Visual-Haptic Perception of Compliance: Fusion of Visual and Haptic Information

Martin Kuschel, Martin Buss*
Institute of Automatic Control Engineering,
Technische Universität München,
Arcisstraße 21,
80290 Munich, Germany

Franziska Freyberger, Berthold Färber†
Human Factors Institute,
Universität der Bundeswehr München,
Werner-Heisenberg-Weg 39,
85577 Munich, Germany

Roberta L. Klatzky‡
Department of Psychology,
Carnegie Mellon University,
Baker Hall 342,
Pittsburgh, PA 15213, USA

ABSTRACT

Visual-haptic compliance perception demands processing of haptic position and force information as well as visual position information. In this study it was analyzed how different compliances are perceived and how single modality estimates are combined to a visual-haptic percept of compliance. Participants estimated the compliance of a virtual cube displayed by a human system interface. Thereby, psychometric functions were recorded and statistically evaluated. Results indicated that human's ability to discriminate differences relative to a standard compliance decreased as the standard compliance increased. Furthermore, in the conflicting case participants' bimodal percept was close to the modality that captured the less compliant information.

Index Terms: Perception, Compliance, Human System Interface, Sensory Fusion, Maximum-Likelihood Estimation

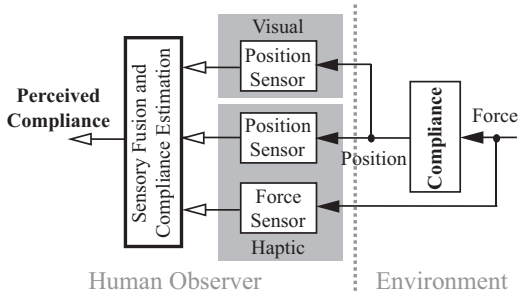

Figure 1: Visual-haptic perception of compliance information: Redundant position information and haptically perceived force information are combined to generate the final estimate.

1 INTRODUCTION

Mechanical environments are mainly perceived by processing position-based information (position, velocity, acceleration) and force information. However, humans do not have distinct modalities for each of these kinds of information. Considering visual-haptic perception, position-based information can be detected by both modalities while force information can only be detected by the haptic (kinesthetic) modality. Furthermore, the obtained information has to be computationally processed to obtain an estimate of the explored mechanical environment. An analysis of the underlying fusion process will contribute not only to psychophysical knowledge but also to the design and control of human system interfaces used to access artificial environments. A sound analysis requires a visual-haptic HSI with high accuracy in displaying mechanical environments and extensive experiments using psychophysical procedures.

Information from different senses is integrated to form a coherent perception of our environment. Several models of integrating multisensory information have been developed and can be subdivided into the following three model types (e.g. [10, 19]): A) Sensory penetration (one modality influences the other during unimodal stimulus processing), B) feedforward convergence and C) feedforward convergence with feedback-loop. Model B of integration is the most popular one (e.g. [10, 19, 24]): Information from different senses is separately processed and converges to a coherent percept at higher processing levels (e.g. [25]). However, modalities do not necessarily contribute equally to the bimodal percept: Visual dominance has often been reported, such that the visual modality captures e.g. the haptic modality (see [21]; for a review see [7, 28]). Different approaches argue that the relative contribution of each sense depends on modality appropriateness (e.g. [28]), effectiveness (e.g. [6]) or on the direction of attention (e.g. [15]). Another approach states that information from different senses are fused in such a way that the percept is the best possible estimate (e.g. [12]). Ernst & Banks point out that under a maximum-likelihood model this estimation depends on the reliability of each modality: The modality with the highest reliability contributes most to the bimodal percept. On the other hand, if the reliability of one modality is reduced, its relative contribution to the bimodal percept decreases (e.g. [11]). Considerable research on integration based on the maximum-likelihood theory (see for further detail 2) has already been done [1, 4, 11, 16]. Most research on bimodal integration has concentrated on one sensory signal, e.g. position [11]. Additional studies have addressed the question of integrating information within a modality (e.g. [9, 17, 18]) and have also shown differences between integration within and between senses (e.g. [16]). However, investigating the integration of more complex variables that induce observers to integrate bimodal signals of different dimensions has not been undertaken yet. Haptic perception of compliance has already been shown to require the combination of force and position cues. This results in a loss of sensitivity (see e.g. [8, 20, 26, 27, 31]).

The contribution of this article is an analysis of visual-haptic compliance integration. According to Hooke's law, compliance is the combination of position and force information. Since position information is redundant (measured by visual and haptic modalities) the perception of compliance must include the fusion of redundant information. As a basis of our investigation it was assumed that humans independently identify the presented compliance visually and haptically. In a second step they integrate the visual and the

*e-mail: [martin.kuschel, mb]@tum.de
†e-mail: [franziska.freyberger, berthold.faerber]@unibw.de
‡e-mail: klatzky@cmu.edu

Symposium on Haptic Interfaces for Virtual Environments and Teleoperator Systems 2008
13-14 March, Reno, Nevada, USA
978-1-4244-2005-6/08/$25.00 ©2008 IEEE

haptic compliance identification result to a single percept. Since humans do not have visual force sensors we assumed that people inferred force cues from the cube itself or from a force prior. We used 2AFC-tasks to record the psychometric functions of visual, haptic, and visual-haptic compliance perception. A test of whether integration of visually and haptically estimated compliance obeyed the maximum-likelihood model [11] was conducted. The results showed that participants had difficulties in identifying compliance from pure visual cues. The reliability of compliance perception decreased with increasing compliance. In our study we could not confirm the maximum-likelihood model.

The article is organized as follows: In Section II the mathematical background of multimodal perception is presented. The Human System Interface used for the experiments is introduced in Section III. In the following sections method (Section IV) and results (Section V) are described. Section IV provides a discussion of the research conducted, and conclusions are drawn in Section VII.

