Combination and Integration in the Perception of Visual-Haptic Compliance Information

Martin Kuschel, Massimiliano Di Luca, Martin Buss, and Roberta L. Klatzky

Abstract—The compliance of a material can be conveyed through mechanical interactions in a virtual environment and perceived through both visual and haptic cues. We investigated this basic aspect of perception. In two experiments, subjects performed compliance discriminations, and the mean perceptual estimate (PSE) and the perceptual standard deviation (proportional to JND) were derived from psychophysical functions. Experiment 1 supported a model in which each modality acted independently to produce a compliance estimate, and the two estimates were then integrated to produce an overall value. Experiment 2 tested three mathematical models of the integration process. The data ruled out exclusive reliance on the more reliable modality and stochastic selection of one modality. Instead, the results supported an integration process that constitutes a weighted summation of two random variables, which are defined by the single modality estimates. The model subsumes optimal fusion but provided valid predictions also if the weights were not optimal. Weights were optimal (i.e., minimized variance) when vision and haptic inputs were congruent, but not when they were incongruent.

Index Terms—Perception, haptic, vision, integration, optimal fusion, combination, human system interface, virtual reality.

1 Introduction

The mechanical properties of an environment (such as compliance/stiffness, density/damping, or center of mass/inertia) are perceived during active manipulation primarily by processing position-based information (position, velocity, acceleration, jerk, etc.) and force cues. See Fig. 1 for an illustration. Position-based cues are provided by both visual and haptic modalities, for example, by the visible displacement of the fingers pressing into a surface and by kinesthetic cues to the change in joint angle. Force cues are primarily haptic, in the form of kinesthetically sensed resistance, although visual cues may arise, for example, from visible surface deformation. A model that accurately captures the perceptual process by which position and force are utilized to produce a compliance judgment will contribute not only to psychophysical knowledge, but also to the design and control of human system interfaces used to access virtual or remote environments.

Psychophysical research into visual-haptic perceptual interactions has frequently shown that both modalities influence perceptual outcomes, but with differential impact, e.g., [1], [2], [3]. Various approaches argue that the relative contribution of each modality depends on modality appropriateness [4], effectiveness [5], or the direction of attention [6]. In [7], it was proposed that bimodal estimates of the same property can be integrated optimally, such that the final percept is a weighted average that minimizes the variance (i.e., maximizes reliability) of the estimate. The optimal fusion model has been broadly invoked, e.g., [7], [8], [9], [10], although exceptions do occur [8].

Most research on bimodal perception has focused on how redundant estimates of the same parameter (e.g., position as estimated by the visual and the haptic modalities) can be fused into a unitary estimate, a process we call *integration*. The estimation of properties that depend jointly on nonredundant, complementary information, which we call *combination*, has generally not been addressed in detail. The perceptual estimation of compliance through haptic exploration involves a combination of position and force information. According to *Hooke's law*, the compliance S is expressed as the division of the indentation x by the force f:

$$S = \frac{x}{f}. \tag{1}$$

Hence, perception of compliance requires processing the combination of both position and force information; the two variables are nonredundant. Of course, when there are redundant estimates of compliance, integration of those estimates could also occur.

Various studies have investigated people's ability to perceive compliance when only haptic information is available (e.g., [11], [12], [13], [14], [15], [16]). These generally conclude that performance in estimating compliance is considerably worse than in estimating position or force from haptic cues. Weber fractions ranging from 12 to 35 percent have been reported for unimodal estimation of compliance. In these experiments, real compliant objects, such as rubber specimens, or virtual objects rendered using a haptic human system interface (HSI), are manipulated without vision. Multimodal studies, which involve the visual modality as well as the haptic one, are rare.

- *M. Kuschel and M. Buss are with the Institute of Automatic Control Engineering (LSR), Technische Universität München, 80333 Munich, Germany. E-mail: martin.kuschel@tum.de, mb@tum.de.*
- *M. Di Luca is with the Max Planck Institute for Biological Cybernetics, Dept Bülthoff, NWG Ernst, 72076 Tübingen, Germany. E-mail: max@tuebingen.mpg.de.*
- *R.L. Klatzky is with the Department of Psychology, Carnegie Mellon University, Pittsburgh, PA 15213. E-mail: klatzky@cmu.edu.*

Manuscript received 29 Sept. 2008; revised 31 May 2009; accepted 1 Feb. 2010; published online 17 Mar. 2010.
Recommended for acceptance by H.Z. Tan.
For information on obtaining reprints of this article, please send E-mail to: toh@computer.org, and reference IEEECS Log Number TH-2008-09-0060.
Digital Object Identifier no. 10.1109/ToH.2010.9.

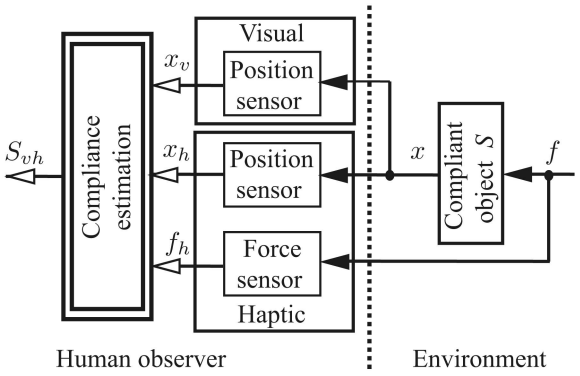

Fig. 1. Visual-haptic perception of compliance S: position information x is measured by the visual and the haptic modality. Force information f is only provided by the haptic modality. The final percept S_{vh} can be obtained by combination and integration processes (Fig. 2).

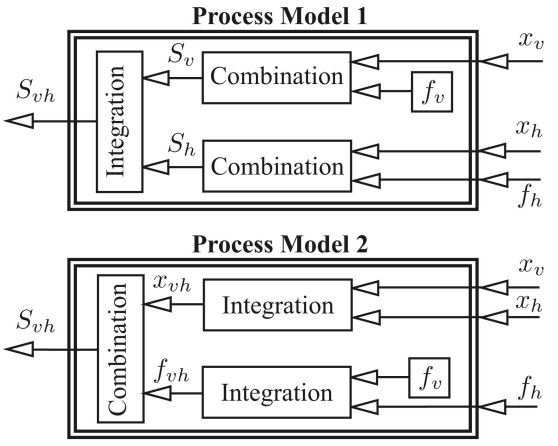

Fig. 2 The two process models of visual-haptic compliance estimation (double lined box of Fig. 1): in Process Model 1, unimodal compliance estimates are integrated to a bimodal percept (combination before integration). In Process Model 2, bimodal position estimates are combined with bimodal force estimates (integration before combination). (The visual force is marked as an internal source as it cannot be sensed directly.)

The first study that analyzed the influence of visual information on haptic compliance perception reported a significant influence of visual information on the final percept [17]. Recent studies extended this result by analyzing cross-modal interactions [18] and discrimination performance [19], [20]. While haptic information is sufficient to perceive compliance (force and position cues are provided), it seems not to be a necessary condition. In [19], evidence was provided for a compliance estimate based entirely on vision. However, this requires visual force information. Since humans have no visual force sensor, the visual force information must be inferred. When a surface is compliant, the deformation pattern (i.e., the profile of the indentation at the contact point) provides a cue to force, e.g., under the assumption that greater force means indentation spreads further from the contact point. Another cue would be the material from which the object appears to be made, e.g., a matte surface might be associated with a compliant material, leading to discounting of the indentation cues, and a shiny surface might be associated with a rigid material, accentuating indentation cues. The appearance of an indentation tool might also be used, for example, a larger or shinier tool might be associated with higher force.

