Visual and haptic perceptual spaces show high similarity in humans

Nina Gaißert
Max Planck Institute for Biological Cybernetics, Tübingen, Germany

Christian Wallraven
Department of Brain and Cognitive Engineering, Korea University, Seoul, Korea

Heinrich H. Bülthoff
Max Planck Institute for Biological Cybernetics, Tübingen, Germany, & Department of Brain and Cognitive Engineering, Korea University, Seoul, Korea

In this study, we show that humans form highly similar perceptual spaces when they explore complex objects from a parametrically defined object space in the visual and haptic domains. For this, a three-dimensional parameter space of well-defined, shell-like objects was generated. Participants either explored two-dimensional pictures or three-dimensional, interactive virtual models of these objects visually, or they explored three-dimensional plastic models haptically. In all cases, the task was to rate the similarity between two objects. Using these similarity ratings and multidimensional scaling (MDS) analyses, the perceptual spaces of the different modalities were then analyzed. Looking at planar configurations within this three-dimensional object space, we found that active visual exploration led to a highly similar perceptual space compared to passive exploration, showing that participants were able to reconstruct the complex parameter space already from two-dimensional pictures alone. Furthermore, we found that visual and haptic perceptual spaces had virtually identical topology compared to that of the physical stimulus space. Surprisingly, the haptic modality even slightly exceeded the visual modality in recovering the topology of the complex object space when the whole three-dimensional space was explored. Our findings point to a close connection between visual and haptic object representations and demonstrate the great degree of fidelity with which haptic shape processing occurs.

Keywords: perceptual spaces, multidimensional scaling, vision, haptics, similarity ratings


Introduction

During the first phase of an infant’s life, the infant is mostly dependent on the visual system to gather shape information about its environment. However, as soon as reaching and grasping becomes available, the infant benefits from an enriched perceptual input about objects in its surrounding (Gibson, 1988). This becomes noticeable as soon as the infant starts walking: Nothing seems to be safe anymore because the infant wants to grasp everything it can reach. Grasping an object enables the infant to turn the object and view it under different angles, but it also provides tactile information about the texture, material, weight, size, and shape of an object. Humans therefore form a visual and haptic representation of their surroundings as soon as active exploration becomes possible. Despite this strong multisensory component of human experience, research on how we learn and process object representations has so far mainly focused on the visual domain. In this paper, we will analyze visual and haptic object representations to enhance our understanding of the haptic modality and its role in “grasping the concepts” of objects. In order to assess the commonalities and differences between the visual and haptic modalities, we based our experiments on a three-dimensional (3D) object space of complex, parametrically defined, 3D objects. These objects were printed on a 3D printer to create tangible objects. Using similarity ratings and multidimensional scaling (MDS) analyses, we were able to visualize and analyze the perceptual spaces of both modalities and thus to provide a better understanding of visual and haptic object representations.

Among the features that can be shared between vision and haptics, we have focused on shape—one of the most salient features for visual processing of objects. In biology, it is well known that shape similarities between different animals or plants can have two origins: homology and convergence. Homology describes shape similarity of a characteristic based on a common evolutionary ancestor, for example, in the case of the forelimb of a human and a monkey. Convergence describes shape similarity formed...
by adaptation to the same environment: for example, the streamlined body of a dolphin and a fish both evolved to reduce flow resistance during swimming. In both cases, function and physical parameters result in shape similarity. Hence, the world is a well-structured place where similar objects tend to function similarly. A cheetah and a jaguar share many features, such as body shape, fur pattern, and prominent canine teeth. If we know that one is a predator, it makes sense to also avoid the other species although both live on different continents. This perception of similarities not only ensures survival but also forms the basis of “cognitive economy” (Rosch, 1988): We are able to generalize from one object to a group of objects that belong together and have the same origins or functions (Shepard, 1987), and to categorize objects into classes (Edelman, 1999; Edelman, Bülthoff, & Bülthoff, 1999; Goldstone, 1994).

Comparing shapes (but also scenes, or events) based on assessing similarities between them therefore is of critical importance for the cognitive processes we rely on (Goldstone & Son, 2005). Multidimensional scaling is an elegant and powerful approach to visualize and study the perceptual spaces that are created by similarities (Shepard, 1962; Torgerson, 1952). Similarity judgments or dissimilarity judgments, or any other measure of pairwise proximity, can be used to represent similarity relations between different entities, which are visualized in an n-dimensional space. The similarity between two objects is inversely related to the distance between two objects in this space, which can be understood as a topological representation of object properties in the brain. This perceptual space contains information about how many dimensions are perceived by humans, whether or not these dimensions correspond to the physical measures of the different entities, and how important the different physical measures are to humans.

In a seminal study, Cutzu and Edelman (1998) used a variety of tasks (including perceptual judgment, delayed matching to sample, and long-term memory recall experiments) to show that the visual domain can fully recover planar configurations of a multidimensional parameter space. The brain thus seems to be able to form metrically veridical representations of similarities among 3D objects. Their study was limited to 3D objects that were presented as 2D pictures on a screen or as passively rotated objects. In our experiments, we analyze whether the recovery of stimulus configurations in a physical object space would also be veridical when participants can actively explore the objects from different viewpoints.

We know that the human visual sense is outstanding in processing object shapes. Recognition of familiar objects appears to be virtually instantaneous (Thorpe, Fize, & Marlot, 1996). Object perception, however, as described in the introductory paragraph, is not exclusive to the visual domain. In everyday life, we integrate visual and haptic information about objects effortlessly—for example, if we decide if an apple is fresh or not. In addition, if we search for a key in our handbag we rely on our haptic modality alone to identify the correct item. Therefore, it does not seem astonishing that identification of everyday objects is fast and accurate when the objects are explored haptically (Klatzky, Lederman, & Metzger, 1985). Other studies have shown that haptic shape identification is also fast and accurate when parametrically defined, low-level object features such as curvature, edges, or texture granularity are explored (Plaisier, Tiest, & Kappers, 2009; van der Horst & Kappers, 2008). The question remains, however, to what degree haptic representation of shape—as a higher level, more global object property—might differ from the visual representation. To study this, we will analyze whether the haptic modality perceives shape similarity in the same fashion as the visual modality and based on this, whether the haptic modality can recover the configurations of a physical object space similarly to the visual modality. Our work builds partly on earlier results from a set of studies that compare visuo-haptic processing of two dimensions: Object shape and texture (Cooke, Jäkel, Wallraven, & Bülthoff, 2007; Cooke, Kannengiesser, Wallraven, & Bülthoff, 2006). In these studies, it was found that, indeed, the underlying perceptual space was highly similar between the visual and haptic modalities. One of the main limitations of this work, however, was that the dimensions of shape (macro-geometry) and texture (micro-geometry) were rather intuitive. Here, we were interested to see whether the results would also hold for a much more complex object space based on a higher dimensional, abstract shape space.

To summarize, in this paper we will compare visual and haptic perceptual spaces of complex, parametrically defined objects embedded in a three-dimensional object space. In addition, we address the question of whether the representations gathered by passive, visual object exploration differ from those gathered by active visual object exploration. Taken together, our experiments shed light on the commonalities and differences between visual and haptic object representations.