2 Perception of Bimodal Information

Systems that extract information about an environment are often equipped with multiple sensors. The mapping from the real world to an internal modal based on sensed information is called the perceptual process. During this process multi-sensory information is fused and extended by a priori information about the environment (See [12] for psychophysical evidence). Information sensed can be complemental (e.g. color and size of a certain object) or redundant (e.g. position of a certain object sensed by visual and proprioceptive sensors). Perception of redundant information should reduce the uncertainty of the final estimate (See [11] for psychophysical evidence). Thereby, single modality estimates

$$\hat{s}_i = f_i(s), \tag{1}$$

are combined to an integrated percept

$$\hat{s} = g(\hat{s}_1, \hat{s}_2, \ldots). \tag{2}$$

The variable s represents the stimulus generated by the environmental property. The index i indicates the modality and the function g defines the perceptual fusion mechanism.

Maximal exploitation of the sensed redundant information can be computed using concepts of mathematical optimization. Thereby, a cost function is minimized with respect to the estimation error of the final percept $\hat{s}$. It is assumed that the fusion mechanism linearly combines the sensed information

$$g(\hat{s}_1, \hat{s}_2, \ldots) = k_1\hat{s}_1 + k_2\hat{s}_2 + \ldots. \tag{3}$$

In the case of visual-haptic, i.e. bimodal, information the optimal integrated estimate $\hat{s}_{opt}$ is distributed normally having a mean

$$\bar{s}_{\text{opt}} = \underbrace{\frac{\sigma_h^2}{\sigma_v^2+\sigma_h^2}}_{k_v}\bar{s}_v + \underbrace{\frac{\sigma_v^2}{\sigma_v^2+\sigma_h^2}}_{k_h}\bar{s}_h. \tag{4}$$

and a standard deviation

$$\sigma_{\text{opt}} = \sqrt{\frac{\sigma_v^2\sigma_h^2}{\sigma_v^2+\sigma_h^2}}. \tag{5}$$

As depicted in the upper diagram of Figure 2 the combined percept yields a more reliable estimate represented by the highly peaked distribution. Furthermore, the mean value of the combined percept is a weighted average of the means of the single modality estimate. For complete derivation of the optimal solution refer to [5].

Human perception is a probabilistic process. Assuming a Gaussian

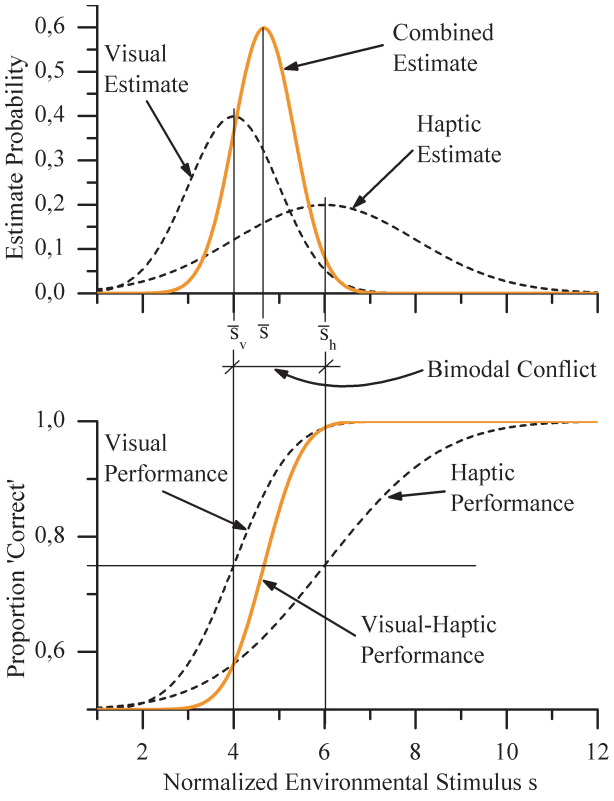

Figure 2: Perception of bimodal information: Minimizing expectation of the quadratic estimation error yields a combined percept with optimal variance (upper diagram). Characteristics of estimates can be obtained by differentiating the psychometric function that represent observer performance in stimulus discrimination (lower diagram).

characteristic, distributions of single and bimodal estimates can be recorded using e.g. a two-alternative-forced-choice task (2AFC) (e.g. [14]). The result, the psychometric function, expresses the observer's performance with respect to a categorical judgment (see lower diagram in Figure 2). In this type of task, the observer has to compare a stimulus s to a standard stimulus s_s. If she/he has decide for one of two alternatives (e.g. *"Was the second stimulus more compliant than the first stimulus?"*) the task is called two-alternative-forced-choice task ('2AFC'). The upper bound of this psychometric function (normally 1.0) represents the observer's performance with a comparison stimulus arbitrarily greater than the standard value, i.e. the observer answers that the comparison is greater on 100% of the trials. The lower bound of this function (usually 0.0) represents the probability of answering that the comparison stimulus is less than the standard stimulus for an arbitarily lesser value. The psychometric function normally has the shape of a sigmoid and can be modeled by a cumulative Gaussian distribution. Hence, the characteristics of the perceptual estimate, i.e. the mean ($\bar{s}_{\text{stimulus}}$) and the standard deviation (σ_{stimulus}), can be empirically estimated by taking the stimulus value and the slope at proportion 0.5. The sloppe at proportion 0.5 corresponds to $1/(\sigma_{\text{stimulus}}\sqrt{2\pi})$. The inverse value of the quadrat of this standard deviation is also called 'reliability' of the perceptual estimate.