This paper proposes a mathematical model for visual-haptic perception of compliant objects, based on two psychophysical experiments. In **Experiment 1**, the order of combination and integration processes was analyzed. The result indicated that initially each modality combines position and force into a unimodal compliance estimate; then, in a successive step, the unimodal estimates are integrated into a bimodal estimate. In **Experiment 2**, the integration process was analyzed in detail. The result indicated that visual and haptic compliance estimates are integrated based on a weighted summation process using weights that are not optimal if conflicts occur.

2 Models for Visual-Haptic Perception of Compliance

2.1 Process Models

As a theoretical basis for the first experiment, we consider process models that describe the perception of compliant information. These differ in the order of the processes that combine and integrate visual and haptic information, as depicted in Fig. 2. Process Model 1 assumes that in a first step, each modality provides a compliance estimate by combining nonredundant position information x and force information f. (As noted above, there is evidence that force information is cued at least to some extent by vision [20].) In a second step, the visual compliance estimate S_v and the haptic compliance estimate S_h are integrated to determine the final percept S_{vh}. Process Model 2 assumes that, in a first step, redundant position information is integrated to produce a composite position estimate x_{vh} and redundant force information is integrated to produce a force estimate f_{vh}. In a second step, position and force estimates are combined to create the final percept of compliance S_{vh}.

2.2 Integration Models

As a theoretical basis for the second experiment, we consider three mechanisms for estimating compliance from independent, redundant sources, called *integration*.

2.2.1 General Assumptions

We assume that the single modality estimates are Gaussian distributed and denoted by the random variables

$$\begin{aligned} X_v &\sim N(\mu_v, \sigma_v), \\ X_h &\sim N(\mu_h, \sigma_h), \end{aligned} \tag{2}$$

where μ is the *mean perceptual estimate* and σ the *perceptual standard deviation*. We further assume that the final estimate is captured by the random variable

$$X_{vh} \sim f(\mu_{vh}, \sigma_{vh}), \tag{3}$$

which is not necessarily Gaussian distributed. The reliability of an estimate or a percept is

$$r = \frac{1}{\sigma^2}. \tag{4}$$

Fusion of multimodal information is said to occur if the reliability of the multimodal estimate is greater than the reliability of the most reliable single modality estimate. That

is, the perceptual standard deviation of the final percept is smaller than the smallest perceptual standard deviation among those provided by the single modality estimates. *Confusion* is said to occur if the perceptual standard deviation of at least one unimodal estimate is smaller than the perceptual standard deviation of the multimodal percept. In this case, the observer is confused by the different sources of information. *No fusion/confusion* is said to occur if the reliability of the final percept is equal to the reliability of the most reliable single modality estimate. This is consistent with the judgment relying on this modality. Note that the terms "fusion" and "confusion" are not defined by a particular model but only by the relation between the unimodal and the multimodal perceptual standard deviations.

2.2.2 Sensory Capture

The trivial way to deal with two sources of redundant information is to disregard one source and concentrate on the other, which will produce the result called above "no fusion/confusion". Here, the perceptual system concentrates on the most reliable modality and the bimodal percept is equal to the estimate of this modality. The process is called *sensory capture*. The model is described by

$$\boxed{X_{vh} = X_q.} \tag{5}$$

Here, the subscript q is either h or v, representing the most reliable modality, haptics or vision.

2.2.3 Weighted Summation

The second basic way to deal with redundant information is to generate a weighted average of the different random variables defined by the different estimates. We call this process *weighted summation* of redundant information. For a bimodal percept the model is described by

$$\boxed{\begin{array}{lcl} X_{vh} & = & w_v X_v + w_h X_h, \\ w_v + w_h & = & 1. \end{array}} \tag{6}$$

The mean of the joint distribution is

$$\mu_{vh} = w_v \mu_v + w_h \mu_h. \tag{7}$$

And the variance becomes

$$\sigma_{vh}^2 = (w_v \sigma_v)^2 + (w_h \sigma_h)^2. \tag{8}$$

Depending on the weights, the model can capture fusion, no fusion/confusion, and confusion. Maximum confusion occurs when the reliability of the bimodal percept is equal to that of the least reliable sensory estimate. Calculating the weights w_v, w_h to minimize the variance according to $\frac{\partial \sigma_{vh}^2}{\partial w_v} = 0$ yields a final estimate with maximal reliability

$$\sigma_{vh}^2 = \frac{\sigma_v^2 \sigma_h^2}{\sigma_v^2 + \sigma_h^2}. \tag{9}$$

Thereby, the optimal weights are only based on the single modality variances[1]

1. This model represents the Kalman filter for the fusion of two static sources.

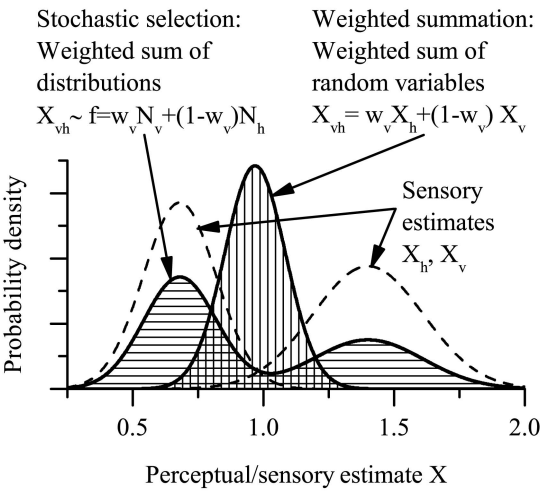

Fig. 3. Nontrivial integration processes of two sensory estimates: under the weighted summation model, the perceptual system performs a convolution of the underlying weighted distributions, which will be Gaussian if the component distributions are Gaussian. Under the stochastic selection model, the perceptual system draws an estimate from one of the two distributions with a certain probability, producing a non-Gaussian distribution of estimates. Both models lead to the same mean value.

$$w_v^* = \frac{\sigma_h^2}{\sigma_v^2 + \sigma_h^2}, \quad w_h^* = \frac{\sigma_v^2}{\sigma_v^2 + \sigma_h^2}. \tag{10}$$

A resulting distribution from a weighted summation process, in this case optimal fusion, is depicted in Fig. 3.

2.2.4 Stochastic Selection

A third basic way to deal with redundant information is to randomly draw with a certain probability from either one of the available sources. Hence, the distribution of the final percept is equal to the weighted sum of the distribution values provided by the sensory estimates. We call this process *stochastic selection* among redundant information. For a bimodal percept, the stochastic selection model is described by

$$\boxed{\begin{array}{lcl} X_{vh} \sim f & = & w_v N(\mu_v, \sigma_v) + w_h N(\mu_h, \sigma_h), \\ w_v + w_h & = & 1. \end{array}} \tag{11}$$

The mean of the joint distribution is again a weighted average

$$\mu_{vh} = w_v \mu_v + w_h \mu_h. \tag{12}$$

The variance is based on the variances of the single modality estimates and on the conflict between the two estimates, i.e., on the difference between the mean values

$$\sigma_{vh}^2 = w_v \sigma_v^2 + w_h \sigma_h^2 + w_v w_h (\mu_v - \mu_h)^2. \tag{13}$$

The stochastic selection model can produce results consistent with the outcomes of no fusion/confusion and confusion. In the case of congruent bimodal information (i.e., no distortion of one modality relative to the other), $\mu_h = \mu_v$; maximum confusion occurs when the reliability of the bimodal percept is equal to that of the less reliable sensory estimate. Hence, fusion is not possible with this model. A resulting distribution is depicted in Fig. 3.