### Stimuli

#### Object space

For the experiments described in this paper, we created a complex space of shell-shaped objects that varied along three shape dimensions. We used shell-shaped objects for several reasons: They resemble natural objects, they are not too familiar to participants, they have several features that can be perceived visually and haptically, and there is a parametric model that describes the shape space of these objects.

The objects were generated using the biologically plausible parametric model described by Fowler, Meinhardt, and Prusinkiewicz (1992; Figure 1a) and the software...
ShellyLib (Randolf Schultz, http://www.shelly.de). The mathematical model is based on the following equation:

\[ r = A \cdot \sin \beta \cdot \cot \alpha. \tag{1} \]

This equation constructs a shell-like shape by shifting an ellipse along a helicon spiral, which then forms the surface of the shell. From this equation, three parameters (\( A, \sin \beta \) and \( \cot \alpha \)) were altered in five defined equidistant steps to construct a three-dimensional object space of \( 5 \times 5 \times 5 = 125 \) objects. The shape parameters can be verbalized as follows: \( A \) changes the distance between aperture and tip of the shell, while \( \sin \beta \) corresponds to the symmetry of the object. Parameter \( \cot \alpha \) corresponds to the number of convolutions. Note that since the shells have a rather complex shape, this verbalization of single dimensions was only possible post hoc, that is, by varying the shape parameters along one dimension of the object space and observing the resulting shape change, in combination with understanding the process of object generation; all of these processes, of course were not transparent to the participants.

Since, for the following visual and haptic experiments, pairwise similarity ratings had to be performed, the amount of stimuli had to be reduced from 125 objects to 21 objects in order to be able to conduct the experiments in a reasonable time period. Following the approach of Cutzu and Edelman (1998), we therefore decided to use three orthogonal planes of the object space with seven objects per plane, instead of all twenty-five objects per plane (Figure 2). These seven objects were arranged in a Y-shaped form that is easily detectable in the MDS maps as a non-accidental configuration. The center stimulus of the object space is the center stimulus of every plane and therefore is present three times in our stimulus set (Figure 1b).

**Two-dimensional objects for visual exploration**

Two-dimensional (2D) pictures of all 21 objects were presented to participants on a screen. To generate the 2D pictures, object meshes from ShellyLib were imported into the 3D modeling software 3D Studio Max. The material of the stimuli was set to a white matte material, resembling the plastic material used by the 3D printer. The camera was positioned at a distance of 50 cm from the object with a field of view of 45 degrees. The lighting was a standard omni-illuminant of 3D Studio Max with an intensity multiplier of 1.1. One 2D view was then rendered of every single object such that all object features were clearly visible (all images are shown in Figure A3). The objects were rendered to \( 1280 \times 1024 \) pixel 2D images on a black background (Figure 1b). The 2D pictures of the models were presented on a Sony Trinitron 21" monitor with a resolution of \( 1024 \times 768 \) pixels using the Psychtoolbox extension for MATLAB (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007; Pelli, 1997). Participants used a chin rest to align the line of sight to the center of the screen. The image size varied in the range of \( 9–12 \times 9–12 \) degrees of visual angle, depending on the object. Hence, participants had about the same visual impression as if one of the 3D plastic objects would lie on a table in front of the participant at about 40 cm distance.

**Three-dimensional objects for visual exploration**

For this task, a setup was used that allowed participants to visually explore the objects from every possible angle. Most importantly, we designed the setup such that participants could manipulate the 3D position of the object.
without being able to touch it (Figure 1d). To achieve this, virtual 3D objects were presented on a head-mounted display (HMD) and participants could move and rotate them by moving and rotating a physical interface, a 42 cm-long stick with tracking markers.

For preprocessing, the virtual objects generated with ShellyLib were imported into 3D Studio Max and the surface was set to a white matte material to imitate the appearance of the plastic material used by the 3D printer. The real-time rendering of the objects was performed using OpenGL. An eMagin z800 HMD was used whose two displays had a resolution of 800 x 600 pixels, a diagonal field of view of 40°, and a refresh rate of 75 Hz. The HMD was adjusted to the head of the participants so that the screens were approximately 2 cm away from the eyes and orthogonal to the line of sight. The movement of the physical interface was replicated on the screen by tracking the stick and the head position with the gaze direction. This was done using an infrared camera tracking system (VICON V8), which consisted of four cameras, a set of markers on the interface, and a set of markers mounted on the participant’s head. The markers were tracked at a rate of 120 Hz. With the screen refresh rate of 75 Hz and the tracking rate of 120 Hz, participants had a synchronous percept of moving the physical interface and the visual feedback without any perceivable delay (Chuang, Vuong, Thornton, & Bülthoff, 2007).

Both screens of the HMD showed the same image, thus the objects were presented without binocular disparity. However, by rotating the physical interface, by moving it left and right, and by moving it closer and farther relative to the head position, participants were able to move the objects accordingly and thus generate vivid motion cues that provided a strong three-dimensional impression. Participants therefore had access to both strong shape-from-motion and shape-from-shading cues to form a good impression of the three-dimensional shape of the objects.

In addition, the setup was adjusted to an average-sized person such that it would present the objects at the same size (range of 9–12 × 9–12 degrees of visual angle) as in the visual 2D conditions, when the stick was held in a relaxed position. Since it was allowed to move the stick closer and farther from the head, the final object size varied, as a natural object would do when manipulated by the participant. In a darkened room, the participants had almost the same experience as if one of the plastic objects floated in front of them, with the advantage that participants never touched the real objects.

Three-dimensional objects for haptic exploration

For the haptic experiments, plastic models of the objects were explored by blindfolded participants with both hands (see Figure 1c). To create the solid objects, the virtual models generated previously were postprocessed in 3D Studio Max to generate printable objects. First, the wall thickness of the objects was increased by 6% using the shell modifier resulting in a wall thickness of about 2 mm of the final plastic models. The surface was then smoothed using two iterations of the meshsmooth modifier (this operation averages the normal direction of two neighboring vertices to smoothen the mesh; the reason for applying
this modifier, however, was only to remove artifacts in the micro-mesh topology for printing—it had no effect on the “visual” appearance). The objects were then printed using the EDEN250TM 16 micron layer 3-Dimensional Printing System of Objet, Belgium. The manufacturing process was performed in “high quality mode” (resolution: X-axis 600 dpi = 42 µm, Y-axis 600 dpi = 42 µm, Z-axis 1600 dpi = 16 µm) with a white acrylic-based photopolymer material, resulting in a hard, white, and opaque plastic model with a smooth surface. The resulting 3D objects weighed about 40 g. The maximum dimensions were 5 cm in depth, 10 cm in height, and 15 cm in width.

Experiments

Design and procedure

In each of the following experiments, 10 persons participated who were naïve to the stimulus set. Participants were undergraduate students and were paid the standard rate of 8 EUR per hour. All participants had normal or corrected-to-normal vision. In addition, participants in the haptic experiments did not report any haptic difficulties.

The task in all experiments was to rate the similarity between two objects on a seven-point scale from low similarity (1) to high similarity (7). Participants were explicitly instructed to use the value of 7 for identical objects. No definition of the word “similarity” was given, and participants had to decide on their own on which object features they wanted to focus.