The mean value $\bar{s}_{\text{stimulus}}$ of the distribution fit to the data is called "point of subjective equality" (PSE) as it defines the value where 50% of the participants decide for the first response alternative and

50% for the second:

$$\mathrm{PSE} := \bar{s} \tag{6}$$

The constant error (CE) is defined as the difference between the PSE and standard stimulus:

$$\mathrm{CE} := \mathrm{PSE} - s_{\mathrm{standard}} \tag{7}$$

A CE is only possible to compute when perceiving information from a single modality or when bimodal information is congruent, i.e. has no conflict.
The bimodal conflict is defined by the difference between the means of the single modality estimates

$$c := |\bar{s}_{\mathrm{v}} - \bar{s}_{\mathrm{h}}| . \tag{8}$$

If the conflict induced by the sensory estimates of the two modalities is too large the perceptual system might not be able to fuse incoming information. In this case the conflict is perceived because the final estimate will not result in a single coherent percept [11].
The relative difference between optimal PSE and empirically estimated PSE is defined by:

$$e_{\mathrm{PSE}} := \left| \frac{\mathrm{PSE}_{\mathrm{opt}} - \mathrm{PSE}_{\mathrm{VH}}}{\mathrm{PSE}_{\mathrm{opt}}} \right| . \tag{9}$$

Respectively, the difference of the slopes is defined by

$$e_{\sigma} := \left| \frac{\sigma_{\mathrm{opt}} - \sigma_{\mathrm{VH}}}{\sigma_{\mathrm{opt}}} \right| . \tag{10}$$

In both equations the subscript VH refers to an perceptual estimate from a bimodal (visual-haptic) stimulus condition.

3 Human System Interface

The rendering of mechanical impedances, like compliance, by a haptic display is a difficult task. Therefore, the details of the hardware and the control are described in the this section.

3.1 Hardware and Software

The human system interface (HSI) provides visual and proprioceptive (haptic) feedback. Furthermore, it measures finger positions and forces. See Figure 3 for a photo of the device and Figure 4 for a sketch of the haptic subsystem.

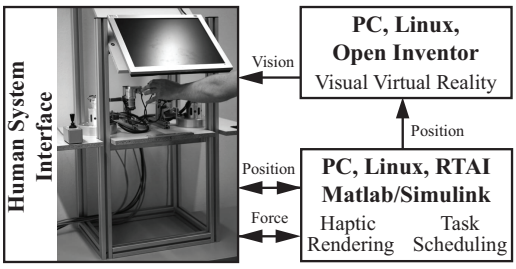

Figure 3: Human system interface and real-time processing unit: Visual and proprioceptive (haptic) information is fed back and positions are sensed.

Proprioceptive information is exchanged via a haptic interface comprised of two SCARA robots providing a single degree of freedom each. The system interacts with index finger and thumb to allow gripping movements. High fidelity components like Maxon motors and Harmonic Drive Gears enable best possible control. Workspace is about 80 mm and maximum force is about 35 N. Position information is measured by angle encoders and force is sensed by strain gauges attached on both robot links. Visual information is provided over a TFT screen. Thereby, the compliant environment is represented by a grey cube squeezed by two orange spheres (on opposed cube sides) representing finger positions (see Figure 5). The TFT screen is slanted by 40° and mounted in the line of sight to the hand enabling participants to look at their hand [1]. The system is connected to a PC running RTAI-

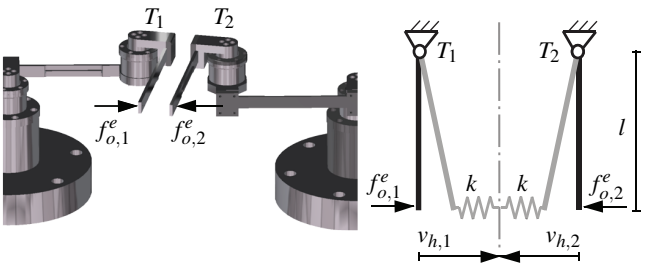

Figure 4: Kinematical structure of the haptic display: Two SCARA robots present compliance information for gripping movements.

RealTime Application Interface for Linux. SCARA sensor signals are recorded by a "Sensoray 626" DAQ-Card providing 16 bit sensing resolution. Signal processing algorithms are implemented as Matlab/Simulink models with real-time code generated automatically. The system operates at 1 kHz sampling frequency. Measured positions are transferred to a second PC running the visual virtual reality programmed in Open Inventor.

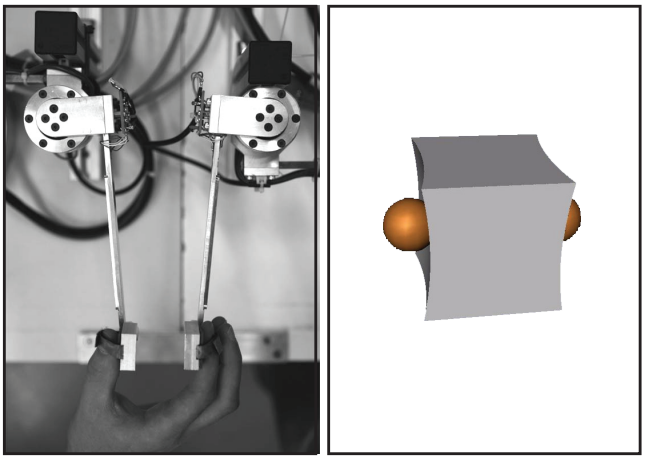

Figure 5: Haptic and visual feedback: The haptic feedback renders a compliant cube to be explored by thumb and index finger. In the visual feedback fingers are replaced by orange spheres.

3.2 Dynamics and Control

The identical robots of the HSI are controlled independently using the same admittance control scheme (see Figure 4 for kinematical configuration). In the following, the concept is explained using a single robot system without loss of generality. Furthermore, the explanation is restricted to translational movements only (kinematical transformation are ignored) since robot links are only moved little when performing the gripping tasks.