3 Experiment 1: Order of Integration versus Combination

To determine which of the process models illustrated in Fig. 2 applies to visual-haptic compliance perception, an experiment was conducted with two conditions. In the first condition, called *active*, participants made a perceptual decision about the compliance of a virtual cube that was explored by pushing the right thumb into its surface. The resulting compliance information I_a was

$$I_a = (x_v, x_h, f_v, f_h). \tag{14}$$

In the second condition, called *resistive*, participants made a perceptual decision about the compliance of a virtual cube that was pushed against the right thumb by the robot. This condition was intended to eliminate haptic position cues and visual force cues. The resulting information I_r was then

$$I_r = (x_v, f_h). \tag{15}$$

If Process Model 1 applies, and perceived compliance is obtained first by unimodal estimates that are then integrated, a compliance percept is possible for the active condition (14) but impossible for the resistive condition (15). This is because by lacking haptic distance and visual force cues neither a haptic-only estimate S_h nor a visual-only estimate S_v can be achieved. On the other hand, if Process Model 2 applies, a reliable compliance percept is possible for both the active condition (14) and the resistive condition (15), since the perceptual system combines force and position information irrespective of the modality they come from.

Since it was difficult to fully eliminate x_h in the experiment (see Section 3.1.3 for a detailed description of the stimuli), a percept was possible in both conditions for both models but presumably differed in the reliability. Hence, we predicted the standard deviation σ of the percepts, which inversely correlates with the reliability. The hypotheses for the experiment were then as follows:

If Process Model 1 holds, the standard deviation σ of the human sensor in condition (15) should be much larger than in condition (14)

$$\sigma_r \overset{!}{>>} \sigma_a, \tag{16}$$

where the indices "r" and "a" denote the resistive and the active condition, respectively. The exclamation mark indicates that this inequality must be true if the hypothesis should not be falsified. If Model 2 holds, the perceptual standard deviation for both conditions should not differ

$$\sigma_r \overset{!}{\approx} \sigma_a. \tag{17}$$

The latter prediction emerges from the assumption that the additional information (x_h, f_v) provided in the active condition (14) is much less reliable, i.e., has a much larger perceptual standard deviation, than the information (x_v, f_h), which is provided in both conditions. Comparisons of JNDs for visual and kinesthetic position sensing [21], [22], [23] indicate that the visual estimation of the position x_v has a much smaller perceptual standard deviation than the haptic estimation of the position (x_h). (However, it should be noted that some studies have found that the relative reliability of vision and haptics is task dependent; visual position may be particularly unreliable in the z-direction of the horizontal plane [24], [25].) Furthermore, estimating force visually f_v should provide little additional information, since force information must be inferred indirectly from visual cues, which were not provided in our set-up. Hence, using the additional information (x_h, f_v) should lead to no decrease or only to a small decrease of the perceptual standard deviation in the active condition compared to the resistive condition (15), if Process Model 2 is applied.

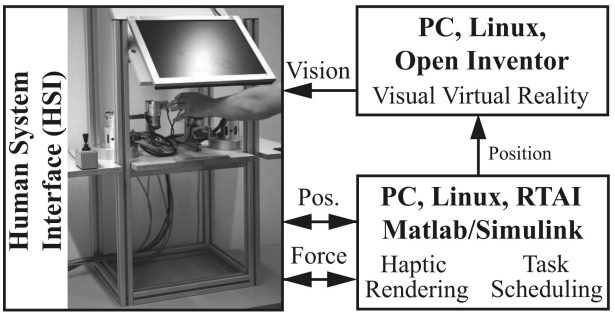

Fig. 4 Experimental set-up: the system consists of a HSI and real-time processing unit. Positions and forces of the SCARA tool tip are measured. The response box can be seen on the left.

3.1 Method

3.1.1 Visual-Haptic Virtual Reality

Hardware and Software: Haptic rendering of compliant objects was obtained using a HSI comprising two self-made selective compliance assembly robot arm (SCARA)-robots, a real-time processing unit, a visual display, and a response box (see Fig. 4). Participants grasped the end effectors of the two SCARA robots between the index and thumb fingers and squeezed them in order to estimate compliance. The SCARAs were built using high fidelity components including *Maxon* motors and *Harmonic Drive* gears enabling very precise control of the interaction. The workspace is approximately 80 mm and the maximum force is approximately 50 N. Position information was measured by angle encoders and force was sensed by strain gauges attached to both robot links. This information was recorded by a *Sensoray626* DAQ-card providing 16-bit resolution. Signal processing algorithms were implemented as *Matlab/Simulink* models with real-time code generated automatically. The system operated at 1 kHz sampling frequency, and it was connected to a PC running real-time application interface *(RTAI)* for Linux. Measured positions were transferred to a second PC running the visual VR programmed in *Open Inventor*. The grasped object was rendered as a gray cube squeezed by two orange spheres. The TFT screen used to display visual information was frontoparallel (slanted by 40 degrees with respect to the vertical direction) and mounted between the subject and the SCARA robots. In this way, from the participant's vantage point, the fingers and the spheres squeezing the cube were completely matched (see Fig. 5). During psychophysical experiments, participants reported their answers using a response box operated with their nondominant hand.

Dynamics and Control: The identical robots of the HSI are controlled independently using the same control scheme.

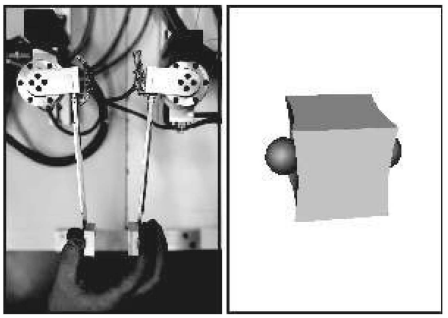

Fig. 5. Virtual stimuli: the haptic feedback renders a compliant cube to be explored by thumb and index finger. Visually, the fingers are represented by orange spheres.

The dynamics of the stimulus compliance is described by the admittance S: force input $\rightarrow$ position output, which represents compliance according to Hooke's law $x_s = Sf$, where S [mm/N] is the compliance whose perception is addressed in this study. The control concept employing inner position control driven by a VR with force reference is called *admittance control*. It is best suitable for rendering nonrigid environments like compliant environments. Minimum compliance (= maximum stiffness) that can be rendered is $S = 0.14$ mm/N. Due to the set-up, the transparency of the system was nearly ideal and the displayed dynamics S can be assumed to be equal to the commanded stimulus $S = S_d$.

3.1.2 Participants

Twenty (20) students of the Technische Universität München took part in a single, 2-h session for pay. Participants had an average age of 24 years and were naïve to the purpose of the study, although most were experienced psychophysical observers.

3.1.3 Stimuli

The stimuli were visually rendered compliant cubes with 80-mm sides. The compliance of the standard stimulus was $S_{ref} = 1.6\ \mathrm{mm/N}$. Eight (8) congruent bimodal comparison stimuli were implemented distributed among the standard stimulus $S = [0.7, 1, 1.2, 1.4, 1.8, 2, 2.2, 2.4]\ \mathrm{mm/N}$.

In the *active* condition, stimuli were provided according to (14). Participants could actively indent the virtual cube by squeezing with the thumb against the index finger. The index finger was fixed mechanically and the indentation was only performed with the thumb (with a maximal excursion of about 2.5 cm). On the visual display, participants saw an indentation tool (orange spheres) indenting a virtual cube (a screenshot of both displays is provided in Fig. 5).

In the *resistive* condition, stimuli were provided according to (15). Haptically, participants were told to maintain their thumb at a constant fixed position (index finger was again fixed and no force was exerted on it) as the robot pressed into their thumb, rendering the commanded compliance. Visually, the cube was pressed into the sphere that represented the thumb. Thereby, it decreased in size on the index finger side. The orange sphere representing the index finger remained at the same position. Since we had to avoid visual force cues, the cube remained flat and did not show a deformation curvature as in Fig. 5, when indented. To avoid haptic position cues, participants were informed by a beep if they moved their finger more than ± 5 mm from the fixed position and data were discarded, but lesser movements were possible. The cube trajectory was a sine-wave with randomized amplitude (ranging from 0.5 cm and 2.5 cm) and period of 1.0 s. On the visual display, participants saw the cube pressing/moving into their thumb representation (orange sphere).