Before the experiment itself started, participants performed 8 non-feedback test trials to become familiar with the task and to get a feeling for the variation inherent in the stimulus set.

In the first three experiments (marked by the suffix “planes”), the three-dimensional object space was divided into three orthogonal planes, each of which contained 7 objects (see Figure 2). In these experiments, every object was compared once with itself and once with every other object of the same plane, resulting in a total of 84 trials \((7 \times (7 - 1) / 2 + 7 = 28\) trials per plane, \(3 \times 28 = 84\) trials in total). It was arbitrarily chosen which of the objects of a pair was presented first. The 84 object pairs were shown randomly in one block. Every participant performed three of these blocks with different randomizations. Participants could take a break after every block.

In Experiment 1 (visual 2D planes), the two-dimensional pictures of the objects were presented on a screen (Figure 1b). Participants had to fixate a cross for 0.5 s before the first object appeared on the screen for 3 s. Then, the screen turned black for 0.5 s before the second object was presented for 3 s. After seeing both objects, participants had to rate the similarity between these two objects by pressing a button between 1 and 7.

Experiment 2 (visual 3D planes) should bridge the gap between visual perception of 2D pictures and free exploration of 3D objects. Therefore, the virtual objects were presented on the HMD (Figure 1d). Participants had 8 s to explore the first object visually from different angles by moving the interface. Then, the second object was presented for 8 s. Again participants were able to move the object freely. Afterward, they had to rate the similarity of the pair of objects by saying a number between 1 and 7 out loud. The rating was recorded by the experimenter.

In Experiment 3 (haptic 3D planes), we compared visual object perception with haptic object perception. Three-dimensional plastic models were used (Figure 1c). Participants were blindfolded and seated in front of a table with a sound-absorbing surface. One object was placed between the hands of the participant. The experimenter gave the signal to start and the participant had 8 s to explore the object with both hands and no restrictions to the exploratory procedure. After exploring the object, the participant had to put the object back on the table and it was replaced by the second object. The experimenter again gave the signal to start and the participant had 8 s to explore the second object. After putting the object back on the table, the participant rated the similarity by saying a number between 1 and 7 out loud. The rating was recorded by the experimenter.

In Experiments 4 and 5 (marked by the suffix “space”), the whole three-dimensional object space was analyzed to investigate whether the additional stimulus pairs would distract the percept of the stimulus configuration of the individual planes or dimensions. Every object was compared once with itself and once with every other object of the object space, resulting in 231 trials \((21 * (21 - 1) / 2 + 21 = 231\). Again, which object of the pair was presented first was arbitrarily chosen. The 231 pairs were shown randomly in one block. Every participant performed three blocks with different randomizations and could take a break after every block.

Experiment 4 (visual 2D space) was performed as described for Experiment 1 but with 231 trials instead of 84 trials.

Experiment 5 (haptic 3D space) was performed according to Experiment 3 but was run on two consecutive days due to the length of the experiment. The first session was started with the regular 8 test trials to make participants familiar with the task and the range of objects. The second session was started with the same 8 test trials to refresh the participants’ memory. Participants were able to take additional breaks within the blocks.

After performing the experiment, participants filled out a questionnaire, where they had to describe and draw the
objects. All participants stated that they based their similarity ratings on overall shape, or listed shape features, like aperture, tip, bulkiness, etc.

**Similarity ratings**

**Analysis**

First, we wanted to compare how visual and haptic object exploration affects the general similarity percept. Therefore, we assessed the consistency of the similarity ratings within and across experiments. Participants’ similarity ratings ranging from 1 to 7 were converted to dissimilarities (by subtracting similarity ratings from 7) and averaged across the three performed blocks. Correlations were calculated between the similarity matrices. Therefore, the average dissimilarity matrix of every participant was correlated with every other participant. For Experiments 1–3, 84 trials were correlated, for Experiments 4–5, 231 trials were correlated, and for comparing the two groups of experiments, the corresponding subset of 84 trials were correlated with each other. The correlations within and across experiments were averaged and the standard error of the mean was calculated. The results of these comparisons are listed in Table 1. Thus, each position of Table 1 gives either a within- or between-experiments correlation. Next, all within correlations were compared to all between correlations by performing a two-sample *t*-test on the values listed in Table 1.

**Results**

The high correlations along the diagonal of Table 1 and the low standard error of the mean values suggest that participants varied little in their strategies to solve the task. Based on this result, we decided to average the similarity ratings across participants for the multidimensional scaling analyses used to determine the 2D topology and the 3D topology described in the next sections (Ashby, Maddox, & Lee, 1994).

The correlations within experiments are as high as the ones between experiments (mean_{within} = 0.871, mean_{between} = 0.868, *t*(16) = 0.278, *p* = 0.785). This shows that the task does not affect the similarity percept of single object pairs. Moreover, it shows that participants perceive the similarities among the objects in a highly similar fashion no matter if the objects are explored visually or haptically (see also Figures A4a and A4b in which the average visual and haptic similarity ratings were correlated and clearly show that visual and haptic object explorations lead to almost the same similarity percept). In addition, the correlations are high regardless of how many objects were compared within the experiments, which already hints at the robustness of the underlying perceptual processing.

**Two-dimensional topology**

**MDS analysis**

To assess the topology of the perceptual spaces, we performed a rank-order MDS analysis by using the non-metric MDS algorithm implemented in MATLAB (MDSCALE). Since non-metric MDS uses the ranks of pairwise distances as input, as opposed to their precise values, it fits the human similarity data better than classical metric MDS (Cooke et al., 2007). Dissimilarity matrices were first averaged across participants, following

<table>
<thead>
<tr>
<th>Visual 2D planes</th>
<th>Visual 3D planes</th>
<th>Haptic 3D planes</th>
<th>Visual 2D space</th>
<th>Haptic 3D space</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visual 2D planes</td>
<td><em>R</em> = 0.872, <em>SEM</em> = 0.006</td>
<td><em>R</em> = 0.874, <em>SEM</em> = 0.003</td>
<td><em>R</em> = 0.848, <em>SEM</em> = 0.004</td>
<td><em>R</em> = 0.883, <em>SEM</em> = 0.003</td>
</tr>
<tr>
<td>Visual 3D planes</td>
<td><em>R</em> = 0.867, <em>SEM</em> = 0.005</td>
<td><em>R</em> = 0.867, <em>SEM</em> = 0.004</td>
<td><em>R</em> = 0.860, <em>SEM</em> = 0.002</td>
<td><em>R</em> = 0.863, <em>SEM</em> = 0.004</td>
</tr>
<tr>
<td>Haptic 3D planes</td>
<td><em>R</em> = 0.861, <em>SEM</em> = 0.006</td>
<td><em>R</em> = 0.861, <em>SEM</em> = 0.004</td>
<td><em>R</em> = 0.861, <em>SEM</em> = 0.004</td>
<td><em>R</em> = 0.861, <em>SEM</em> = 0.004</td>
</tr>
<tr>
<td>Visual 2D space</td>
<td><em>R</em> = 0.906, <em>SEM</em> = 0.003</td>
<td><em>R</em> = 0.906, <em>SEM</em> = 0.003</td>
<td><em>R</em> = 0.892, <em>SEM</em> = 0.003</td>
<td><em>R</em> = 0.892, <em>SEM</em> = 0.003</td>
</tr>
<tr>
<td>Haptic 3D space</td>
<td><em>R</em> = 0.861, <em>SEM</em> = 0.005</td>
<td><em>R</em> = 0.861, <em>SEM</em> = 0.003</td>
<td><em>R</em> = 0.831, <em>SEM</em> = 0.003</td>
<td><em>R</em> = 0.831, <em>SEM</em> = 0.003</td>
</tr>
</tbody>
</table>

