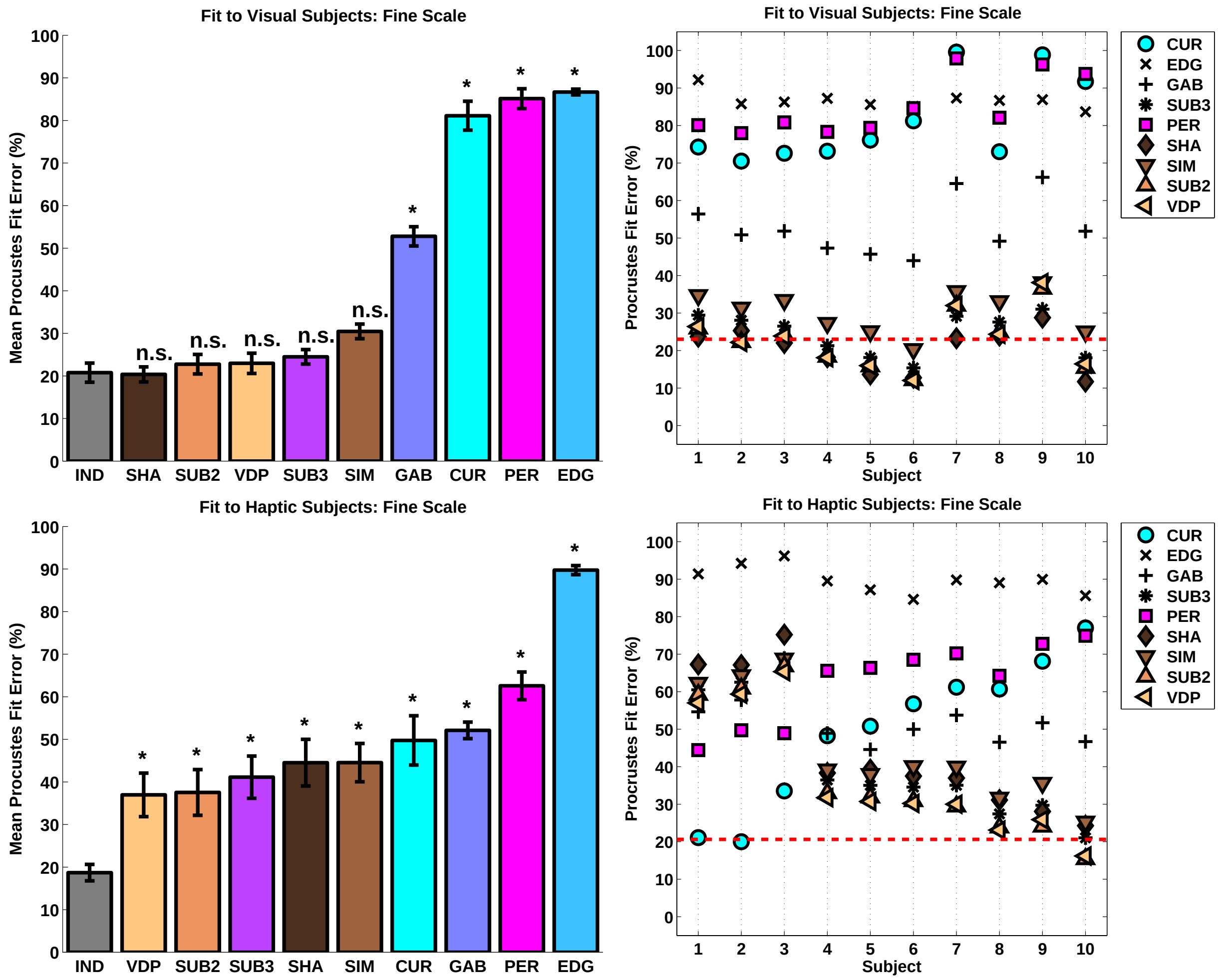


Fig. 9. Fits between computational measures and human maps when features are computed using *finer* resolution object data. Mean fit over all subjects (left) and fits to individual subjects (left). Fits relative to visual data (top) and haptic data (bottom). Error bars represent standard error. * = significant difference compared to mean cross-fitting error (IND); n.s. = not significant. In the right-hand figures, the dashed red line is drawn at one standard error away from the mean cross-fitting error. (CUR: 3D curvature estimate; EDG: 2D edge detection; GAB: 2D Gabor jets; SUB3: 3D subtraction; PER: 3D perimeter; SHA: 3D shape histograms; SIM: 2D Structural Similarity; SUB2: 2D subtraction; VDP: 2D Visual Difference Predictor)

data; further 3D features need to be tested in order to investigate this possibility.