[1] The tool transformation has no influence on the dynamics of the gripping movement, if participants are given a learning phase (e.g. see [3])

For dynamics consider a mechanical robot with a single translational degree-of-freedom. The dynamical equation is given by

$$M_h \dot{v}_h + n_h = g_h - f_o^e \tag{11}$$

Where $M_h \in \mathbb{R}$ and $n_h \in \mathbb{R}$ denote mass and nonlinearities of the robot. Robot force $g_h \in \mathbb{R}$ depends on motor torque T and on link length l, respectively (Figure 4). The velocity of the tool tip is denoted by v_h. Input-output linearization [23] is achieved by commanding

$$g_h = f_h^m + n_h. \tag{12}$$

The resulting linear dynamics are

$$M_h \dot{v}_h = f_h^m - f_o^e. \tag{13}$$

Where f_h^m is the new motor force of the linearized HSI.
A velocity controller, $C_v : U \rightarrow M$, realizes the command signal f_h^m according to

$$f_h^m = G_h(C_v(v_v - v_h)). \tag{14}$$

Where G_h represents the dynamics of the actuator, which can be reduced to the dynamics of the current control.
The HSI is serially connected to the human operator, whose fingers are described by the dynamics $Z_o : U \rightarrow M$. The velocity of the HSI and the velocity of the operator's fingers are opposite

$$v_h = -v_o. \tag{15}$$

The dynamics of the robot interacting actively with the human operator are described by

$$f_o = Z_o(v_o) + f_o^m. \tag{16}$$

Where f_o^m is the force actively intended by the human operator impeded by the force f_o that mediates the virtual reality (VR).
The dynamics of the VR is described by the admittance $Y_v : U \rightarrow M$, which represents pure stiffness yielding

$$v_v = Y_v(k, f_m^e) = k^{-1} \frac{df_m^e}{dt}. \tag{17}$$

Where k [N/mm] is the stiffness coefficient (compliance k^{-1}) whose perception is addressed in this publication. The control concept employing inner velocity control driven by a virtual reality with force reference is called admittance control. It is best suitable for rendering non-rigid environments like compliant environments (see [2] for detailed information). Minimum compliance (= maximum stiffness) that can be rendered is $k^{-1} = 0.2$ mm/N.
A block diagram of the human operator interacting with the haptic human system interface is depicted in Figure 6. Hollow arrows depict physical interactions, filled arrows are used for signal interactions. All subsystems are considered to be linear and time-invariant. The fidelity of the VR depends on dynamics and control of the HSI. The robot is light weighted, dynamics of the motor current control are negligible, and velocities are small (i.e. friction effects negligible). Consequently, the transparency of the system can be assumed as nearly ideal and the displayed dynamics Y_d can be considered equal to the dynamics of the VR

$$Y_d = Y_v. \tag{18}$$

4 Method

4.1 Participants

Fifty-eight students of the Technische Universität München and the Ludwigs-Maximilian-Universität, München took part in this study and were paid for participation. Participants were randomly assigned to one of the experimental groups according to modality in which stimuli were explored, see Section 4.2. Thirteen participants had to be excluded from further analysis, because no psychometric function could be fitted to the data; most of theses participants had been assigned to group V (n=8). Eleven persons (6 men, 5 women) with an average age of 24 years explored the stimuli without vision (group H), and 11 participants (5 men, 6 women) with an average age of 26 years explored without haptic feedback (group V). The other two groups tested the cube bimodally: 12 persons (6 men, 6 women) with an average age of 24 years tested congruent stimuli (group HV), whereas 11 participants (4 men, 7 woman) with an average age of 25 years explored incongruent stimuli (group HVS). All participants were right-handed and had normal or corrected to normal vision.

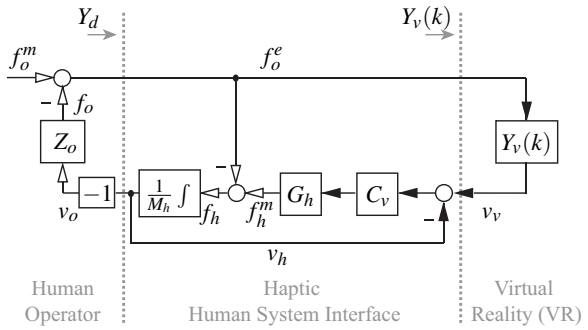

Figure 6: Admittance control: The haptic virtual reality is induced using different stiffnesses. High performance is achieved through light weighted robot.

4.2 Stimuli

All stimuli were virtual rendered compliant cubes of 80 mm edge length having a certain compliance s. They were either unimodally, i.e. visually or haptically, or bimodally, i.e. visual-haptically, displayed by the HSI (see Section 3). The bimodal stimuli were presented congruently, i.e. haptic and visual information were equal or as conflicting stimuli, i.e. haptic and visual information differed according to equation (8). No visual feedback was given during the unimodal haptic presentation nor was haptic feedback given in unimodal visual presentation (except the possibility to move the grasp device in order to move the visual fingertips).
Five standard stimuli were selected. The values were 0.2, 0.4, 0.9, 1.4, and 2.1 [mm/N]. Seven comparison stimuli were selected to be linearly distributed above the standard stimuli. The comparison stimulus with the lowest compliance was the same as the standard stimulus. Since the detection threshold is known to be around 30% (see [13] for detailed information), the most compliant comparison stimulus was twice the standard value to assure that nearly all participants would detect the difference. The reason why the standard stimuli were not centered in the middle of the comparison stimuli is due to the fact that it is difficult to haptically display relatively compliant environments accuratly by a human system interface, especially when the standard stimuli are less compliant themselves, which was the case here.
Additionally, bimodal conflicting stimuli were generated using a reference modality, which defined the standard stimulus. With the haptic modality as reference the (haptic, visual) combinations were (0.2, 0.4), (0.9, 1.6), (1.4, 2.1) in [mm/N]. With the visual modality as reference the (haptic, visual) combinations were (0.4, 0.2), (1.3, 0.9), (2.1, 1.4) in [mm/N].

4.3 Design and Experimental Paradigm

Design. The psychometric function for discriminating compliance was assessed within the following modality conditions: Unimodal haptic (H), unimodal visual (V), bimodal congruent and bimodal