Table 1. Cross-correlations. The average dissimilarity matrix of every participant was correlated with every other participant. These correlations were averaged within and between experiments. The mean values as well as the standard error of the mean values are listed. For the conditions marked with space, the first row gives the correlation for 84 trials, the second for 231 trials.
To determine how many dimensions were necessary to explain the data, the stress value was calculated for every plane using one to five dimensions and then plotted (Figure 3). To determine how many dimensions are sufficient to explain the data, it is a common heuristic to look for the “statistical elbow” in the plot (Cox & Cox, 2001). This is the point along the curve where the decrements in stress begin to be less pronounced. A stress level of zero is not mandatory as the human data are expected to contain some noise (Borg & Groenen, 2005; Cox & Cox, 2001; Kruskal & Wish, 1978). For all experiments and all planes, the elbow is clearly visible at two dimensions. Thus, we performed the following MDS for two dimensions.

To visualize the physical object space, the positions of the stimuli were marked with colored “×” symbols and stimulus number (Figure 4). The Y shape of the physical stimulus space was made visible by linking the stimulus positions with dotted lines. The perceptual MDS output map was then fit to the physical space using the Procrustes function of MATLAB. This function fits the points of the perceptual space to the points of the physical space by performing a linear transformation (translation, reflection, and orthogonal rotation). Note that these are valid operations, since MDS does not yield any absolute positions but relative positions in space). Its output consists of the distance of the points of the perceptual space to the physical space as the sum of squared errors (this value is indicated as “procrustes value” in the following sections. Low values indicate a better fit than high values, with a value of 0 for a perfect fit; all values are listed in Table 2). The resulting stimulus positions in perceptual space were marked with “stars” using the same stimulus colors as in the physical space (Figure 4).

To visualize the variation across participants and to quantify the stability of the MDS solutions, the data were plotted in a novel way. We performed MDS 500 times using matrices created by randomly perturbing the average matrices. This Monte-Carlo-like technique was used to define confidence ranges around the location of object in the MDS maps. The average matrix was taken and for every similarity rating a random value within the range of ±1 standard error of the mean across participants was added. Each of the 500 resulting matrices was used for MDS and the resulting MDS solutions was fit to the MDS solution generated with the unperturbed matrix using the Procrustes function. The outermost contour line in Figures 4, A1, and A2 encloses 80% of the calculated perturbed MDS solutions. When overlaying two plots of different MDS maps (as done for Figures A1 and A2), the confidence ranges give a good impression of how different the two perceptual spaces really are. In addition, this method allows us to calculate a quantitative measure, that is, the overlap between two maps as determined by the enclosed confidence areas.

In order to test the “quality” of the resulting topology, that is, to determine the degree with which participants were able to perceptually reconstruct the topology of the underlying physical object space, the following tests were performed.

**Neighborhood test:** In order to get a measure of the faithfulness of the perceptual reconstruction, we determined the nearest neighbors of the perceptual and the physical spaces. That is, we took object 1 of the physical space, calculated the distance between this point and all seven points of the average perceptual space and determined which object was nearest. The number of points for which the correspondence was correct was then determined for each plane and each experiment.

**Random configuration:** Second, we arbitrarily distributed seven points within a circular area with a diameter of six (similarity ratings ranged between 1 and 7, resulting in a maximum distance of 6) to create 500 different random point configurations. The resulting configurations were then fit to the positions of the objects in physical space using the Procrustes function and the procrustes values...
were determined. This distribution of fit values was then compared to the one obtained by fitting the experimental data to the physical space. Since the two distributions are not normally distributed, a Mann–Whitney U-test was performed to compare the fit values. The same was repeated with 500 Y-shaped configurations. Therefore, seven points were randomly distributed along a Y-shaped form.

**Topology change:** Finally, we tested how the procrustes values would change, when participants would introduce local errors in the topology, such as changing the order of neighboring points in the topology. For this, the 500 MDS representing participants’ data were shuffled by swapping two neighboring points. These configurations were again fit to the physical space and the two fit value distributions were compared using non-parametric tests.

In addition, the procrustes distributions were used to determine possible differences between the three different planes and the five different experiments.

### Results

To analyze the 2D topology of the perceptual spaces, we looked at the similarity ratings within each plane. Therefore, the corresponding 84 trials of all five experiments were analyzed. As can be seen in Figure 3, which plots the stress values for MDS solutions from 1 to 5 dimensions, for all different stimulus planes, the elbow lies at 2 dimensions. This means that each two-dimensional plane of the physical object space is also perceived as containing two perceptual dimensions in both visual experiments, as well as in the haptic experiment. This may sound trivial at first, but having a look at seven stimuli of, e.g., plane 1 (see Figure A3) shows how hard it actually is to determine the parameters used for stimulus generation visually. The fact that our haptic modality is actually is to determine the parameters used for stimulus generation visually. This may sound trivial at first, but having a look at seven stimuli of, e.g., plane 1 (see Figure A3) shows how hard it actually is to determine the parameters used for stimulus generation visually. The fact that our haptic modality is equally well suitable to reconstruct the shape parameters makes this finding even more interesting.

The results of the 2D MDS are visualized in Figure 4 after being fit to the underlying physical stimulus configuration. This figure, together with the corresponding low procrustes values (listed in Table 2), already indicates that the perceptual spaces reflect the topology of the physical spaces.

For the Neighborhood test, we found for plane 1 of the first experiment, for example, that the correspondence between the perceptual and physical spaces was perfect, meaning that the nearest neighbor for object 1 was 1, for object 2 it was 2, and so on. In total, we found correspondence to be perfect for 9 out of 15 planes (marked in bold in Table 2). For plane 2 of Experiment 1, only six of seven objects had the correct correspondence. Given that the misassigned object P9 was positioned between the two correctly assigned objects P8 and P10 (that is, P9–P8 < P10–P8 and P9–P10 < P8–P10), we concluded that the topology was conserved here as well. All other misassigned objects were surrounded by correctly assigned objects as well. This analysis showed that the broad topology of the Y-shaped form was conserved across all experiments. Participants therefore were not only able to perceive the relevant stimulus dimensions, but they also retained the neighborhood relations correctly and were able to recover the topology of the physical space. Again, this finding is even more striking when considering that in previous experiments (not reported here) less than 30% of the participants were able to reconstruct the topology of every plane correctly when explicitly asked to do so (see Figure A3).

Accordingly, the comparison of the procrustes distribution of 500 participants’ MDS with the randomly generated spaces shows a highly significant difference in fit quality, suggesting that it is highly unlikely that the observed perceptual spaces are randomly structured (medianMDS = 0.111, medianrandom = 0.781, U = 3375, p = 0.000, see Figure A5). Moreover, it is highly unlikely that random Y-shaped forms were perceived (medianMDS = 0.111, medianY = 0.772, U = 25,753, p = 0.000).