## 4. GENERAL DISCUSSION

### 4.1 Effect of modality on perceptual similarity

*Visual similarity* In visual similarity judgments, shape was the dominant perceptual dimension, whereas texture variation played a lesser role. This finding agrees with the idea that shape is a key determinant of similarity relationships between objects [Edelman 1999]. It is also consistent with the notion that the extraction of global form is one of the visual system's areas of expertise [Klatzky and Lederman 2003]. Distinct clusters of stimuli based on shape appeared in the visual similarity map, hinting at the formation of shape-based categories in similarity space. This observation is interesting given debate surrounding the question of whether similarity relationships form the basis for perceptual categorization [Hahn and Ramscar 2001] and coincides with evidence for a special role of shape in the formation of category structure. For instance, young children have been shown to use shape as a basis for naming generalization, ignoring other properties such as size and tex-

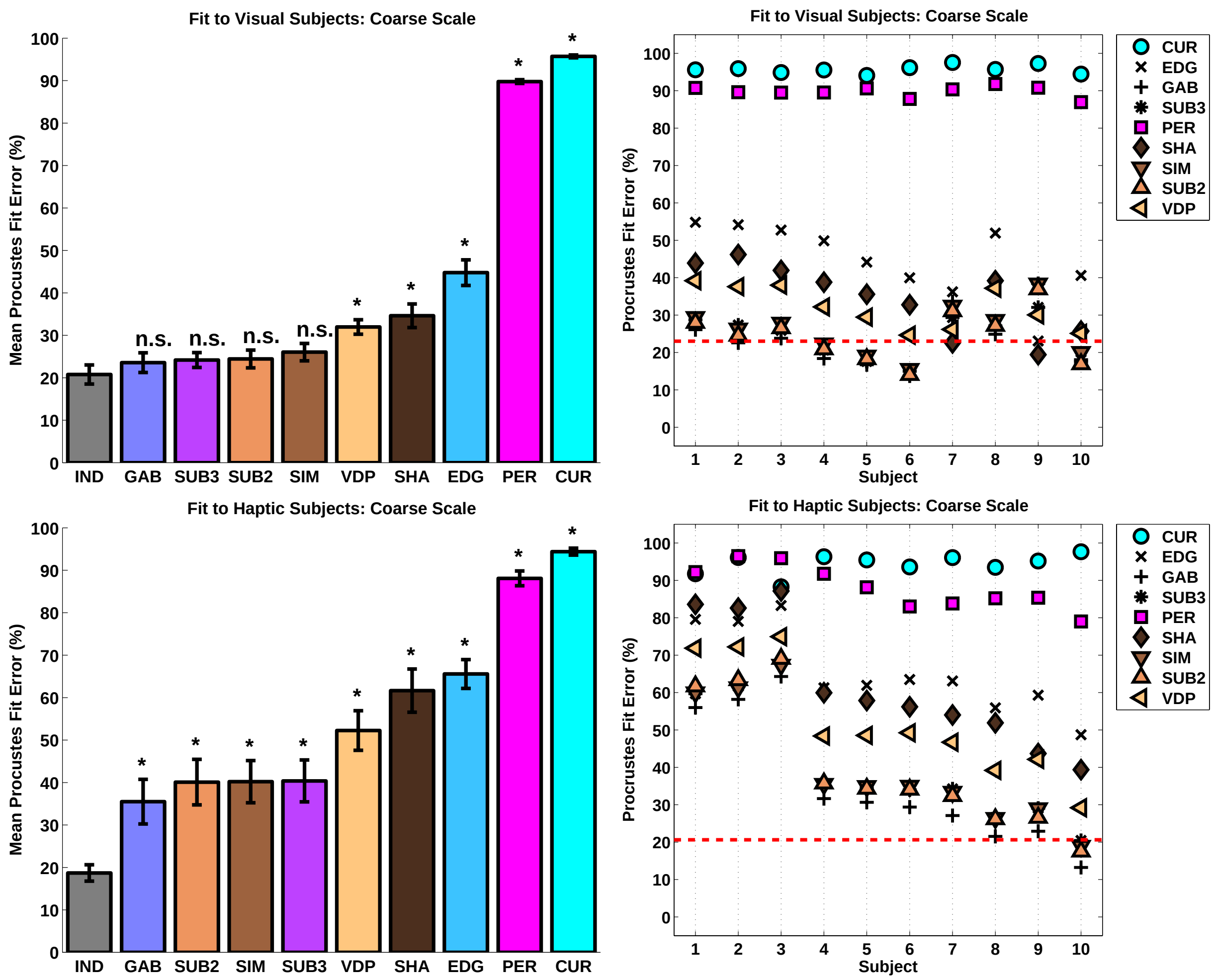


Fig. 10. Fits between computational measures and human maps when features are computed using *coarser* resolution object data. Mean fit over all subjects (right) and fits to individual subjects (left). Fits relative to visual data (top) and haptic data (bottom). Error bars represent standard error. * = significant difference compared to mean cross-fitting error (IND); n.s. = not significant. In the right-hand figures, the dashed red line is drawn at one standard error away from the mean cross-fitting error. (CUR: 3D curvature estimate; EDG: 2D edge detection; GAB: 2D Gabor jets; SUB3: 3D subtraction; PER: 3D perimeter; SHA: 3D shape histograms; SIM: 2D Structural Similarity; SUB2: 2D subtraction; VDP: 2D Visual Difference Predictor)

ture [Landau et al. 1998]. Models of visual object categorization have also been developed on the basis of shape primitives [Biederman 1987] as well as on the basis of similarity relationships amongst shape primitives [Edelman 1999]. An interesting question is whether shape also plays a special role in category formation when objects are first experienced through touch. We have recently found evidence that *both* shape and texture dimensions are capable of determining spontaneous category structure in vision as well as in touch [Cooke et al. 2006]. In addition, we found that the relative importance of shape/texture in determining category structure was the same as the relative weight in judging similarities (shape was more important than texture for vision, while for haptics, shape and texture were roughly equally important).