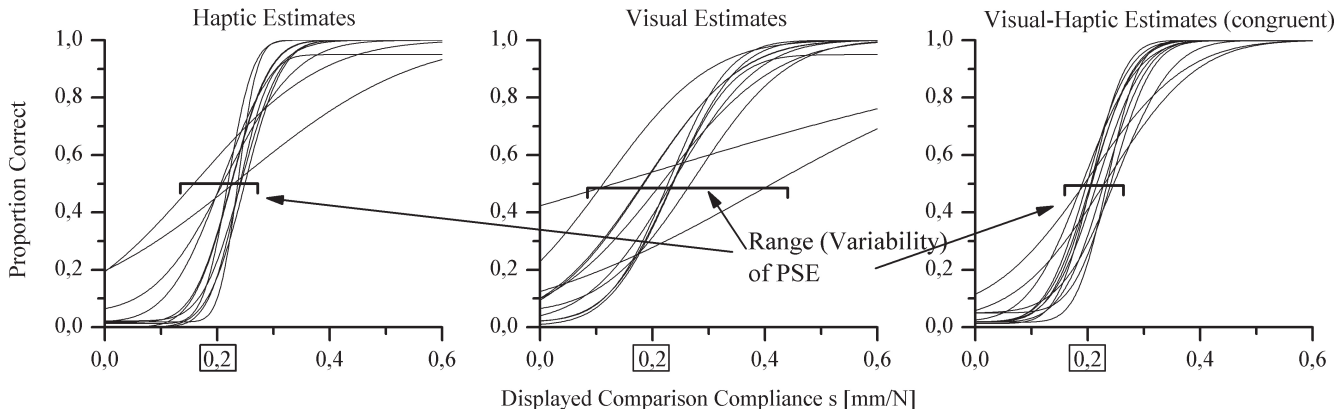

Figure 7: Unimodal and bimodal compliance estimates (referenced to the standard compliance 0.2 N/mm): Haptic estimates and visual-haptic estimates show nearly identical slopes and variabilities. Visual estimates show higher variabilities and a lower slopes. (Due to the experimental design values smaller than the standard compliance were extrapolated by using the symmetric characteristic of the sigmoid.)

incongruent stimulus presentation with one modality being the reference modality (VH). Because testing time is relatively long using this method, modality was chosen to be a between-participants variable. Each of the above described five standard stimuli was recorded using the 2AFC-tasks (e.g. [14]; also used by [11]); standard compliance was a within-participants variable.

Experimental Paradigm. One trial consists of the sequential presentation of two stimuli: the standard and the comparison stimulus; the sequencing of standard stimulus and comparison stimulus differed randomly. Duration of each stimulus presentation was 2 s with an inter-stimulus interval of 2 s and an inter-trial interval of 4 s. Discrimination performance of each standard stimulus was assessed during one block, within which, each combination of this standard and the seven comparison stimuli was randomly presented 8 times.

The task was a 2AFC-task. That is, subjects had to compare both sequentially presented stimuli and to decide whether the second stimulus was more compliant than the first one. Because only stimuli that were more compliant than the standard stimulus had been chosen for this experiment (see Section 4.2) and position of the standard stimulus was randomly varied the resulting psychometric function represents the proportion of correctly detected differences (further referred to as 'proportion correct'). In our design, this is equivalent to the proportion of stimuli rated greater than the standard.

4.4 Experimental Procedure

Participants were seated in front of the HSI with their dominant hand grasping the device and with 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 had to be completed, before the test session started.

At the end of the experiment, participants were asked to fill in questionnaires assessing their demographical data. Additionally, because the ability to be drawn into a book, film or VE, better known as *immersive tendency* (see [30]), has been assumed to contribute to performance within virtual environments. An additional 12-item questionnaire was included to control for that variable ([30], translated by [22]): Immersive tendency was assessed by answering 12 items building the two factors, *tendency to get emotionally involved* and *degree of involvement* (see [22]).

4.5 Data Analysis

Questionnaire data were first analyzed (see Section 5.1).

For each participant and each standard stimulus a psychometric function was fitted using "psignifit" software, version 2.5.6, described in [29] and Matlab/Simulink (see Section 4.3). The following parameters were computed by fitting the psychometric function to the experimental data: Means (PSE_H, PSE_V, PSE_{VH}) and standard deviations (σ_H, σ_V, σ_{VH}) of the perceptual estimate were computed for each standard stimulus by taking stimulus level and slope at proportion 0.5 (see also Section 2). Additionally, the CE was computed.

Whether standard compliance affected discrimination performance (see Results: Subsection 5.2) and whether visual-haptic integration occurs (see Results: Subsection 5.3) were tested separately. The first question was tested with the CE and standard deviation of the PSE (which will further be referred to as η) and additionally with the slope of the psychometric function, σ.

The second question was tested by first estimating the optimal mean (PSE_{opt}) as well its standard deviation (σ_{opt}) according to equation (4) and equation (5). Results were descriptively analyzed and compared to the experimentally derived bimodal parameters (PSE_{VH} (congruent ans incongruent), σ_{VH} (congruent and incongruent)).

5 Results

5.1 Questionnaire Data

Immersive tendency. Both factors, emotional involvement and degree of involvement, were computed for each participant. Rating of emotional involvement ranged between 26.4 and 29.9, and degree of involvement between 17.5 and 21.0 within all 4 groups. These values did not statistically significantly differ from those reported by Scheuchenpflug [22], indicating that the participants are a good sample of the general population. No correlation between emotional involvement and individual thresholds and their standard deviation could be observed. This indicates that no immersive tendency influenced participants.

Additional rating of group VH and VHS. Participants of both bimodal groups rated which modality they attended to: Most attended to the haptic (group VH: n=9, group VHS: n=7), the visual (group VH: n=2, group VHS: n=1) or both modalities (group VH: n=1, group VHS: n=3).

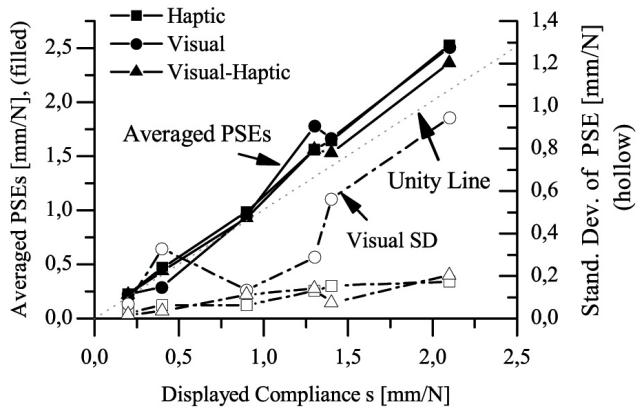

Figure 8: Unimodal and bimodal compliance estimates: PSEs (averaged across subjects, solid line) show low constant error independent of displayed compliance and modality. Standard deviation of PSEs (dashed line) across subjects increases with higher compliance. The visual standard deviation, in particular, is highly variable indicating the difficulty when estimating compliance visually.