3.1.4 Procedure

The psychophysical function for discriminating compliance was assessed by the method of constant stimuli. A two-conditions within-subject design was used. The active and resistive conditions were presented in different blocks that consisted of 8 comparisons and 10 repetitions. The sequence for each of the blocks was randomly chosen for each participant. Participants sat in front of the HSI, looking at the screen and grasping the device with their dominant hand. After being instructed, participants were trained using a set of stimuli which were not repeated in the experiment until they felt confident they understood the task. In total, $2 \times 20 = 40$ psychophysical functions were recorded.

One trial consists of the sequential presentation of two stimuli: a standard and a comparison stimulus. Participants explored each stimulus as directed, active or resistive, to assess their compliance of the two stimuli and reported which of the two appeared more compliant. The duration of each stimulus presentation was 2 s with an interstimulus interval of 2 s. The next trial began 2 s after the answer was given using the response box. Exploratory movements were generally quite variable in amplitude even within a trial. No systematic difference in exploration across the two conditions was observed. After the experiments, the participants judged the quality of the display and their immersion into the virtual reality and stated them in a written report. This was done to draw conclusions about the participants' ability in combining and integrating the information presented by our display.

3.1.5 Data Analysis

Psychophysical functions were fitted with a cumulative Gaussian function using "psignifit" software, version 2.5.6, described in [27] and Matlab/Simulink. The following parameters were computed by fitting the psychophysical function to the experimental data to describe the final percept S_{vh}: the point of subjective equality for the two conditions and the control condition was computed as the stimulus level at proportion 0.5 yielding $\mathrm{PSE} = \mathrm{F}_{\Psi}^{-1}(0.5)$, where F_{Ψ} denoted the cumulative Gaussian function. The just noticeable difference for each condition (JND_a, JND_r) was computed by taking the stimulus difference between PSE and proportion 0.84, yielding $\mathrm{JND} = \mathrm{F}_{\Psi}^{-1}(0.84) - \mathrm{PSE}$.

These empirical values are converted to theoretical estimates of the mean perceptual estimate, or average percept, and the variability around that estimate, as follows: the *mean perceptual estimate* was measured by the observed PSE

$$\mu = \mathrm{PSE}. \tag{18}$$

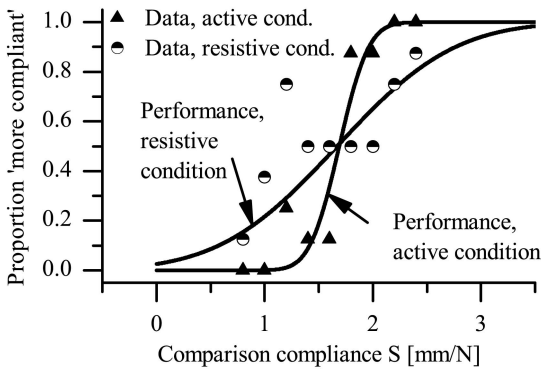

Fig. 6. Perception of visual-haptic compliance in the active and the resistive condition: shown for a sample participant are the experimental data and the psychophysical function. Active perception (14) led to better discrimination performance than resistive perception (15).

The *perceptual standard deviation* was calculated from the observed JND, treating it as the JND of a reminder task according to

$$\sigma = \sqrt{2}\mathrm{JND}, \tag{19}$$

where the multiplier $\sqrt{2}$ represented the performance correction for two-interval reminder tasks (method of constant stimuli).[2]

3.2 Results

Discriminating compliance with our visual-haptic virtual reality was found to produce psychophysical functions of the expected shape. (The data of only one participant had to be removed on this basis.) The data and the psychophysical functions of a single participant are depicted in Fig. 6. Participants' subjective post-experimental reports confirmed that the display created an immersive experience of the virtual object as a compliant cube, characterized by both visual and haptic cues. Although the presentation of the stimuli was quite abstract compared to real world stimuli, participants appeared to have no difficulty in treating the display as a compound visual-haptic stimulus.

A paired-sample t-test performed on the JNDs obtained in the two conditions indicated that the JND in the resistive condition was significantly larger than in the active condition ($t(18) = 3.32, p = 0.001$ one-tailed). For the active condition (14), the perceptual standard deviation σ_a resulted in 0.46 mm/N (28.91 percent relative to the standard stimulus S_{ref}). For the resistive condition (15), σ_r resulted in 1.32 mm/N (82.67 percent). The perceptual standard deviation in the resistive conditions was 2.86 times the perceptual standard deviation of the active condition.

3.3 Discussion

In general, the reliabilities of the compliance estimates were smaller than the reliabilities for the perception of position information or force information as reported in the literature, e.g., [11]. The large difference between the perceptual standard deviations can be explained only by Process Model 1 according to (16). In this model, perception of compliance involves combining nonredundant position and force information independently for each sensory modality, then integrating the different compliance estimates. To be effective, this process requires that each modality provides both position and force information, as in the active condition described above. These results raise the question of whether Process Model 1 underlies performance because it is advantageous from a processing perspective, relative to Model 2. Certain advantages can be suggested. One is that because Model 1 assumes the integration of independent estimates, compliance values that are directly inferred from other modalities or from experience can seamlessly be incorporated into its structure. A second advantage is that as each compliance estimate is sensor specific, the compliance percept could be computed directly from the sensor input. This allows for a more robust and less noisy transmission of the information to higher levels of processing. (This kind of preprocessing is mandatory in electrical measurement and signal processing.)

Our conclusion differs from the claim by Srinivasan et al. [17], namely, that humans' compliance percept is based "on the relationship between the visual position information and the indentation force sensed tactually" (p. 555). However, their experiment was in principle not suited to differentiate between the two ways to compute compliance information described by the process models. The observation that participants' ability to haptically perceive compliance decreased, when conflicting visual position information was provided concurrently, does not indicate whether compliance was estimated separately for each modality (Process Model 1) or whether visual and haptic displacements were integrated before compliance was estimated (Process Model 2).

The finding that compliance could be estimated to some extent in the resistive condition, under the assumptions of Process Model 1, can be explained by the fact that we did not eliminate finger movements, i.e., haptic position cues x_h, smaller than ± 5 mm. Combined with haptic force information f_h, a haptic compliance percept was apparently possible but had low reliability. In future work, it would be valuable to directly measure the JNDs for haptic position and visual force under the experimental conditions to assess their potential contributions to compliance.

4 Experiment 2: Visual-Haptic Integration of Compliance Information

The results of the first experiment supported the hypothesis that compliance perception is performed according to Process Model 1; that is, integration of visual compliance information and haptic compliance information takes place only after the combination of force and position to produce a modality-specific estimate of compliance. The second experiment was designed to test Process Model 1 in more detail. Specifically, it manipulated visual cues to position and force, holding haptic cues constant, and used psychophysical evaluation to determine how visual distortions affected the final compliance estimate. According to Process Model 1, the visual channel should use the distorted cues to

2. See [29], pp. 180-181 for a detailed description of the reminder paradigm. The design was not a 2AFC; hence, the performance correction was applied because participants were assumed to use a discrimination strategy when comparing between standard and comparison stimulus. This would be unnecessary with a decisionally separable strategy, where participants would essentially memorize the standard stimulus. However, given the complexity of the compliance percept, we assumed that participants did not ignore the standard stimulus but processed it anew on each trial.