Comparing the medians of participants’ fit values with those from the topology change condition, in

---

<table>
<thead>
<tr>
<th></th>
<th>Plane 1</th>
<th>Plane 2</th>
<th>Plane 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visual 2D planes</td>
<td>0.0465</td>
<td>0.1394</td>
<td>0.2586</td>
</tr>
<tr>
<td>Visual 3D planes</td>
<td>0.0807</td>
<td>0.0949</td>
<td>0.1979</td>
</tr>
<tr>
<td>Haptic 3D planes</td>
<td>0.1542</td>
<td>0.1641</td>
<td>0.0523</td>
</tr>
<tr>
<td>Visual 2D space</td>
<td>0.0929</td>
<td>0.0984</td>
<td>0.2400</td>
</tr>
<tr>
<td>Haptic 3D space</td>
<td>0.0620</td>
<td>0.0978</td>
<td>0.0995</td>
</tr>
</tbody>
</table>

Table 2. Procrustes values. Using participants’ similarity ratings and performing 2D MDS, the perceptual spaces were generated. The perceptual spaces were then fit to the physical object space using the Procrustes function. The distances between the points of the perceptual and physical spaces were calculated and are listed here. Low values indicate a better fit than high values, with a value of 0 for a perfect fit. Moreover, the topology of the perceptual space was tested by determining the nearest neighbors. All planes marked in bold showed perfect correspondence. In all other cases, the order within the Y-shaped form was conserved as well since misassigned objects were surrounded by correctly assigned objects (see text for more details).
which one object was exchanged with its next neighbor, showed a significant effect as well (median_{MDS} = 0.111, median_{shuffled} = 0.238, \( U = 41,790,684, p = 0.000 \)). This test provides an even stronger indication that participants were, indeed, able to reconstruct the Y-shaped stimulus space and that the neighborhood relations were correctly identified.

In the following, we will take a closer look at the reconstruction quality for the different planes and experiments. To assess the differences between the three planes, the procrustes distributions of all five experiments were pooled and compared over planes (see Figure A5). We found that plane 3 had a significantly higher median (median_{plane3} = 0.145) than plane 1 (median_{plane1} = 0.087, \( U = 1,391,413, p = 0.000 \)) and plane 2 (median_{plane2} = 0.108, \( U = 2,218,117, p = 0.000 \) indicating that the combination of \( \sin \beta \) and \( A \) seems to be less intuitive than the combination of \( \sin \beta \) and \( \cot \theta \) or \( \cot \theta \) and \( A \). This can also be seen in the size of the confidence ranges displayed in Figure 4, which approximate the stability of the MDS solutions. The confidence ranges are highest for plane 3 in the visual 2D and 3D planes experiment again pointing out that the parameter combination of \( \sin \beta \) and \( A \) seems to be less intuitive.

For plane 3, the procrustes values (Table 2) are notably higher in the visual 2D planes and visual 2D space experiments (0.26 and 0.24, respectively) and the Y shape is less clearly visible in the MDS output maps. This might indicate that access to the 3D shape of the objects (either visually or haptically) can help in reconstructing the underlying physical space. Note, however, that this reconstruction works rather well for the other images as the conserved topology of plane 1 and 2 show.

In statistics, a visual check for determining whether two means are significantly different consists of verifying that their confidence intervals do not overlap. Following this idea, we visualized the overlap of the calculated density fields by plotting the outermost contour line of all experiments or all planes for every single stimulus into one figure (see Figures A1 and A2, respectively). The visualized confidence ranges correspond to altitude profiles and thus enclose volumes not areas. For every stimulus, we calculated the overlapping volume between experiments and planes and normalized these volumes by the sum of the compared volumes. Using this technique, we found that the overlap of the different experiments is with 60% significantly higher than the overlap for the different planes (overlap planes = 30%, \( t(313) = -9.989, p = 0.000 \)). This shows, on the one hand, that the different ways of object exploration in the different experiments lead to rather similar perceptual spaces. On the other hand, the three different planes are less congruent and thus topologically more different. The overlaps between the different planes (mean overlap plane 1 versus plane 2: mean = 26%, plane 1 versus plane 3: mean = 30%, plane 2 versus plane 3: mean = 33%) were compared in two-tailed \( t \)-tests across experiments but did not reach significance (\( t(68) = -0.517, p = 0.606, t(68) = -1.011, p = 0.315, t(68) = -0.512, p = 0.610 \), respectively). Whereas the average data seem to show a difference between the planes, these results indicate that the inherent variance in the data across all conditions is too large to find a major effect.

In more detail, the high degree of overlap (66%) of the density fields for Experiment 1, where 2D views were presented, and Experiment 2, where 3D objects were actively explored, show that participants can reconstruct the three-dimensional shape from the two-dimensional pictures—or at least that the relevant features for similarity judgments of the 3D shape are clearly visible in the 2D pictures. Similarly, when we compare Experiment 3, in which the objects were explored haptically, with Experiments 1 and 2, in which objects were explored visually, we can again see the high degree of overlap of the density fields (59% and 71%, respectively). This demonstrates that a highly similar perceptual space is formed in both modalities.

Finally, we compare Experiments 1 and 2 with Experiments 4 and 5 to test the stability of the results. For analyzing the 2D topology of the perceptual spaces, only 84 similarity ratings were taken into account, while in Experiments 4 and 5 actually 231 similarity ratings were performed. The additional trials can be used to test the robustness of the perceptual spaces. The MDS analysis of the full set of trials did not reveal any changes in the resulting topology. For all experiments, the Y-shaped stimulus configuration was recovered, and the order within the Y-shaped form was conserved both by the visual as well as the haptic modality. In addition, there was no significant change in fit quality in the visual modality (median_{exp1} = 0.134, median_{exp4} = 0.104, \( U = 1,095,052, p = 0.207 \)). Interestingly, the additional trials in the haptic modality seemed to result in an improvement in reconstruction quality (median_{exp2} = 0.121, median_{exp5} = 0.088, \( U = 565,572, p = 0.00 \)).

In summary, our results show that participants were able to perceive the three dimensions underlying stimulus generation in all conditions with a surprising degree of fidelity. In addition, both the visual and haptic modalities recovered the topology of the physical object space remarkably well.

### Three-dimensional topology

#### MDS analysis

The previous results were gathered by looking at similarity ratings within each plane. Given that the original physical space is three-dimensional with the three tested planes in an orthogonal configuration, it still might
be that the three perceptual spaces are perceived separately. In order to investigate the degree to which participants were able to perceive the structure of the whole three-dimensional object space, the full set of similarity ratings from Experiments 4 and 5 were analyzed.

Similar to the previous analysis, all ratings were averaged across blocks and participants to form an average dissimilarity matrix. The diagonal was set to zero and non-metric MDS was performed. The stress value was calculated for one to fifteen dimensions and plotted (Figure 5). For the haptic modality, the elbow that is commonly used to determine the sufficient amount of dimensions is visible at three dimensions. For the visual modality, the elbow is less clearly determinable, lying at either two or three dimensions. For a better comparison of the visual and haptic perceptual spaces, we decided to perform the following analyses for both modalities first for three-dimensional solutions.