5.2 Perception of Different Compliances

The first result reveals that participants had difficulty in discriminating compliance which was only presented visually: As can be seen from Figure 7 the variability of the CE, is small in group H and group VH, but larger in group V. This can also be seen from Figure 8 which shows the PSE and the standard deviation of the PSE, η, plotted against the standard compliance: Although the CE (difference between PSE and unity line) is small for nearly all standard stimuli, η from the visual estimate is higher for nearly all standard stimuli. Consequently, the mean CE across all standard compliances in group H amounted to 0.2 N/mm (standard deviation $\eta = 0.1$ N/mm), 0.17 N/mm ($\eta = 0.08$ N/mm) in group VH, and in group V 0.24 N/mm ($\eta = 0.4$ N/mm).
Additionally, as has been reported in Section 4.1, several participants had difficulty in performing the task and had to be excluded form further analysis.
The second result shows that participants' ability to generally estimate compliances decreases with increasing compliance. It was tested whether the CE changed depending on the standard compliance (separately for all groups). With increasing standard compliance, there was an increase in CE in group H (Greenhouse Geisser corrected: $F(2.7, 26.8) = 26.64, p < 0.05$; partial $\eta^2 = 0.727$) as well as in group VH (Greenhouse Geisser corrected: $F(3.0, 32.6) = 20.18, p < 0.05$; partial $\eta^2 = 0.647$), but not in group V (Greenhouse Geisser corrected: $F(1.8, 18.4) = 0.98, p = 0.389$). Descriptively, this can be seen in the second part of Figure 8 relating the standard deviation of the PSEs, η, to the displayed compliance. It can be seen that all curves show a positive trend. Especially, η of the visual estimates showed a strong increase with higher standard compliances.
The consistently positive value of the CE is due to the fact that comparison compliance was always greater than the standard compliance.

The third result confirms the second result showing that participants' estimate of compliance became less reliable with decreasing compliance. As depicted in Figure 9, the slope of the psychometric function decreases with increasing standard compliance. That means the standard deviation σ of the perceptual estimate increases with larger compliances. Whether this influence is statistically significant was tested separately for all modalities. A statistically significant influence of standard compliance was found in group H (Greenhouse Geisser corrected: $F(1.4, 14.3) = 8.11, p < 0.05$; partial $\eta^2 = 0.448$), in the bimodal congruent presentation group VH (Greenhouse Geisser corrected: $F(2.4, 26.4) = 41.80, p < 0.05$; partial $\eta^2 = 0.792$), and in the bimodal conflict group VHS (Greenhouse Geisser corrected: $F(1.2, 11.6) = 10.43, p < 0.05$; partial $\eta^2 = 0.510$). In each case there was an increase in σ with increasing compliance. No influence of compliance was found to affect the σ in group V (Greenhouse Geisser corrected: $F(1.4, 14.3) = 8.11, p = 0.309$).

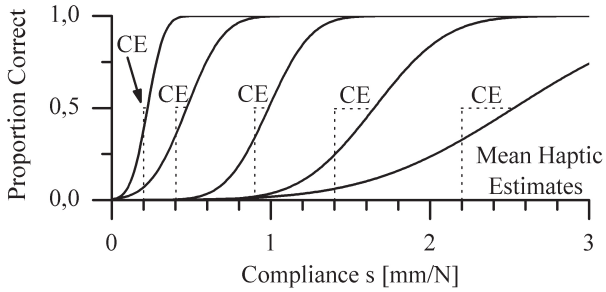

Figure 9: Psychometric functions of different compliance estimates (standard compliances indicated by dotted-lines): Haptic estimation of compliance became less reliable with increasing standard compliance. The same results were discovered for the bimodal groups. A comparable increase in standard deviation was not observed in the visual groups, where performance was low for all compliance values. The consistently positive value of the CE was due to the fact that participants were only presented comparison stimuli that were higher than the standard.

Standard		Estimates			Calculated	
s_h	s_v	PSE_H	PSE_V	PSE_{VH}	PSE_{opt}	e_{PSE}
A) congruent standard stimulus [mm/N]						
0.2	0.2	0.2	0.2	0.2	0.2	2.4%
0.9	0.9	1.0	1.0	0.9	1.0	5.7%
1.4	1.4	1.6	1.7	1.5	1.6	7.2%
B) conflicting standard stimulus - haptic reference [mm/N]						
0.2	0.4	**0.2**	0.3	**0.2**	0.2	4.6%
0.9	1.6	**1.0**	1.8	**1.0**	1.1	11.9%
1.4	2.1	**1.6**	2.5	**1.6**	1.9	17.4%
C) conflicting standard stimulus - visual reference [mm/N]						
0.4	0.2	0.5	**0.2**	**0.2**	0.3	32.1%
1.3	0.9	1.6	**1.0**	**1.1**	1.1	3.7%
2.1	1.4	2.5	**1.7**	**1.7**	2.1	23.0%

Table 1: Experimentally-estimated and theoretically-calculated PSEs (PSE_{opt} and e_{PSE} calculated according to equations (4),(9) ; values rounded): PSEs of the combined percept (PSE_{VH}) were similar to the PSEs of the single-modality percepts (PSE_H, PSE_V) that relayed the more compliant information (connection emphasized by bold values). Optimal PSEs according to the maximum-likelihood model differed up to 32.2% from the empirically estimated PSEs of the combined percept (PSE_{VH}).

5.3 Fusion of Compliance Information

Evidence of perceptual fusion is obtained if the PSE of the combined percept (PSE_{VH}) is situated between the single modality PSEs. Furthermore, the reliability of the combined percept should be higher resulting in a smaller standard deviation σ_{VH} compared to the single modality parameters σ_H, σ_V. Maximum likelihood estimates were calculated according to the formulas given in Section 2.