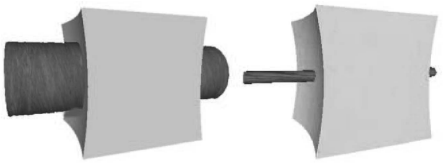

Fig. 7. Visual force: to generate different visual forces f_v, different indentation tools were used. The metal tool had a large diameter and a metal-colored surface. The wood tool had a small diameter and wooden-colored surface.

arrive at a compliance estimate which is then integrated with the haptic estimate.

In this experiment, we modified the visual position information relative to the haptic by three ratios: $2, 1, 0.5$ (used as subscripts in the following expressions), producing conditions x_{v,S_2}, x_{v,S_1}, and $x_{v,S_{0.5}}$, respectively. Physically, the upscaling of the participants' finger movements (x_{v,S_2}) leads to a higher compliance S_v and shortening the indentation ($x_{v,S_{0.5}}$) leads to smaller visual compliance.

Furthermore, we attempted to manipulate the visual cues to force by varying the visual indentation tool. Cylinders instead of spheres were used to represent the indentation tool in the visual VR. We varied their size and surface appearance to simulate metal (subscript "M") and wood (subscript "W"), as depicted in Fig. 7. Using the wood tool was expected to yield a higher compliance estimate than using the metal tool, assuming that it implies that less force is used to reach a given indentation.

Specific hypotheses about the changes in the mean and the variability of the responses were formulated in the context of a set of models, as described below. Here, we describe more general hypotheses. For any level of visual cuing, integration of S_v with S_h, i.e., confusion, suboptimal fusion, or optimal fusion, is indicated by a weighted mean perceptual estimate

$$\mu_{S_m} \stackrel{!}{=} w_h \mu_h + w_v \mu_{v,S_m}, \tag{20}$$

where m denotes the visual ratios of the different bimodal conditions 0.5, 1, or 2. The subscripts h and v denote unimodal haptic and visual information, respectively. Furthermore, the weights sum to unity, $w_h + w_v = 1$ and none of them is zero. Hence, the mean bimodal perceptual estimate should not shift for the congruent vision case, but should shift downward for $\mu_{S_{0.5}}$ and upward for μ_{S_2}. Hence, assuming integration, the mean perceptual estimates can be denoted as

$$\mu_{S_{0.5}} \stackrel{!}{<} \mu_h \stackrel{!}{=} \mu_{S_1} \stackrel{!}{<} \mu_{S_2}. \tag{21}$$

Furthermore, if the compliance information is integrated according to the optimal weighted summation model, the perceptual standard deviations will be smaller in the bimodal conditions. Hence, assuming optimal or suboptimal fusion the perceptual standard deviations can be denoted as

$$\sigma_{S_{0.5,1,2}} \stackrel{!}{<} \sigma_h. \tag{22}$$

4.1 Method

4.1.1 Visual-Haptic Virtual Reality

See the first experiment, Section 3.1, for details.

4.1.2 Participants

Twenty-three (23) students of the Technische Universität München took part in this study for pay and gave informed consent. The average age was 24 years.

4.1.3 Stimuli

The stimuli were virtually rendered compliant cubes with 80-mm sides. Participants were able to squeeze the virtual cube displayed by the HSI using two fingers. The haptic compliance of the standard stimulus was $S_{ref} = 0.7\ \mathrm{mm/N}$. This was considerably lower than in the first experiment to accommodate the expected increase of the mean perceptual estimate due to the visual distortions. The visual compliance of the standard stimulus S_v was defined by the three ratios introduced above. Additionally, we presented a unimodal condition with no visual information S_h and recorded the mean perceptual estimate μ_h and the perceptual standard deviation σ_h.

Furthermore, in the visual stimulus, the fingers were represented not by orange spheres as in Experiment 1, but by two cylinders. The cylinders were differently shaped and colored as described above. In the first configuration, the cylinders had a small diameter (1 cm) and the surface resembled wood (denoted by the subscript "W"). In the second configuration, the cylinders had a large diameter (5 cm) and the surface resembled metal (denoted by the subscript "M"). The conditions are depicted in Fig. 7.

Eight (8) congruent bimodal comparison stimuli were distributed with the standard stimulus at the center of the distribution $S = [0.14, 0.22, 0.36, 0.5, 1, 1.15, 1.3, 1.44]\ \mathrm{mm/N}$.

4.1.4 Procedure

As in the first experiment, the method of constant stimuli was used to assess performance and a within-subject design was used. Eight (8) conditions were tested, constituting the combination of three levels of visual compliance, a haptic-only control condition, and two levels intended to influence the visual force f_v. Each of 23 participants was tested in all conditions. Hence, $4 \times 2 \times 23 = 184$ psychophysical functions were recorded. One trial consisted of the sequential presentation of two stimuli: the standard and the comparison stimulus. Duration of each stimulus presentation was 2 s with an interstimulus interval of 2 s. The next trial began 2 s after the answer was given by the response box. Participants could elect to have a break whenever they wanted by not answering. Each condition was presented in one block consisting of the eight (8) comparison stimuli presented in random order ten (10) times. Hence, 80 trials were necessary to obtain a psychophysical function. The sequence of blocks was randomly chosen for each participant.

Participants were seated in front of the HSI with their dominant hand grasping the device and with the direction of gaze essentially perpendicular to the screen. They were carefully instructed according to the group to which they were randomly assigned. A short training was completed before the session started.

4.1.5 Data Analysis

See the first experiment, Section 3.1, for details.

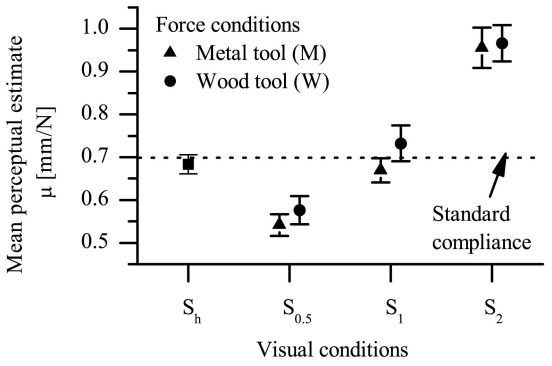

Fig. 8. Mean perceptual estimates as measured by the PSE, (18), with error bars: the bimodal mean perceptual estimate shifted to lower compliance when the visual indentation was half the haptic indentation (condition $S_{0.5}$) and to higher compliance when the visual indentation was twice the haptic indentation (condition S_2). The effect of simulated tool was consistent with the hypothesis that the metal tool would imply larger visual force than the wood tool, but the difference was not significant.

4.2 Results

As in the first experiment, psychophysical functions could be fit to the data in all conditions. (The data of four participants had to be removed, because they produced outliers in a single condition.)

4.2.1 Analysis of the Mean Perceptual Estimate

First, we focus on the mean perceptual estimate μ, as measured by the PSE (18), to make an initial selection among the three basic models explained in Section 2.2. The changes of the mean perceptual estimate induced by varying visual position and indentation tool information are depicted in Fig. 8. A two-way repeated-measures ANOVA was performed on the μ-values obtained from the individual fit of subjects' responses, with factors indentation tool (thick metal, thin wood) and visual cue (S_h, $S_{0.5}$, S_1, S_2). There was a significant effect of visual cue ($F(3,54) = 38.35$, $p < 0.001$), but the effect of the indentation tool cue was not significant ($F(1,18) = 2.68$, n.s.), nor was the interaction ($F(3,54) = 0.23$, n.s.). The first finding confirmed that visual position information influenced the bimodal percept, as would be predicted if integration had occurred, according to (21). The second result indicated that to the extent that force information can be conveyed by simulated material and tool thickness, it did not significantly influence the mean perceptual estimate, contrary to the hypothesis. After merging the data of the indentation tool conditions, the average values were $\mu_h = 0.68$, $\mu_{S_{0.5}} = 0.56$, $\mu_{S_1} = 0.70$, and $\mu_{S_2} = 0.96$ [mm/N].