*Haptic similarity* When subjects touched the objects, they weighted the relative importance of shape and texture differently than when they saw the objects: instead of shape dominating their similarity judgments, *both* shape and texture were important. Given that local material properties are thought to be more accessible to the haptic system than global geometric properties [Klatzky and Lederman 2003], one might have expected haptic similarity ratings to be more strongly affected by texture differences. [Klatzky et al. 1987] found that haptic free sorting of wafer shapes based on similarity was performed preferentially on the basis of material properties as opposed to geometrical properties. However, in a follow-up study [Lederman et al. 1996], in which stimuli were fully 3D and shape variation was no longer limited to the edges, geometric properties played a *more* important role than material properties in haptic similarity judgments. This result indicates that the distribution of geometrical features (e.g., 2D vs. 3D shape information) has an influence on the relative weights of object properties. In our stimulus set, shape information is 3D, but most variation in shape features can be captured in the frontal 2D plane [Cooke et al. 2005]. In light of the two aforementioned studies, it could be that this "2 1/2 D" distribution of shape features contributed to the even weighting of shape and texture properties in this experiment. Experiments involving stimuli with controlled variations in the distribution of shape features would be required to test this hypothesis.

Another difference between our study and [Lederman et al. 1996] is that subjects freely explored the objects and used a variety of different hand movements, whereas subjects in this experiment were restricted to contour-following. Although contour-following allows for the extraction of both shape and texture properties, it is thought to be optimal for extracting precise shape information. We are currently investigating how controlled variation in exploratory procedure affects relative property weightings for this stimulus set.