The fourth result is that the data of this experiment did not confirm that the PSE of the combined bimodal percept was an average of single-modality percepts. In particular, there seemed to be no optimally weighted fusion of uniformly distributed information as predicted by the maximum-likelihood theory described in Section 2. It can be seen in Table 1 that the PSEs of the combined percepts were close to the modality that presented the less compliant information. (Optimal PSEs (PSE_{opt}) and relative error between optimal and estimated PSEs (e_{PSE}) were calculated according to equations (4),(9).) As indicated by the bold values in Table 1 participants seemed to be limited by the modality that presented the less compliant information when perceiving bimodal information. The other modality does not seem to add information since the combined percept cannot be denoted as being 'between' the PSEs of the two modalities. Consequently, the PSEs of the incongruent estimates were smaller than the optimal values. Differences were between 4.6% and 32.1% (mean 15.5%). The PSEs of the congruent estimates did not differ notably from the optimal values, but this does not indicate fusion since optimal PSE and single-modality PSEs are the same in the congruent case.

The fifth result confirms the results above based upon the standard deviation of the perceptual estimates. The standard deviations σ that describe the inverse reliability of the estimates are listed in Table 2. (Optimal σ's (σ_{opt}) and relative error between optimal and estimated σ's (e_σ) were calculated according to equations (5),(10).) In the congruent case, participants' reliabilities of the bimodal percept seemed to be closely tied to the reliabilities of the haptic modality (which was reported in the first result to be more reliable than the visual modality when presenting compliance information). In the incongruent case participants' reliabilities of the bimodal percept seemed to be tied to the reliabilities of the modality that presented the less compliant information: In all cases where the reliabilities of the single modality percept differ more than 0.1 mm/N participants' combined estimate was close or equal to the reliabilities of the smaller single-modal percept. Additionally, combined estimates for congruent and incongruent standard stimuli (σ_{VH}) were larger than the values σ_{opt} based on the maximum likelihood theory. According to maximum-likelihood model the slope σ of the combined percept has to be smaller than the smallest slope of the two single-modality percepts indicating that information is added by both modalities and reliability increases. However, in this study, the slopes of the combined estimates σ_{VH} are not smaller than the smallest slope of single modality estimates (only in one case). This indicates that there seemed to be no fusion of the bimodal information.

Stimuli		Estimates			Calculated			
s_h	s_v	σ_H	σ_V	σ_{VH}	σ_{opt}	e_σ	k_h	k_v
A) congruent standard stimulus [mm/N]								
0.2	0.2	**0.1**	0.2	**0.1**	0.1	7.4%	0.8	0.2
0.9	0.9	**0.2**	0.5	**0.2**	0.2	6.7%	0.8	0.2
1.4	1.4	**0.4**	0.8	**0.4**	0.3	28.2%	0.8	0.2
B) conflicting standard stimulus - haptic reference [mm/N]								
0.2	0.4	**0.1**	0.4	**0.1**	0.1	28.2%	1.0	0.0
0.9	1.6	**0.2**	0.6	**0.3**	0.2	27.2%	0.9	0.1
1.4	2.1	0.4	0.5	0.5	0.3	82.4%	0.6	0.4
C) conflicting standard stimulus - visual reference [mm/N]								
0.4	0.2	0.2	0.2	0.1	0.1	33.4%	0.5	0.5
1.3	0.9	0.7	**0.5**	**0.5**	0.4	33.4%	0.3	0.7
2.1	1.4	0.7	0.8	0.9	0.6	71.1%	0.5	0.5

Table 2: Experimentally-estimated and theoretically-calculated slopes (σ_{opt} and e_σ calculated according to equations (5),(10); values rounded): In the congruent case, participants' reliabilities of the bimodal percept seemed to be closely tied to the reliabilities of the haptic modality. In the incongruent case participants' reliabilities of the bimodal percept seemed to be tied to the reliabilities of the modality that presented the less compliant information. All relations are marked bold. Optimal σ's according to the maximum-likelihood model differed up to 82.4% (mean 32.3%) from the empirically estimated σ's of the combined percept (σ_{VH}).

6 Discussion

To summarize the results, we found that discrimination performance increases with decreasing compliance. Furthermore, the combined percept of visual and haptic compliances could not be predicted by a maximum-likelihood estimation of uniformly distributed information. Rather, participants' percept was close to the modality that presented the less compliant stimulus. We now expand on those points.

Discrimination of compliance with our visual-haptic human system interface was found to produce reliable psychometric functions. The estimate of the mean and standard deviation of the perceptual estimate of compliance indicated that people were less able to discriminate compliance, and perceived compliance became more variable, as the baseline level of compliance increased. This was true for stimuli presented haptically and bimodally; performance with purely visual stimuli was relatively poor at all compliance levels. Thus overall, these results indicate that haptic cues to compliance, as simulated by our interface, are not only effective but necessary.

The more accurate performance in detecting rigid (less compliant) environments than detecting soft (compliant) environments might be explained by the fact that exploration of rigid environments leads to a smaller change in finger position than is the case for soft environments. Consequently, the position estimate in compliant environments might be more reliable, with the result that the visual and haptic compliance estimates are more reliable as well.

Our experiments not only tested the efficacy of the interface for conveying compliance, but they also provided tests of the hypothesis that visual and haptic cues would be fused. Contradicting this hypothesis, the results indicated that fusion did not occur, and in particular, the combined percept of visual and haptic compliances could not be predicted by maximum-likelihood of uniformly distributed compliance information. Instead, participants' percept was close to the modality that presented the less compliant stimulus.

The PSEs of the incongruent estimates were smaller than the values predicted from optimal (i.e., maximum-likelihood) integration, instead lying close to the values of the modality presenting the more compliant level of the stimulus. The estimated reliabilities of the percept were greater than optimal, and particularly for haptic stimuli, tended to be closer to the haptic value (which was the less compliant).