The resulting three-dimensional perceptual spaces of both modalities were fit to the physical space using the Procrustes function and the distance between the points was calculated. In addition, we performed the three tests described in the previous section to verify the reconstruction quality of the resulting perceptual spaces: the Neighborhood test, the test against a random configuration, and the topology change test were repeated for two-dimensional perceptual spaces. Most importantly, the comparison between the visual and haptic perceptual spaces was performed for two-dimensional MDS solutions as well.

Results

As the elbow in Figure 5 shows, participants use three main dimensions to form their perceptual space when all objects of the three-dimensional object space were compared to all other objects using the haptic modality. Figure 5 shows a higher noise level as Figure 3, since 231 comparisons had to be made instead of 84. This higher noise level makes it harder to determine the elbow, especially for the visual modality where the elbow can be either determined to lie at two or three dimensions. For easier comparison to the haptic data and especially since the data analysis of the previous experiments resulted in three two-dimensional perceptual spaces (see previous section), the following analysis first compares results for a three-dimensional solution.

The three-dimensional perceptual spaces were calculated for the visual and haptic modalities and were fit to the positions of the objects within the physical space using the Procrustes function. As described for the 2D topology in the previous section, the correspondence between the points of the physical and perceptual spaces was determined using the Neighborhood test. For the visual modality, 17 out of 21 objects showed direct correspondence. For the haptic modality, 18 objects showed direct correspondence. Given that the misassigned objects were surrounded by correctly assigned objects, we concluded that both for the visual and haptic modalities the topology and the neighborhood relations were correctly recovered. Given the highly complex objects that make up our three-dimensional physical object space (see Figure A3) and the large number of test trials involved, this result again points to a very good reconstruction quality.

We then calculated the distance between the positions of the stimuli in the perceptual spaces and the physical object space. For the visual perceptual space, the procrustes value was 0.186. The fit was even better for the haptic perceptual space (0.147). The procrustes values
of the random spaces (21 points were distributed 500 times within a sphere that had a diameter of 6) again were highly different from those of the observed spaces (median \(_{\text{random}} = 0.900\), \(p_{\text{visual}} = 0.000\), \(p_{\text{haptic}} = 0.000\), Figure A6). In the topology change test, in which the 21 objects were randomly distributed along three orthogonal Y-shaped configurations, we also found highly different procrustes values compared to those of the observed perceptual spaces (median \(_Y = 0.898\), \(p_{\text{visual}} = 0.000\), \(p_{\text{haptic}} = 0.000\)).

Finally, we tested how the procrustes values would change when participants would “change” the order of neighboring points. Again, this analysis led to a significant increase in the procrustes values both for the visual and haptic modalities (median \(_{\text{visual}} = 0.219\), \(p = 0.000\), median \(_{\text{haptic}} = 0.181\), \(p = 0.000\)) and shows (as already stated in the 2D topology section) that participants also locally identified the neighborhood relations correctly.

After comparing the perceptual spaces with the physical object space, the two perceptual spaces were compared, by calculating the distance between the two spaces. The procrustes value was very low (0.088) and thus the distance between the two perceptual spaces is even smaller than the distance of the visual or the haptic perceptual space to the physical space (0.186 and 0.147, respectively). Considering the fact that the visual modality is often considered as superior to the haptic modality, the similarity of the two perceptual spaces is quite large.

Since the planar configurations of the three planes were analyzed in detail in the previous section, we wanted to know how these three planes were integrated to form a three-dimensional perceptual space. Therefore, we took the visual and haptic three-dimensional MDS and fitted planes 1, 2, and 3 into the space and calculated the angles between the planes (see Movies 1 and 2). In the haptic modality, the reproduction of the three-dimensional physical object space is better, stressing again the surprising fidelity of the haptic modality in perceiving the underlying physical shape changes: Planes 1 and 2 span an angle of 39°, while planes 1 and 3 and planes 2 and 3 are almost perfectly orthogonal to each other (99° and 87°, respectively). In the visual perceptual space, however, plane 1 is tilted to plane 2 by 64°, plane 1 is tilted to plane 3 by 47°, and planes 2 and 3 are separated by an angle of 68°. Whereas in the haptic modality, the three stimulus dimensions are more clearly recovered, it seems that in the visual domain the separation of the dimensions is less clear. Although the three planes are not perfectly orthogonal, the resulting quality of the reconstruction suggests that humans are able to form a three-dimensional perceptual space when they fully explore the complex shell-shaped objects haptically, and that the third dimension also helps to recover the topology of the object space when exploring the objects visually.

Indeed, Figure 5 shows that the stress values can also support a two-dimensional solution for the visual data. Because of this, we repeated the analyses for two-dimensional perceptual spaces. The fit quality between the visual perceptual space (0.182) is again significantly better than for randomly distributed points (median \(_{\text{random}} = 0.937\), \(p = 0.000\)) and also significantly better than for random Y-shaped configurations (median \(_Y = 0.935\), \(p = 0.000\)). Introducing changes into the topology also led to
significantly higher procrustes values (median\textsubscript{shuffled} = 0.225, \(p = 0.000\)). The neighborhood test was performed comparing always seven points per plane and recovered 16 of 21 correctly assigned objects, with the misassigned objects surrounded by correctly assigned objects. Thus, the two-dimensional perceptual space in the visual domain still recovers the Y-shaped configurations correctly. For the two-dimensional haptic perceptual space, we also found a better fit than for 21 arbitrarily distributed points as well as random Y-shaped configurations (median\textsubscript{random} = 0.937, \(p = 0.000\), median\textsubscript{Y} = 0.935, \(p = 0.000\)). For changes in the order on the Y-shaped configuration (median\textsubscript{shuffled} = 0.224, \(p = 0.238\)), however, the fit did not seem better. In the neighborhood test, 16 out of 21 objects were correctly assigned. Among the other 5 objects, however, the immediate neighborhood was less well preserved, indicating a slightly worse reconstruction of the full topology in the two-dimensional haptic perceptual space.

Nevertheless, comparing the visual and the haptic two-dimensional spaces resulted in a low procrustes value of 0.042. Hence, the two-dimensional MDS solutions of the visual and the haptic perceptual spaces are, similarly to the three-dimensional reconstructions, also highly congruent.

We can therefore conclude that the haptic modality is able to recover the topology of the physical space surprisingly well. The quality of the haptic reconstruction seems even better than that of the visual modality, as the visual stress values do not point to a three-dimensional perceptual space as clearly as the haptic stress values and as both two-dimensional and three-dimensional visual reconstructions have some (small) topological errors. Most importantly, however, regardless of whether we compared two- or three-dimensional MDS solutions, the visual and the haptic modalities result in very similar perceptual spaces, pointing to a highly similar shape percept for the two modalities.

Discussion

In five experiments, the perceptual representations of a parametrically defined, three-dimensional object space were analyzed. Whether participants explored 2D pictures or 3D virtual objects visually, or whether they explored 3D plastic models haptically, we found that the topology of the underlying physical object space was represented with a surprising fidelity.