A contrast was then used to test the specific hypotheses that the bimodal mean perceptual estimate would increase in direct proportion to the visual component of the stimulus, and that the mean for unimodal haptics would equal the mean for the congruent case (i.e., contrast weights were -0.625, -0.125, -0.125, and 0.875 corresponding to conditions $S_{0.5}$, S_h, S_1, and S_2). The contrast showed that the hypothesis accounted for more than 99 percent of the ANOVA sum of squares for the visual-cue factor. Least significant difference (LSD) tests were used to compare the mean perceptual estimates across visual cues, and all means that were predicted to differ according to the hypothesized ordering (21) did differ significantly by this test (i.e., all paired comparisons except for S_h versus S_1). Regarding the different integration models introduced in Section 2.2, the result of this test supports the weighted summation model (6) and the stochastic selection model (11), which predict the same results for the μ-values, (7), (12). It is in contrast to the predictions of the sensory capture model (5), since both modalities were used.

Based on (20), the weights w_v, w_h participants used to integrate the bimodal information can be calculated (both models provide the same equation for the mean perceptual estimate). Two independent calculations can be performed based on the data of each incongruent condition. (It is mathematically not possible to calculate the weights for the congruent condition, based on (20).) To specify the visual mean perceptual estimates for the calculations, we assumed a linear, unbiased transduction of the commanded visual compliance (i.e., the commanded position change) to the visual mean perceptual compliance estimate (see [30] for the support of this assumption)

$$\mu_{v,S_m} = mS_{ref}, \tag{23}$$

where $m = [0.5, 1, 2]$. The calculated values for the visual mean perceptual estimates are $\hat{\mu}_{v,S_{0.5}} = 0.35\ \mathrm{mm/N}$ and $\hat{\mu}_{v,S_2} = 1.4\ \mathrm{mm/N}$. Following (20), (23), the calculations of the average visual weights for the incongruent conditions according to

$$w_{v,S_m} = \frac{\mu_{S_m} - \mu_h}{mS_{ref} - \mu_h}, \tag{24}$$

converge on essentially the same result for both conditions: for condition $S_{0.5}$, the average weights were calculated to be $w_v = 0.36$ and, therefore, $w_h = 0.64$ (note: $w_v + w_h = 1$). For the condition S_2, the average weights were calculated to be $w_v = 0.37$, and $w_h = 0.63$. Hence, the weights were found to act "sticky"; that is, they appeared to be set to a constant value despite differences in the visual distortion across conditions.

4.2.2 Analysis of the Perceptual Standard Deviation

Second, to make a final decision between the two remaining models, weighted summation (6) and stochastic selection (11), we focus on the perceptual standard deviation as measured by the JND, (19). The deviation is inversely correlated with the reliability according to (4). The changes of the perceptual standard deviation induced by visual position and visual force information are depicted in Fig. 9. The same ANOVA reported above for the μ-values was repeated on the σ-values. It showed a significant effect of visual position information ($F(3,54) = 7.49$, $p < 0.001$), whereas the effect of the indentation tool information did not reach significance ($F(1,18) = 0.11$, n.s.). Averaging data over tool conditions, LSD tests were used to compare the σ-values obtained in the incongruent visual conditions with the unimodal condition. These tests indicated that in the bimodal condition S_2 the perceptual standard deviation was greater than the unimodal S_h; the other comparisons of unimodal versus $S_{0.5}$ and unimodal versus S_1 were not significant. Finally, in pursuit of the models described above, we did an LSD test for differences between the σ-values in the incongruent bimodal conditions; the σ-value for $S_{0.5}$ was significantly less than for S_2. Although the analysis of the mean perceptual estimate, Section 4.2.1, revealed that integration occurred in all bimodal conditions,

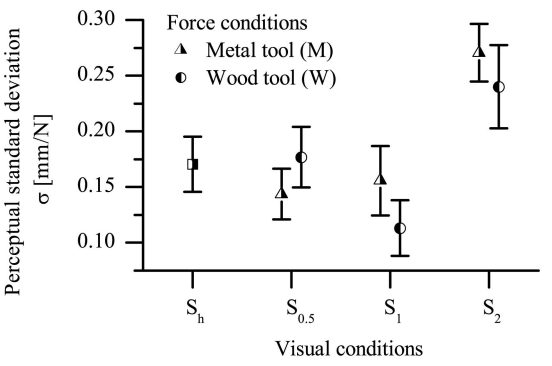

Fig. 9. Perceptual standard deviation σ measured by recording the JND, (19) with error bars: an effect of position information was found between conditions S_h and S_2 and between $S_{0.5}$ and S_2. An effect of visual force information was not found.

optimal integration could not explain why the perceptual standard deviation in condition S_2 was significantly greater than in the unimodal condition. Therefore, hypothesis (22) had to be rejected. As in the above analysis, the ANOVA did not show an influence of the visual force conditions. After merging these data, the average values were $\sigma_h = 0.17$, $\sigma_{S_{0.5}} = 0.16$, $\sigma_{S_1} = 0.13$, and $\sigma_{S_2} = 0.26$ [mm/N].

To test the agreement of the models with the experimental data and the effects reported by the ANOVA, we used the equation of each model that defines the perceptual standard deviation. First, we calculated the visual standard deviation from the experimental data.

Considering stochastic selection, (13), the visual perceptual standard deviation can be identified by calculating

$$\hat{\sigma}_{v,S_m} = \sqrt{\frac{\sigma_{S_m}^2 - (1-w_v)\sigma_h^2 \quad -(1-w_v)w_v(\mu_v - \mu_h)^2}{w_v}}. \quad (25)$$

In this equation, the average values were used and the visual weight $w_v = 0.37$ was derived from the μ-data. The term that represents the mean difference in (13) dropped out, since in condition S_1 the visual and haptic stimuli were physically congruent, and the equivalence of the μ-values indicated that they were perceptually equivalent. The prediction is indicated by a hat $\hat{}$.

Based on this calculation, the stochastic selection model was ruled out, since in all conditions the visual perceptual standard deviation was imaginary. This result was caused by the argument of the square root in (25) (the visual perceptual variance) taking on a negative value. To test that the negative values were statistically reliable, we conducted both parametric (95 percent confidence interval) and nonparametric (binomial) tests on the variances derived from individual subject data. These confirmed that with the stochastic-selection model, the argument of the square root is significantly less than zero for the conditions $S_{0.5}$ and S_2 (the test for S_1 did not exclude a positive value). Further examination of (25) suggests why the imaginary values for the unimodal visual standard deviation are obtained under stochastic selection: the greater the μ-difference (the bimodal conflict), the greater the bimodal perceptual standard deviation must be in order for the argument of the square root to remain positive. In general, the stochastic selection model predicts a bimodal perceptual standard deviation that is too large in comparison to our experimental data, for the situations with intersensory conflict.

Next, consider the weighted summation model according to (8). The visual standard deviation is given by

$$\hat{\sigma}_{v,S_m} = \sqrt{\frac{\sigma_{S_m}^2 - \left((1-w_v)\sigma_h\right)^2}{w_v^2}}. \quad (26)$$

For this calculation, the same average parameters were used as in the test of the stochastic selection model reported above. The predicted values for the mean visual standard deviation were calculated to be $\hat{\sigma}_{v,S_{0.5}} = 0.32$ mm/N, $\hat{\sigma}_{v,S_1} = 0.22$ mm/N, and $\hat{\sigma}_{v,S_2} = 0.64$ mm/N. All values are real numbers. All values are within a plausible range for a visual perceptual standard deviation for compliance, between $[0.22 - 0.64]$ mm/N. The visual perceptual standard deviation for the condition S_2 is greater than the visual perceptual standard deviations for $S_{0.5}$ and S_1. (A comparison between $S_{0.5}$ and S_1 is not allowed, since the ANOVA did not report a significant difference between the bimodal perceptual standard deviations.)