*Differences and commonalities in haptic and visual similarities* We found a larger amount of individual variation in the dimension weights derived from haptic similarity data than in those derived from visual similarity data. This may have arisen due to differences in exploration time: in the visual condition, viewing time was controlled by a software program and was kept constant (500ms) for all subjects, whereas in the haptic condition, subjects were allowed to explore the object for up to 10s. Actual exploration time varied from individual to individual and was as short as 3s per object. [Lakatos and Marks 1999] found that that local and global geometrical properties had comparable effects on haptic similarity ratings for short exploration times (0.5s to 4s), but that the role of global properties increased with exploration time. This suggests that subjects who took longer to explore the objects may have been more shape-biased than those who used shorter exploration times. However, in a follow-up study [Cooke et al. 2006], in which exploration times were kept between 3 and 5s, we found the same pattern of dimension weights: visual subjects were consistently shape-dominated, while haptic subjects exhibited variable weights, with shape and texture being equally weighted on average. One remaining explanation is that since subjects have less experience with haptic similarity ratings, they tend to invoke cognitive strategies and rules more often when access is provided via touch instead of vision, which in turn leads to greater variability in dimension weights.

Despite the differences we found between visually and haptically-derived representations, there were also important commonalities. In both visual and haptic conditions, subjects were able to extract the two kinds of parametric variation which were used to create the stimuli. These two stimulus variations, which we initially referred to as changes in "macrogeometry" and "microgeometry" were consistently referred to by subjects as changes in "shape" and "texture." The fact that subjects were able to extract systematic variation along these two dimensions is a non-trivial ability given the high-dimensionality of the visual and haptic measurement spaces. For instance, assuming gaze fixation, the visual measurement space might be approximated by the number of pixels in the images of the stimuli. The haptic measurement space might be approximated by the 3D forces exerted on the finger plus the relevant joint angles and positions, taken over the course of the contour-following procedure. Furthermore, the two stimulus manipulations may have had non-linear effects on measurements of object data. In spite of this, maps derived from mean human similarity data exhibit clear, regular responses to the two stimulus manipulations and ordinal relationships present in the stimulus set are recovered. Understanding how the visual and haptic systems deliver such close results despite large differences in the anatomical structure of receptors and pathways which convey object information to the brain is a key motivation for our work.

### 4.2 Similarity-based feature validation

In this paper, we have proposed a criterion for feature validity based on the fit error between maps based on the fit between computational and human similarity measures. However, as discussed above, the results of our human experiments indicate that perceptual similarity between objects can vary as a function of the modality used to experience the objects. Therefore, feature validity, when based on perceptual similarity, depends on the modality assumed to be used for perception. This underscores the importance of specifying the modality (or combination of modalities) to be used when evaluating feature validity.

*Visual feature validation* Several of the features we implemented met our criterion for perceptual validity relative to the visual modality (2D/3D subtraction, 3D shape histograms, VDP and SSIM). These results are in accordance with those reported in [Watson et al. 2001]. The strong performance of the VDP and SSIM, which is are industry standards for assessing image differences, is to be expected. In this sense, they can also be considered benchmarks against which the performance of the other measures can be compared. Surprisingly, the much simpler subtraction-based measures yielded comparable stimulus maps and fit errors. One apparent difference between the shape-dominated measures and the human visual data lies in their response to texture changes: the human data (Figure 4) do not exhibit the same hypersensitivity to high texture levels observed in some of the computational maps (Figures 6 and 7).

The fact that the 3D subtraction map met our criterion for perceptual validity could be taken as an indication that the human visual system reconstructs 3D geometry from 2D images; however, as pointed out earlier, 2D and 3D subtraction measures likely yielded similar results on our stimuli since most of the variation among stimuli occurs in the image plane. A stronger test of whether 3D measures are indeed perceptually valid for the visual modality would require the use of a stimulus set in which variation occurs in depth.

Perimeter and curvature measures provided poorer fits to human visual maps. This is mainly due to the insensitivity of these measures to changes in shape. This result shows that the visual system does not rely solely on curvature or perimeter estimates (at least not as we have implemented them) to judge similarities. This is not as trivial as it may seem: it is indeed possible to extract object perimeter from 2D images and, since perimeter can also be extracted in the haptic modality, it could serve as a convenient feature for sharing information between vision and touch. Curvature can also be extracted from 2D images; in fact, the visual system could use shading-related changes in pixel intensities to estimate both local curvature (texture-from-shading [Todd et al. 2004]) and global curvature (shape-from-shading [Blake and Bülthoff 1991]). Our findings do not rule out the possibility that the visual system uses these features, but they indicate that neither perimeter nor curvature (as we computed them) is sufficient to explain our human visual similarity data.