Our study takes inspiration from the work of Cutzu and Edelman (1998) who showed that vision can fully recover planar configurations of a multidimensional parameter space and thus supports metrically veridical representations of similarities among 3D objects. In Experiment 1, we were able to reproduce this finding using our three-dimensional object space of complex shell-shaped objects whose Y-shaped configurations were recovered surprisingly well. In addition, the study of Cutzu and Edelman used only static pictures or passively rotated objects. In Experiment 2, we used a virtual reality setup and were able to show that self-determined exploration also leads to a veridical representation of the object space.

In this context, we would like to discuss why we found only slight differences between the passive and active exploration conditions in our experiments, especially for plane 1 and plane 2. Object motion (such as rotation around an axis) can recover the 3D shape of an object via a process called shape-from-motion (Ullman, 1979). Harman, Humphrey, and Goodale (1999) showed that active control of the visual input facilitates later recognition of objects compared to passive exploration. In our experiments, however, participants did not seem to benefit from the additional information that was available during active exploration. One possible reason for this might be that our setup was not able to properly convey the relevant depth information due to, for example, faulty tracking or wrong calibration of the HMD. However, no participant reported to have problems with the setup and we took care to adjust the HMD to each participant minimizing any errors due to the hardware setup. Another reason could be that the 2D pictures of the shells show the shape features so clearly that all information necessary to form a veridical perceptual space were available. Indeed, humans often rely on the impression of several planar views and seem to even ignore other views (James, Humphrey, & Goodale, 2001). It is possible that the pictures we presented in the visual 2D experiments show the same view that participants paid most attention to when they freely explored the objects and thus did not benefit from the other views (Stone, 1999). We thus suggest as the most likely explanation that humans, based on their prior knowledge about shape in general (Langer & Bülthoff, 2001) or even about shell-like objects in particular, can form a reasonably accurate shape representation already from the 2D pictures we used. Active exploration of the 3D shapes, however, does seem to help in recovering the topology of plane 3, which is the least intuitive plane. It remains for future work to investigate the influence of these view-based factors in more detail.

Given that visual object exploration recovers the Y-shaped topology with a high fidelity, in Experiment 3 we wanted to analyze whether haptic exploration might also be able to recover this complex topology. Cooke et al. (2007) showed that the haptic modality can form an accurate perceptual representation when the underlying physical object space consists of two dimensions. However, the objects used in their study varied in texture and shape—two rather intuitive dimensions for both vision and haptics. With the shell-shaped objects, we used complex shape parameters that are neither intuitive to vision nor to haptics, and additionally, parameters differed along three dimensions. Despite this increased complexity, participants were still...
able to form a perceptual space based on haptic exploration that has a high degree of similarity to the physical object space.

Both the visual and haptic modalities showed good reconstruction quality in recovering the planar Y-shaped configurations of the three planes. In Experiments 4 and 5, we wanted to test whether this finding still holds when the whole three-dimensional object space is explored. Surprisingly, the haptic modality recovered the three-dimensional structure slightly better than the visual modality, shown by the good fit of the three-dimensional perceptual space to the physical space. In addition, the stress value clearly points to a three-dimensional perceptual space for the haptic modality, while the visual data could be explained by two- or three-dimensional perceptual spaces based on the stress value. This is also evident in the weighting of the three shape parameters in each modality: Haptically, the first, and thus most important dimension of the perceptual space, is $\epsilon^{cota}$, the number of convolutions (mostly referred to as “bulkiness” by the participants), while visually $\sin \beta$, or the symmetry, corresponds to the first dimension. Haptically, the second dimension corresponds to $\sin \beta$, while visually it is $\epsilon^{cota}$. For both modalities, parameter $A$, the distance between aperture and tip, corresponds to the third dimension. This parameter is well recovered in the haptic data, whereas for the visual modality this parameter plays only a minor role given that two dimensions might already be enough to explain the data. When going to the three-dimensional solution, however, the third dimension does result in a more faithful reconstruction of the physical space. Since we found that active exploration of the visual stimuli helped to better recover plane 3, which is varying along this parameter, such a condition might also result in a better, fully three-dimensional visual reconstruction. Further experiments will need to be run to confirm this hypothesis.

Following Shepard (2001), neural circuits of the brain have been shaped by natural selection specifically to provide a veridical representation of the world. Continuing along this line of argumentation, we can observe that the visual and the haptic sensory systems (as well as the other sensory modalities) co-evolved in the same physical environment while striking a balance between orthogonality of the measured physical properties of the world and redundancy: Orthogonality is needed to extend the amount of information about the world our brain can receive. Redundancy—that is, being able to measure similar physical properties across two modalities—not only guarantees continued functioning in case of failure in one modality but also the capability to put the different sensory modalities into correspondence (the so-called “binding problem”; Calvert, Campbell, & Brammer, 2000). The highly similar representations of the physical object space by the visual as well as the haptic domain and by other species (Sugihara, Edelman, & Tanaka, 1998) provide additional evidence for Shepard’s theory that there are “universal laws” in perception that enable similar generalization capabilities in modalities (or species) when they evolved in the same environment.

Comparing the visual and haptic perceptual spaces, the high similarity of the recovered topologies might lead one to ask whether touch and vision in this case might form one common representation that is shared between the modalities. Given that our experiments and the ones by Cooke et al. (2007) both resulted in highly congruent haptic and visual spaces, we might indeed assume that one underlying perceptual space is formed that is accessible to both modalities. Excellent cross-modal priming behavior observed between vision and haptics also points into this direction (Bushnell & Baxt, 1999; Reales & Ballesteros, 1999). If one multimodal representation is formed, however, one might expect cross-modal shape comparisons to be equivalent in performance to unimodal shape comparisons. In this context, Norman, Norman, Clayton, Lianeckhammy, and Zielke (2004) reported high accuracy but also significant performance differences between cross-modal and unimodal shape comparisons, concluding functionally overlapping but maybe distinguishable representations. Similarly, experiments on haptic face recognition have found asymmetric transfer between vision and haptics (Dopjans, Wallraven, & Bülthoff, 2009). Further psychophysical experiments will have to show the degree to which the perceptual representation might be shared between the two modalities.

Although the exact nature of the object representations is not yet clear, it is known that visual and haptic object representations activate overlapping brain regions as shown in several neuroimaging studies (Amedi, Malach, Hendler, Peled, & Zohary, 2001; James et al., 2002; Reed, Shoham, & Halgren, 2004). The lateral occipital complex (LOC) shows activation during visual and tactile object recognition. Visual imagery may be a possible explanation but leads to less activation than either tactile or visual object recognition (Amedi et al., 2001) leading to the assumption that the LOC is a multimodal area that is involved in perceptual representations of object shape (Kourtzi, Erb, Grodd, & Bülthoff, 2003). In addition, similar objects evoke similar response patterns in LOC whereas shapes perceived as more different are associated also with more different response patterns (Op de Beeck, Torfs, & Wagemans, 2008), which points toward a possible implementation of how the perceptual space might be represented.