Assuming optimal fusion, it is now possible to derive the weights for all three conditions using (10). For condition $S_{0.5}$, the average weights were calculated to be $w_v^* = 0.22$ and, therefore, $w_h^* = 0.78$ (* indicates optimality). For condition S_1, the average weights were calculated to be $w_v^* = 0.37$, and $w_h^* = 0.63$. And for condition S_2, the average weights were calculated to be $w_v^* = 0.07$, and $w_h^* = 0.93$. The weights so derived for the congruent condition S_1 proved to be essentially the same as the weights identified in Section 4.2.1 for the incongruent conditions. Hence, together with the results of Section 4.2.1, we found that three independent estimates of the weights converged to essentially the same values, further supporting the concept of sticky weights. That is, the weight set for the congruent condition is maintained in the incongruent conditions. The implication is that participants integrated optimally in the congruent condition S_1. However, integration was suboptimal, or even confused participants in the incongruent conditions, because their weights remained sticky and consequently were not optimal.

Based on the foregoing analysis, it is possible to predict the bimodal standard deviations at all possible weights (visual weight between $[0, 1]$) according to

$$\hat{\sigma}_{S_m} = \sqrt{\left(m w_v \hat{\sigma}_{v,S_1}\right)^2 + \left((1-w_v)\sigma_h\right)^2}, \quad (27)$$

where $m = [0.5, 1, 2]$. The result is depicted in Fig. 10. It can be seen that the weighted summation model predicts parabolas whose minima (optimal percepts) are defined by the unimodal standard deviations. The greater the difference between the unimodal standard deviations, the smaller the difference between the optimal bimodal standard deviation (minimum) and the smallest unimodal standard deviation, as can be seen for the prediction of the incongruent condition S_2. Consequently, if humans were to fuse optimally, in highly incongruent conditions their perception would resemble the sensory capture model, i.e., integration would break down. However, since they stick to the weights that lead to an optimal percept in the congruent case, they cannot be optimal in the incongruent conditions, given that integration occurs.

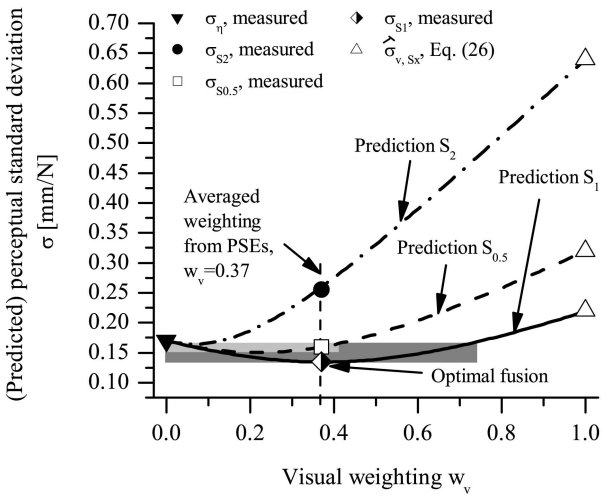

Fig. 10. Predictions of the weighted summation model (6) using measurements and the predictions for the visual perceptual standard deviations, according to (27): the model predicts parabolic curves with a minimum (optimal integration). As can be seen by the gray-shaded areas, fusion would be constrained to a rather small area if the unimodal standard deviations differ substantially. The largest fusion area would be obtained if the two unimodal standard deviations were equal. Outside the fusion area, observers were confused by the integration of the two sources of information.

4.3 Discussion

The influence of visual compliance information on the final compliance percept was analyzed in an experiment with conditions varying the visual compliance by changing visual position and visual force. Visual force was varied by depicting the visual VR with either thin wooden or thick metal tools. A stiffer cube was expected to be reported for the thick metal tool, and although the trend was as predicted, the effect of tool appearance on the bimodal mean perceptual estimate did not reach significance. A more compelling visual rendering might be more effective.

Visual position was varied by doubling and by halving the movement of the fingers. When visual indentation doubled, perceived compliance increased compared to unimodal, whereas if visual position information, i.e., visual indentation, halved, perceived compliance decreased. Congruent visual position information produced the same mean perceptual estimates for compliance as in the unimodal case. Hence, it can be concluded that visual compliance information is integrated with haptic compliance information into a bimodal compliance percept.

Further analysis focused on evaluating models of the integration mechanism. The data analysis supported the weighted summation of random variables (6) as the underlying process for visual-haptic compliance perception. Independent estimates of the weights in three different conditions were found to converge on essentially identical values. The data ruled out models that assumed only a single modality was considered, either across the experiment or stochastically, trial by trial. The weighted summation model closely predicted the average percept μ in all three conditions. Furthermore, it predicted successfully that the perceptual standard deviation in bimodal conditions can be larger than the unimodal, even if integration occurred (confusion). The model further specified that integration in the congruent condition was optimal, but not in the incongruent conditions.

Taken as a whole, these results are inconsistent with optimal fusion (9) as a generally applicable model for compliance. Instead, the weighted summation model, which subsumes the optimal fusion model as a special case, provides a more general paradigm for the integration of visual-haptic compliance information. Depending on the weights, it predicts optimal fusion, suboptimal fusion, and confusion (bimodal standard deviation greater than one of the unimodal deviations). The tendency toward invariant, i.e., sticky weights, likely reflects long-term experience in congruent (i.e., real world) environments. In general, the weights encode how much information a certain modality can contribute to a multimodal percept, i.e., a weight represents the reliability of a certain sensor estimating a certain environmental property. Participants on average adopt weights that lead to an optimal percept in congruent situations. However, they cannot adapt their weights when visual information is incongruent with haptic, if they lack experience of the incongruent situation. This leads to suboptimal fusion or even to confusion.

5 Conclusion

The experiments lead to a model in which each modality combines nonredundant position and force information to estimate compliance, following which the estimates are integrated by a weighted summation process. The first experiment (Section 3) ordered the component processes as combination first, and then integration. The second experiment (Section 4) provided evidence that the general underlying process of bimodal compliance integration is the weighted summation model (6), which contains the optimal fusion model (9) as a special case. Further, the mean data from Experiment 2 suggest that the weights are set to optimal values for congruent visual/haptic inputs and then remain resistant to change, and hence nonoptimal, when the visual input is distorted relative to the haptic.

Although the present work enhances our understanding of visual-haptic compliance perception, questions still remain about the generality of the present model. Experiment 1 indicates that observers cannot combine reliable position and force cues from different modalities into a compliance estimate, suggesting that the weighted-summation model applies to redundant, but not independent, cues. Moreover, previous work using a matching task indicated that observers could ignore discrepant visual compliance information [18], indicating that integration might be task dependent. Further research is needed to determine how weighted summation applies to the perception of other environmental properties, such as position, viscosity, volume, etc. Another question that needs further investigation is how the weights given by participants to the two sources of information are assigned and whether they can be varied with experimental manipulations as in [7].

Acknowledgments

This research was partly funded by the German National Science Foundation (DFG). The authors would like to thank Professor Robert H. Swendsen, Department of Physics, Carnegie Mellon University, for his invaluable comments on the models. The authors kindly appreciated the suggestions of the reviewers.

REFERENCES

[1] J. Driver and C. Spence, "Multisensory Perception: Beyond Modularity and Convergence," *Current Biology,* vol. 10, no. 13, pp. 731-735, 2000.