*Haptic feature validation* None of the measures we tested met our criterion for perceptual validity relative to the haptic modality. This is due in part to the strong shape or texture bias of the measures we implemented, which meant that good fits were not obtained for subjects with intermediate shape/texture weightings. Although identifying such a feature would help to address this problem, one would still need to account for intra-subject variability in haptic similarity judgments. For this, it may be necessary to implement an individually-adjusted combination of features.

At the individual subject level, good fits were obtained for subjects who were strongly biased either towards shape or texture. A surprising finding was that despite the fact that subjects explored the objects via a contour-following procedure, the map based on the perimeter measure did not yield good fit values for either of the two scales we tested. Possible explanations for this could be that subjects do not compute perimeter during contour following, or that it is computed but not used to judge similarity, e.g., because the estimate is not statistically reliable [Ernst and Banks 2002].

Finally, contrary to our expectations, we did not find that measures computed on 3D data provided generally better fits to haptic data than measures computed on 2D data. Studies involving objects with greater 3D variation are needed to further test whether 2D features are indeed sufficient to model haptic object representations.

### 4.3 Variation of object scale

In this study, we computed features using object data at two different scales and looked for differences in fits to human data caused by this variation. For our object set, the main effect of presenting coarser data was to make it more difficult for measures to recover texture variation. This had little effect on fits to human visual data, since the recovery of shape information is the critical factor in determining a good fit. For edge detection, downsampling the data led to a better recovery of shape variation and fit error was lower in the coarse than in the fine condition. Texture-dominated measures (curvature and perimeter) were strongly affected by downsampling; the measures no longer captured texture variation in the stimuli and, as a result, fits to texture-dominated haptic subjects worsened. The fact that our curvature measure was strongly affected by downsampling object data is interesting in light of findings that human haptic curvature estimation is also scale-dependent [Louw et al. 2000; Nefs and Kappers 2003]. Further studies are needed to systematically investigate how the scale of object data affects *perceptual* similarity judgments and compare this to the sensitivity of computed measures. In turn, this type of knowledge can help to design effective haptic, visual, and multimodal interfaces. More generally, we have demonstrated how our method can be used to assess the sensitivity of features to computational parameters (in this case, data resolution). We envisage that this procedure could be performed iteratively in an optimization setting, e.g., to find the resolution of object data for which feature X provides the best fit to human data.

### 4.4 Summary of findings and outlook

In this work, we have shown how a similarity-based approach can be used to assess the perceptual validity of features relative to a specific sensory modality. With the example of scale variation, we also showed how the method can be used to assess the sensitivity of features to changes in the way they are computed. This work is just a first step, however, towards developing a rigorous test of perceptual validity. In particular, our stimulus set only contained 25 objects, meaning that the fits between computational and perceptual maps were based on relatively few data points. Another limitation is that we limited our study to a single stimulus class. That having been said, we were able to provide a "plausibility test" for perceptual validity for a set of standard 2D and 3D features. We found a number of perceptually plausible features for the visual condition, with the critical factor being the features' ability to recover shape variation in the stimulus set. The features we tested did not provide good overall fits to the haptic data, although good fits were found for individuals who were particularly biased toward either shape or texture variation in the stimuli. A more sophisticated 3D feature or an individually-adjusted combination of features may be required to model our haptic data.

Although similarity-based methods have already been applied to compare perception in different modalities (e.g., [Garbin 1988]), our *combination* of similarity measures and *parametrically-related* stimuli differentiates our approach and allows us to compare how different computations or modalities recover high-level, topological relationships in the stimulus set. In addition to the rich qualitative information contained in the MDS maps, the method provides two important quantitative metrics: 1) weightings of the dimensions which span the output space generated by a given modality or computational measure and 2) a goodness-of-fit measure between two stimulus configurations in the output space.

We suggest that the method can be used for three distinct purposes:

(1) First, it can be used to visualize and quantify changes in human similarity-based representations of objects when different or multiple modalities are used to explore objects. The effects of varying, adding, or removing stimulus properties as well as the effects of stimulus-independent manipulations (e.g., changes in viewpoint or illumination) on the structure of human object representations can be studied. The method also makes it possible to identify object properties which are important for unimodal perception, but are not good predictors of multimodal perception.

(2) Second, the method can be used to visualize and quantify how well a computed object feature is able generate human-like stimulus maps and to compare the relative performance of different features on a given set of object data.