Multidimensional scaling tries to embed stimuli in a space, in which stimuli perceived as very similar are close to each other, and stimuli that are perceived as less similar are farther apart. Again, following Shepard’s universal law of generalization, the generalization between objects increases with decreasing distance in perceptual space (Shepard, 1987). At the same time, objects that are close in perceptual space are more likely to be confused than objects far away from each other. Following DiCarlo and
Cox (2007) one might envision the representation of single objects or categories of objects within this space to be separated by hyperplanes (see also Jäkel, Schölkopf, & Wichmann, 2008)—this could be done by assigning a collection of views to the corresponding object and subsequently assigning similar shapes to the corresponding category (see also Edelman, 1999; Graf, Schwaninger, Wallraven, & Bülthoff, 2002). Knowing the structure of such a perceptual space should therefore make it possible to predict how a set of objects will be categorized by humans. In the study of Cooke et al. (2007), for example, it was shown that human categorization behavior correlated with the predictions based on the perceptual space. Again, objects varied in shape and texture and categorization was prone to follow easily verbalized rules. In contrast, the shell-shaped objects are complex and hard to describe (several participants reported that they could not describe the object structure in detail). It will therefore be more difficult to find and successfully predict reproducible categorization boundaries. The computational modeling of these categories based on the existing similarity data from the present experiments is one of the current research topics that we are pursuing.

The shell-shaped stimuli described in this paper cannot only be used to better understand the link between perceptual spaces and categorization behavior; it will also be possible to investigate another conjecture by Shepard (2001) that states that categories should form connected regions (manifolds) in representational space. Following his theory, morphing two objects of the same category into each other, no object can be generated that is outside of this category. As soon as we know how humans categorize the natural, but unfamiliar shell-shaped objects, we can generate intermediate morphs and test this conjecture and the degree to which it holds for both visual and haptic processing.

In this study, we showed that humans can form veridical perceptual spaces from planar configurations of complex shell-shaped objects for both passive and active visual explorations. Moreover, we showed that the haptic modality can reproduce the underlying physical object space with high fidelity and that, overall, the reconstructed, perceptual topologies were highly similar in both modalities. Surprisingly, the haptic modality even slightly exceeded the visual modality in recovering the structure of the underlying object space when the whole three-dimensional space was presented both concerning the correct identification of the dimensionality, as well as the faithfulness of the topology recovery.

We conclude that haptic shape processing of natural, complex shapes can, indeed, compete with visual shape processing. This finding is supported by the fact that recognition of highly familiar objects is almost as good in haptics as in vision (Klatzky et al., 1985) and that even unfamiliar but natural objects can easily be discriminated haptically (Norman, Clayton, Norman, & Crabtree, 2008). The sense of touch thus seems to be even more suited for shape processing than previously thought—even though humans use haptic shape identification less often than visual shape identification in everyday life. By reevaluating the importance of the haptic modality and its potential cross-talk to vision, we can think about novel application areas such as human–machine interfaces, training of shape perception, and even teaching of alphabets to children using haptics (Bara, Gentaz, Cole, & Sprenger-Charolles, 2004).

Appendix A

Figure A2. Overlay of all planes: The outermost contour line of all stimuli of all three planes was plotted into one figure per experiment to visualize the overlap of the confidence areas. On average, the overlap of the confidence areas is 30%. Plane 1: red, plane 2: green, plane 3: blue. The difference of the topology of every single plane is clearly visible.

Figure A3. Object images: For the experiments, 21 three-dimensional objects were generated. For every object, one image was rendered showing the object features clearly. The images presented in the visual 2D planes and the visual 2D space experiments are shown here. Objects are not sorted according to stimulus number to illustrate the difficulty of extracting shape parameters. When participants were asked to sort the objects of one plane according to their position in the physical object space, less than 30% were able to correctly recover the Y-shaped form, although the MDS output maps clearly show that participants are implicitly aware of two shape parameters. (The first row shows all objects of plane 1: 1, 6, 7, 2, 4, 3, 5. The second row shows all objects of plane 2: 10, 14, 8, 13, 12, 9, 11. The third row shows all objects of plane 3: 18, 16, 17, 19, 15, 20, 21. Stimulus numbers from left to right. Compare to Figure 2 for object positions within the physical object space.)
Figure A4. (a) Correlation between visual and haptic similarity ratings in the *planes* conditions. To show the high correlation of visual and haptic similarity perceptions, the similarity ratings of visual 2D planes were averaged over participants and correlated with the similarity ratings of haptic 3D planes, which were averaged over participants as well. (b) Correlation between visual and haptic similarity ratings in the *space* conditions. To show the high correlation of visual and haptic similarity perceptions, the similarity ratings of visual 2D space were averaged over participants and correlated with the similarity ratings of haptic 3D space, which were averaged over participants as well.
Three-dimensional topology: The variation across participants was analyzed by generating perturbed MDS. These were fit to the physical space using the Procrustes function. The corresponding procrustes values are displayed here. MDS: procrustes value of the visual (0.186, red cross) and the haptic (0.147, blue cross) perceptual spaces to the physical object space. Shuffled: every object of the perceptual spaces was exchanged once with its direct neighbor, the procrustes values between the resulting shuffled perceptual spaces and the physical space are shown. Y: 21 objects were randomly distributed 500 times on three orthogonal Y-shaped forms. Random: 21 objects were randomly distributed 500 times within a sphere with a diameter of 6. The perceptual spaces were significantly different from the random distribution as well as from the Y-shaped distributions showing that participants did not recover the topology by chance. Moreover, when the visual perceptual space is compared to the shuffled visual perceptual space and the haptic perceptual space is compared to the shuffled haptic perceptual space, the effect is also highly significant, showing that participants correctly recovered the neighborhood relations of the objects (red line: median, blue box: lower and upper quartile values, whiskers: most extreme values).

Figure A6. Three-dimensional topology: The variation across participants was analyzed by generating perturbed MDS. These were fit to the physical space using the Procrustes function. The corresponding procrustes values are displayed here. MDS: procrustes value of the visual (0.186, red cross) and the haptic (0.147, blue cross) perceptual spaces to the physical object space. Shuffled: every object of the perceptual spaces was exchanged once with its direct neighbor, the procrustes values between the resulting shuffled perceptual spaces and the physical space are shown. Y: 21 objects were randomly distributed 500 times on three orthogonal Y-shaped forms. Random: 21 objects were randomly distributed 500 times within a sphere with a diameter of 6. The perceptual spaces were significantly different from the random distribution as well as from the Y-shaped distributions showing that participants did not recover the topology by chance. Moreover, when the visual perceptual space is compared to the shuffled visual perceptual space and the haptic perceptual space is compared to the shuffled haptic perceptual space, the effect is also highly significant, showing that participants correctly recovered the neighborhood relations of the objects (red line: median, blue box: lower and upper quartile values, whiskers: most extreme values).
Acknowledgments

The authors would like to thank M. Breidt for assistance in stimulus production, B. Arsene for help in carrying out the haptic experiments, and J. Schultz for helpful comments on software and the manuscript. This manuscript has benefited from conscientious peer review.

This research was supported by a PhD stipend from the Max Planck Society. Part of this research was also supported by the World Class University (WCU) program through the National Research Foundation of Korea funded by the Ministry of Education, Science and Technology (R31-2008-000-10008-0).

Commercial relationships: none.
Corresponding author: Christian Wallraven.
Email: wallraven@korea.ac.kr.

References


Jäkel, F., Schölkopf, B., & Wichmann, F. A. (2008). Generalization and similarity in exemplar models of