[2] D.W. Massaro, "Speechreading: Illusion or Window into Pattern Recognition," *Trends in Cognitive Science,* vol. 3, no. 8, pp. 310-317, 1999.

[3] B.E. Stein and M.A. Meredith, *The Merging of the Senses.* MIT Press, 1993.

[4] R.B. Welch and D.H. Warren, "Intersensory Interactions," *Handbook of Perception and Human Performance: Sensory Processes and Perception,* vol. 1, John Wiley & Sons, 1986.

[5] G.A. Calvert, M.J. Brammer, and S.D. Iversen, "Crossmodal Identification," *Trends in Cognitive Sciences,* vol. 2, no. 7, pp. 247-253, 1998.

[6] S. Guest and C. Spence, "What Role Does Multisensory Integration Play in the Visuotactile Perception of Texture?," *Int'l J. Psychophysiology,* vol. 50, pp. 63-80, 2003.

[7] M.O. Ernst and M.S. Banks, "Humans Integrate Visual and Haptic Information in a Statistically Optimal Fashion," *Nature,* vol. 415, pp. 429-433, 2002.

[8] J.M. Hillis, M.O. Ernst, M.S. Banks, and M.S. Landy, "Combining Sensory Information: Mandatory Fusion within, But Not between, Senses," *Science,* vol. 298, pp. 1627-1630, 2002.

[9] J.M. Hillis, S.J. Watt, M.S. Landy, and M.S. Banks, "Slant from Texture and Disparity Cues: Optimal Cue Combination," *J. Vision,* vol. 4, no. 13, pp. 1-24, 2004.

[10] D.C. Knill and J.A. Saunders, "Do Humans Optimally Integrate Stereo and Texture Information for Judgements of Surface Slant?," *Vision Research,* vol. 43, pp. 2539-2558, 2003.

[11] L.A. Jones and I.W. Hunter, "A Perceptual Analysis of Stiffness," *Experimental Brain Research,* vol. 79, pp. 150-156, 1990.

[12] H.Z. Tan, Pang, X.D. Pang, and N.I. Durlach, "Manual Resolution of Length, Force, and Compliance," *Proc. Winter Ann. Meeting of the Am. Soc. of Mechanical Engineers Advances in Robotics,* 1992.

[13] H.Z. Tan, N.I. Durlach, G.L. Beauregard, and M.A. Srinivasan, "Manual Discrimination of Compliance Using Active Pinch Grasp: The Roles of Force and Work Cues," *Perception and Psychophysics,* vol. 57, pp. 495-510, 1995.

[14] N. Dhruv and F. Tendick, "Frequency Dependence of Compliance Contrast Detection," *Proc. ASME Dynamic Systems and Control Division Conf.,* 2000.

[15] R.H. LaMotte, "Softness Discrimination with a Tool," *J. Neurophysiology,* vol. 83, pp. 1777-1786, 2000.

[16] W. Bergmann-Tiest and A. Kappers, "Kinaesthetic and Cutaneous Contributions to the Perception of Compressibility," *Proc. Int'l Conf. Haptics: Perception, Devices and Scenarios,* pp. 255-264, 2008.

[17] M.A. Srinivasan, G.L. Beauregard, and D.L. Brock, "The Impact of Visual Information on the Haptic Perception of Stiffness in Virtual Environments," *Proc. ASME Dynamics Systems and Control Division Conf.,* 1996.

[18] F. Freyberger, M. Kuschel, B. Färber, M. Buss, and B. Färber, "Visual-Haptic Perception of Compliance: Direct Matching of Visual and Haptic Information," *Proc. IEEE Int'l Workshop Haptic Audio Visual Environments (HAVE) and Applications,* 2007.

[19] M. Kuschel, F. Freyberger, M. Buss, B. Färber, and R.L. Klatzky, "Visual Haptic Perception of Compliance: Fusion of Visual and Haptic Information," *Proc. Symp. Haptic Interfaces for Virtual Environments and Teleoperator Systems,* 2008.

[20] V. Varadharajan, R.L. Klatzky, B. Unger, R. Swendson, and R. Hollis, "Haptic Rendering and Psychophysical Evaluation of a Virtual Three-Dimensional Helical Spring," *Proc. 16th Symp. Haptic Interfaces for Virtual Environments and Teleoperator Systems,* 2008.

[21] R. Teghtsoonian, "On the Exponents in Stevens' Law and the Constant in Ekman's Law," *Psychological Rev.,* vol. 78, no. 1, pp. 71-80, 1971.

[22] B.R. Brewer, M. Fagan, R.L. Klatzky, and Y. Matsuoka, "Perceptual Limits for a Robotic Rehabilitation Environment Using Visual Feedback Distortion," *IEEE Trans. Neural Systems and Rehabilitation Eng.,* vol. 13, no. 1, pp. 1-11, Mar. 2005.

[23] H.Z. Tan, M.A. Srinivasan, C.M. Reed, and N.I. Durlach, "Discrimination and Identification of Finger Joint-Angle Position Using Active Motion," *ACM Trans. Applied Perception,* vol. 4, no. 2, pp. 1-14, 2007.

[24] R. van Beers, A. Sittig, and J.J. Danier van der Gon, "The Precision of Proprioceptive Position Sense," *Experimental Brain Research,* vol. 11, pp. 367-377, 1998.

[25] R. van Beers, D. Wolpert, and P. Haggard, "When Feeling Is More Important Than Seeing in Sensorimotor Adaptation," *Current Biology,* vol. 12, pp. 834-837, 2002.

[26] E. Brenner and J.B.J. Smeets, "Fast Corrections of Movements with a Computer Mouse," *Spatial Vision,* vol. 16, pp. 364-376, 2003.

[27] F.A. Wichmann and N.J. Hill, "The Psychometric Function: I. Fitting, Sampling, and Goodness of Fit," *Perception and Psychophysics,* vol. 63, pp. 1293-1313, 2001.

[28] N.A. Macmillan and C.D. Creelman, *Detection Theory—A User's Guide,* second ed. Lawrence Erlbaum Assoc., Inc., 2005.

[29] M. Teghtsoonian and R. Teghtsoonian, "Seen and Felt Length," *Psychonomic Science,* vol. 3, pp. 465-466, 1965.

Martin Kuschel is a project manager for lightweight robotics and human-robot collaboration at KUKA Roboter GmbH, Augsburg, Germany. He studied electrical engineering and wrote the dissertation about visual-haptic perception and compression of haptic information at the LSR, Technische Universität München, Germany. He was a recipient of a fellowship at the Department of Psychology, Carnegie Mellon University, Pittsburgh.

Massimiliano Di Luca received the master degree in psychology from the University of Trieste and the PhD degree in cognitive sciences from Brown University. He is a research scientist at the Max Planck Institute for Biological Cybernetics in Tübingen, Germany. His main research interests include 3D-shape, haptic, and multimodal perception.

Martin Buss is a full professor, head of the Institute of Automatic Control Engineering (LSR), Technische Universität München. Since 2006, he is the coordinator of the DFG Excellence Research Cluster *Cognition for Technical Systems—CoTeSys*. His main research interests include automatic control, mechatronics, robotics, intelligent control, multimodal human-system interfaces, haptic systems, optimization, nonlinear, and hybrid systems.

Roberta L. Klatzky received the BS degree in mathematics from the University of Michigan and the PhD degree in experimental psychology from Stanford University. She was a member of the faculty at the University of California, Santa Barbara. She is a professor of psychology at Carnegie Mellon University, where she is also on the faculty of the Center for the Neural Basis of Cognition and the Human-Computer Interaction Institute. She is the author of more than 200 articles and chapters and has authored or edited five books.