(3) Third, the method makes it possible to assess the sensitivity of features to changes in inputs (e.g., resolution) or algorithm parameters. The method offers a criterion which can used to optimize such parameters relative to human perception.

The next step in our work is to develop more rigorous tests of perceptual validity, which can be added as a second stage once perceptual plausibility has been established using the method presented in this paper (Figure 11). This second stage of perceptual validity testing has two important components: 1) a test of generalization to a larger number of more complex stimulus classes (left-hand boxes) and 2) a test of generalization to a larger number of points within each stimulus space (right-hand boxes). A rigorous evaluation of perceptual validity which incorporates these components should provide substantial benefit for developing efficient artificial representations of objects and also help to elucidate the computational mechanisms underlying human perception.

## 5. ACKNOWLEDGMENTS

The authors wish to thank the following persons for their help with the work presented here, a part of which was also presented in [Cooke et al. 2005]: Rafal Mantiuk and Karol Myszkowski for graciously providing code to implement the VDP, Florian Steinke for providing code to compute 3D curvature, Martin Breidt and Michael Renner for assistance with stimulus production, and Karin Bierig for helping to run the haptic experiments.

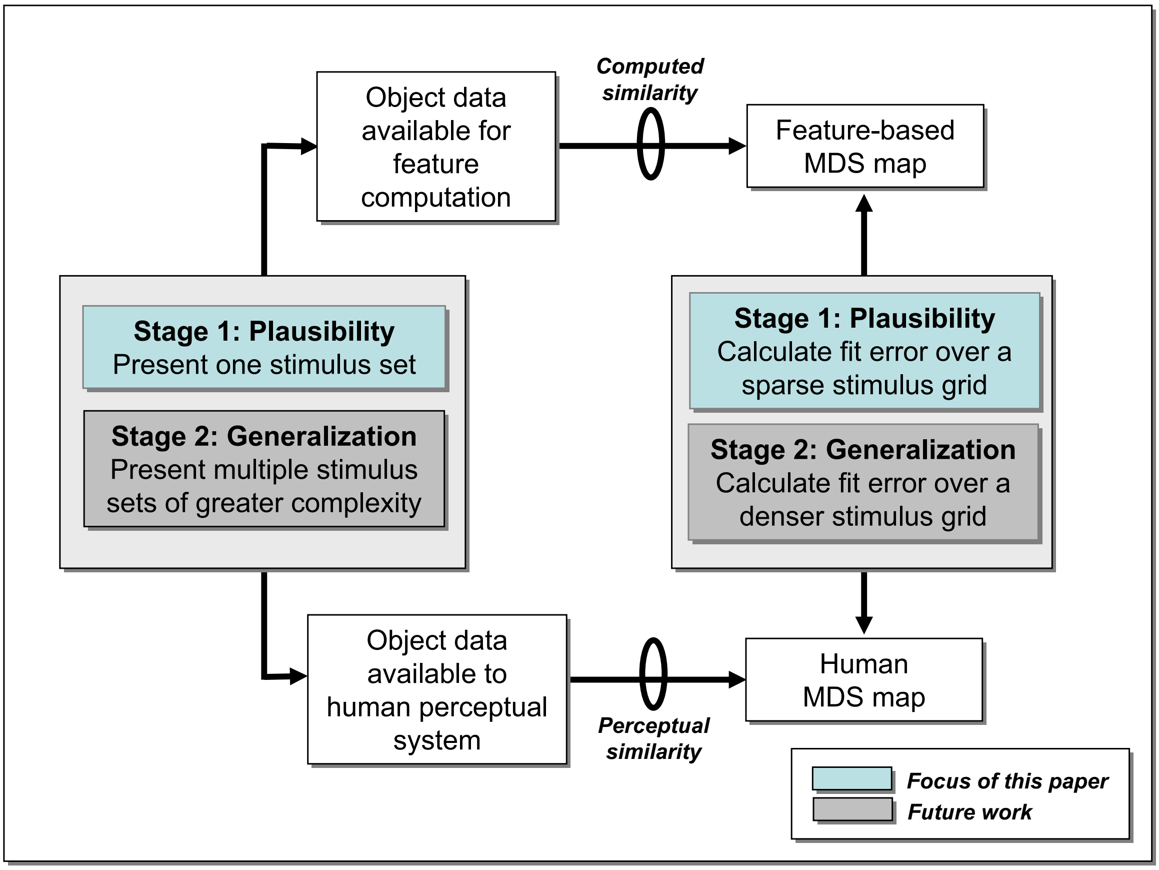


Fig. 11. A two-stage model for perceptual feature validation. The first stage involves a less demanding, but relatively straightforward plausibility test. In the second stage, perceptual validity is tested more rigorously by testing whether goodness-of-fit holds for a larger number of points in the stimulus space and to more complex stimulus classes.