The results of the bimodal conditions, in short, clearly violate optimal fusion as described by the maximum-likelihood model. Instead, performance appears to be limited by the discriminability and reliability of the less compliant component of an incongruent bimodal stimulus. Why should this occur? One answer to this question can be derived from our data, which clearly indicate that the lower the compliance being perceive, the more discriminable and reliable the perceptual estimate. If two independent stimulus estimates are produced for a discrepant stimulus, with some signal of discriminability and reliability, a higher-order processor may use the better estimate in making the discrimination judgment. This would lead the lower-compliance stimulus level to dominate judgments, as we observe.

Further research is needed to test these hypotheses, but our results clearly indicate a departure from optimal fusion and hence advance our knowledge of compliance judgments using a bimodal interface.

ACKNOWLEDGEMENTS

This research was funded by the German National Science Foundation (DFG) within the Collaborative Research Center on *High-Fidelity Telepresence and Teleaction* (SFB453). We appreciate the help of students Norbert Gattinger and Pit Hoffmann for implementing and conducting the experiments.

REFERENCES

[1] W. J. Adams, E. W. Graf, and M. O. Ernst. Experience can change the 'light-from-above' prior. *Nature Neuroscience*, 7(10):1057–1058, 2004.

[2] M. Überle and M. Buss. Control of kinesthetic haptic interfaces. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2004.

[3] E. Brenner and J. B. J. Smeets. Fast corrections of movements with a computer mouse. *Spatial Vision*, 16:364–376, 2003.

[4] J.-P. Bresciani, F. Dammeier, and M. O. Ernst. Vision and touch are automatically integrated for the perception of sequences of events. *Journal of Vision*, 6:554–564, 2006.

[5] R. G. Brown and P. Y. C. Hwang. *Introduction to Random Signals and Applied Kalman Filtering*. John Wiley & Sons, 1997.

[6] G. A. Calvert, M. J. Brammer, and S. D. Iversen. Crossmodal identification. *Trends in Cognitive Sciences*, 2(7):247–253, 1998.

[7] G. A. Calvert, C. Spence, and B. E. Stein. *The handbook of multisensory processes*. Cambridge: MIT Press, 2004.

[8] N. Dhruv and F. Tendick. Frequency dependence of compliance contrast detection. In *Proccedings of the ASME Dynamic Systems and Control Division*, 2000.

[9] K. Drewing and M. O. Ernst. Integration of force and position cues for shape perception through active touch. *Brain Research*, 1078:92–100, 2006.

[10] J. Driver and C. Spence. Multisensory perception: Beyond modularity and convergence. *Current Biology*, 10(13):731–735, 2000.

[11] M. O. Ernst and M. S. Banks. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415:429–433, 2002.

[12] M. O. Ernst and H. H. Bülthoff. Merging the senses into a robust percept. *Trends in Cognitive Sciences*, 8(4):162–169, 2004.

[13] F. Freyberger, M. Kuschel, R. L. Klatzky, B. Färber, and M. Buss. Visual-haptic perception of compliance: Direct matching of visual and haptic information. In *Proceedings of the IEEE International Workshop on Haptic Audio Visual Environments and their Applications (HAVE)*, 2007.

[14] G. A. Gescheider. *Psychophysics - Method and Theory.* Hillsdale: J. Wiley & Sons, 1976.

[15] S. Guest and C. Spence. What role does multisensory integration play in the visuotactile pereption of texture? *International Journal of Psychophysiology*, 50:63–80, 2003.

[16] J. M. Hillis, M. O. Ernst, M. S. Banks, and M. S. Landy. Combining sensory information: Mandatory fusion within, but not netween, senses. *Science*, 298:1627–1630, 2002.

[17] J. M. Hillis, S. J. Watt, M. S. Landy, and M. S. Banks. Slant from texture and disparity cues: Optimal cue combination. *Journal of Vision*, 4 (13):1–24, 2004.

[18] D. C. Knill and J. A. Saunders. Do humans optimally integrate stereo and texture information for judgements of surface slant? *Vision Research*, 43:2539–2558, 2003.

[19] D. W. Massaro. Speechreading: illusion or window into pattern recognition. *Trends in Cognitive Science*, 3(8):310–317, 1999.

[20] M. K. O'Malley and M. Goldfarb. The implications of surface stiffness for size identification and perceived surface hardness in haptic interfaces. In *Proceedings of the 2002 IEEE International Conference on Robotics and Automation, Washington DC*, pages 1255–1260, 2002.

[21] F. Pavani, C. Spence, and J. Driver. Visual capture of touch: Out-of-the-body experience with rubber gloves. *Psychological Science*, 11(5):353–359, 2000.

[22] R. Scheuchenpflug. Measuring presence in virtual environments. In M. J. Smith, G. Salvendey, and M. R. Kasdorf, editors, *HCI International 2001, New Orleans*, pages 56–58, 2001.

[23] L. Sciavicco and B. Siciliano. *Modelling and Control of Robot Manipulators*. Springer, 1 edition, 2003.

[24] S. Shimojo and L. Shams. Sensory modalities are not separate modaltities: Plasticity and interactions. *Current Opinion in Neurobiology*, 11:505–509, 2001.

[25] B. E. Stein and M. A. Meredith. *The merging of the senses*. Cambridge: MIT Press, 1993.

[26] 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 & Psychophysics*, 57(4):495–510, 1995.

[27] S. A. Wall and S. A. Brewster. Scratching the surface: Preliminary investigations of haptic properties for data representation. In *Proceedings of Eurohaptics 2003, Dublin, Ireland*, pages 330–342, 2003.

[28] R. B. Welch and D. H. Warren. *Handbook of Perception and Human Performance: Sensory Processes and Perception, Volume 1*, chapter Intersensory Interactions. New York: J. Wiley & Sons, 1986.

[29] F. A. Wichmann and N. J. Hill. The psychometric function: I. fitting, sampling, and goodness of fit. *Perception and Psychophysics*, 63:1293–1313, 2001.

[30] B. G. Witmer and M. J. Singer. Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and Virtual Environments.*, 7(3):225–240, 1998.

[31] S. Yamakawa, H. Fujimoto, S. Manabe, and Y. Kobyayshi. The necessary conditions of the scaling ratio in master-slave systems based on human difference limen of force sense. *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans*, 35(2):275–282, 2005.