## REFERENCES

E Acosta, B Temkin, JA Griswold, SA Deeb, T Krummel, RS Haluck, and LR Kavoussi. Heuristic haptic texture for surgical simulations. *Stud Health Technol Inform*, 85:14–16, 2002.

M Ankerst, G Kastenmüller, HP Kriegel, and T Seidl. 3D shape histograms for similarity search and classification in spatial databases. In R Güting, D Papadias, and F Lochovsky, editors, *Advances in Spatial Databases, 6th International Symposium, SSD'99*, volume 1651, pages 207–228, Hong Kong, China, 1999. Springer.

I Biederman. Recognition-by-components: a theory of human image understanding. *Psychol Rev*, 94(2):115–147, Apr 1987.

A Blake and HH Bülthoff. Shape from specularities: Computation and psychophysics. *Philosophical Transactions of the Royal Society (London) Series B*, 331:237–252, 1991.

I Borg and P Groenen. *Modern multidimensional scaling*. Springer, 1997.

DH Brainard. The psychophysics toolbox. *Spat Vis*, 10:433–436, 1997.

HH Bülthoff and S Edelman. Psychophysical support for a 2-d view interpolation theory of object recognition. *Proceedings of the National Academy of Science*, 89:60–64, 1992.

M Cadik and P Slavik. Evaluation of two principal approaches to objective image quality assessment. In *Proceedings of the 8th International Conference on Information Visualisation*, pages 513–518, Washington, DC, USA, 2004. IEEE Computer Society.

T Cooke, F Jäkel, C Wallraven, and HH Büelthoff. Multimodal similarity and categorization of novel, three-dimensional objects. *Neuropsychologia*, Mar 29:[Epub ahead of print], 2006.

T Cooke, F Steinke, C Wallraven, and HH Bülthoff. A similarity-based approach to perceptual feature validation. In *APGV 2005 - Symposium on Applied Perception in Graphics and Visualization*. ACM SIGGRAPH, August 2005.

TF Cox and MA Cox. *Multidimensional Scaling*, volume Number 59 in Monographs on Statistics and Applied Probability. Chapman & Hall, London, 2nd edition, 2001.

S Daly. *Digital images and human vision*. MIT Press, Watson, AB (ed.), 1993.

E Deibert, M Kraut, S Kremen, and J Jr Hart. Neural pathways in tactile object recognition. *Neurology*, 52(7):1413–1417, 1999.

S Edelman. *Representation and recognition in vision*. MIT Press, 1999.

MO Ernst and MS Banks. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415:429–433, 2002.

T Funkhouser, P Min, M Kazhdan, J Chen, A Halderman, D Dobkin, and D Jacobs. A search engine for 3d models. *ACM Trans. Graph.*, 22(1):83–105, 2003.

CP Garbin. Visual-haptic perceptual nonequivalence for shape information and its impact upon cross-modal performance. *J Exp Psychol Hum Percept Perform*, 14(4):547–553, 1988.

MA Giese. *A neural model for biological movement recognition: a neurophysiologically plausible theory*, pages 443–470. Kluwer Academic Publishers, Norwell, MA, USA, 2004.

U Hahn and M Ramscar, editors. *Similarity and categorization*. Oxford University Press, 2001.

JP Jones and LA Palmer. An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. *J Neurophysiol*, 58(6):1233–1258, 1987.

RL Klatzky, S Lederman, and C Reed. There's more to touch than meets the eye: the salience of object attributes for haptics with and without vision. *J Exp Psychol Gen*, 116(4):356–369, 1987.

RL Klatzky and SJ Lederman. *Experimental psychology*, volume 4 of *Handbook of Psychology*, chapter Touch, pages 147–176. Healy, AF and Proctor, RW and Weiner, IB (eds.), Wiley, 2003.

JJ Koenderink. The structure of images. *Biol Cybern*, 50(5):363–370, 1984.

S Lakatos and LE Marks. Haptic form perception: relative salience of local and global features. *Percept Psychophys*, 61(5):895–908, 1999.

B Landau, L Smith, and S Jones. Object perception and object naming in early development. *Trends Cogn Sci*, 2(1):19–24, 1998.

SJ Lederman and RL Klatzky. Extracting object properties through haptic exploration. *Acta Psychol*, 84:29–40, 1993.

SJ Lederman, C Summers, and RL Klatzky. Cognitive salience of haptic object properties: role of modality-encoding bias. *Perception*, 25(8):983–998, 1996.

T Lindeberg. *Scale-Space Theory in Computer Vision*. Kluwer Academic Publishers, Norwell, MA, USA, 1994.

S Louw, AM Kappers, and JJ Koenderink. Haptic detection thresholds of Gaussian profiles over the whole range of spatial scales. *Exp Brain Res*, 132(3):369–374, 2000.

R Mantiuk, S Daly, Kl Myszkowski, and H Seidel. Predicting visible differences in high dynamic range images - model and its calibration. In *Human Vision and Electronic Imaging X, IS&T/SPIE's 17th Annual Symposium on Electronic Imaging*, volume 5666, pages 204–214, San Jose, USA, 2005. SPIE.

D Marr. *Vision: a computational investigation into the human representation and processing of visual information*. W. H. Freeman, San Francisco, 1982.

HT Nefs and JJ Kappers, AML Koenderink. Detection of amplitude modulation and frequency modulation in tactual gratings: a critical bandwidth for active touch. *Perception*, 32(10):1259–1271, 2003.

O Nestares, R Navarro, J Portilla, and A Tabernero. Efficient spatial domain implementation of a multiscale image representation based on gabor functions. *Journal of Electronic Imaging*, 7:166–173, 1998.

SC Pont, AM Kappers, and JJ Koenderink. Similar mechanisms underlie curvature comparison by static and dynamic touch. *Percept Psychophys*, 61(5):874–894, 1999.

CL Reed, S Shoham, and E Halgren. Neural substrates of tactile object recognition: an fMRI study. *Hum Brain Mapp*, 21(4):236–246, 2004.

M Riesenhuber and T Poggio. Hierarchical models of object recognition in cortex. *Nat Neurosci*, 2(11):1019–1025, 1999.

E Rosch, CB Mervis, WD Gray, DM Johnson, and P Boyes-Braem. Basic objects in natural categories. *Cognit Psychol*, 8:382–439, 1976.

K Salisbury, F Conti, and F Barbagli. Haptic rendering: Introductory concepts. *IEEE Computer Graphics and Applications*, 24(2):24–32, 2004.

K Sathian, AW Goodwin, KT John, and I Darian-Smith. Perceived roughness of a grating: correlation with responses of mechanoreceptive afferents innervating the monkey's fingerpad. *J Neurosci*, 9(4):1273–1279, 1989.

MA Srinivasan and RH LaMotte. *Encoding of shape in the responses of cutaneous mechanoreceptors*, chapter 5, pages 59–61. Wenner-Gren International Symposium Series. MacMillan Press, 1991.

F Steinke, B Schölkopf, and V Blanz. Support vector machines for 3D shape processing. In *Computer Graphics Forum (Proceedings of EUROGRAPHICS 2005)*, volume 24, 3, 2005.

JT Todd, AHJ Oomes, JJ Koenderink, and AML Kappers. The perception of doubly curved surfaces from anisotropic textures. *Psychol Sci*, 15(1):40–46, 2004.

S Ullman. *High-level Vision*. MIT Press, 1996.

S Ullman, M Vidal-Naquet, and E Sali. Visual features of intermediate complexity and their use in classification. *Nat Neurosci*, 5(7):682–687, 2002.

Z Wang, A Bovik, H Sheikh, and E Simoncelli. Image quality assessment: From error measurement to structural similarity. *IEEE Trans. Image Processing*, 13:600–612, 2004.

B Watson, A Friedman, and A McGaffey. Measuring and predicting visual fidelity. In *SIGGRAPH*, pages 213–220. ACM Press, 2001.