Grasping Multisensory Integration: Proprioceptive Capture after Virtual Object Interactions

Johannes Lohmann (johannes.lohmann@uni-tuebingen.de)
Department of Computer Science, Cognitive Modeling, Sand 14
Tübingen, 72076 Germany

Jakob Gütschow (jakob.guetschow@student.uni-tuebingen.de)
Department of Computer Science, Cognitive Modeling, Sand 14
Tübingen, 72076 Germany

Martin V. Butz (martin.butz@uni-tuebingen.de)
Department of Computer Science, Cognitive Modeling, Sand 14
Tübingen, 72076 Germany

Abstract

According to most recent theories of multisensory integration, weighting of different modalities depends on the reliability of the involved sensory estimates. Top-down modulations have been studied to a lesser degree. Furthermore, it is still debated whether working memory maintains multisensory information in a distributed modal fashion, or in terms of an integrated representation. To investigate whether multisensory integration is modulated by task relevance and to probe the nature of the working memory encodings, we combined an object interaction task with a size estimation task in an immersive virtual reality. During the object interaction, we induced multisensory conflict between seen and felt grip aperture. Both, visual and proprioceptive size estimation showed a clear modulation by the experimental manipulation. Thus, the results suggest that multisensory integration is not only driven by reliability, but is also biased by task demands. Furthermore, multisensory information seems to be represented by means of interactive modal representations.

Keywords: Multisensory Integration; Multisensory Conflict; Object Interaction; Virtual Reality

Introduction

Adaptive interaction with the environment requires the combination of various sensory signals. According to theories of predictive coding, this integration is driven by a desire for consistency between internal models and the external world (Friston, 2010), as well as by a desire for consistency across different internal models (Butz, Kutter, & Lorenz, 2014; Ehrenfeld, Herbort, & Butz, 2013). Research on the mechanism of multisensory integration has shown that this consistency is achieved in terms of a maximum likelihood integration which combines different sensory signals based on their respective reliability estimates, resulting in a Bayesian estimate about the state of the external world (Ernst & Banks, 2002; Ernst & Bülthoff, 2004). It is still debated, however, whether this estimate is represented by means of an integrated representation (Cowan, 2001) or by means of separate, modality specific representations which are integrated on demand (Baddeley & Hitch, 1974). Experimental results show strong interactions between modalities in the internal representation, for instance between visual and auditory working memory (Morey & Cowan, 2005). Furthermore, unimodal retrieval from a multisensory representation is affected by previous modal encodings (Thelen, Talsma, & Murray, 2015). Quak, London, and Talsma (2015) suggest that task requirements typically determine whether a unimodal or a complex, multisensory representation is formed.

Our aim in the present study was two-fold. First, we wanted to investigate whether multisensory integration is modulated by task relevance. Second, we wanted to probe the nature of the stored representations. To investigate these questions, we combined an object interaction task involving multisensory conflict with a size estimation task. We let participants perform a grasp-and-carry task in an immersive virtual reality, by tracking the hands of the participants. Conflict was introduced in terms of a visual offset, either expanding or shrinking the visual grip aperture, thereby dissociating vision and proprioception. Moreover, we augmented the object interaction with vibrotactile feedback, which signaled when the relevant object was grasped. After the object interaction, we let participants judge the size of the object they interacted with either visually or based on the grip aperture. If vision and proprioception are integrated, visual estimates should be biased in the same way as proprioceptive estimates. On the other hand, if there was no bias in visual estimates, this would imply an independent storage of modal information.

Method

Participants

Twenty students from the University of Tübingen participated in the study (seven males). Their age ranged from 18 to 34 years ($M = 22.1$, $SD = 3.9$). All participants were right-handed and had normal or corrected-to-normal vision. Participants provided informed consent and received either course credit or a monetary compensation for their participation. Three participants could not complete the experiment due to problems with the motion capture system, only the data of the remaining 17 participants was considered in the data analysis.

Apparatus

Participants were equipped with an Oculus Rift© DK2 stereoscopic head-mounted display (Oculus VR LLC, Menlo

Park, California). Motion capture was realized by the combination of a Synertial IGS-150 upper-body suit and an IGS Glove for the right hand (Synertial UK Ltd., South Brighton, United Kingdom). Rotational data from the suit's and glove's inertial measurement units was streamed to the computer controlling the experiment via a Wifi connection. The data was then used to animate a simplistic hand model in a virtual reality. Since the IGS system only provides rotation data, we used a Leap Motion© near-infrared sensor (Leap Motion Inc, San Francisco, California, SDK version 2.3.1) to initially scale the virtual hand model according to the size of the participants' hands. To allow participants to confirm their size estimates without manual interactions, participants were equipped with a headset. Speech recognition was implemented by means of the Microsoft Speech API 5.4. The whole experiment was implemented with the Unity® engine 5.0.1 using the C# interface provided by the API. During the experiment, the scene was rendered in parallel on the Oculus Rift and a computer screen, such that the experimenter could observe and assist the participants.

To provide the participants with vibrotactile feedback during object interactions, we used two small, shaftless vibration motors attached to the tip of the thumb and the index finger of the participants. The diameter of the motors was 10 mm, the height was 3.4 mm. The motors were controlled via an Arduino Uno microcontroller (Arduino S.R.L., Scarmagno, Italy) running custom C software. The microcontroller was connected to the computer via a USB port which could be accessed by the Unity® program. If a collision between the virtual hand model and an object was registered in the VR, the respective motor was enabled with an initial current of 2.0 V. The deeper the hand moved into the object, the higher the applied current (up to 3.0 V) and the according vibration. At a current of 3.0 V, the motors produced a vibration with 200 rotations per second, the resulting vibration amplitude was 0.75 g. The wiring diagram as well as additional information regarding the components are available online. [1]

Virtual Reality Setup

The VR scenario put participants in a small clearing covered with a grasslike texture, surrounded by a ring of hills and various trees. A stylized container was placed in the center of the scene and served as target for the transportation task (see Fig. 1, left panel). The to-be-grasped and carried object was a cube rendered with a marble texture. The size of the cube varied from trial to trial but the cube always appeared at the same position in the scene. Textual information, like trial instructions and error feedback were presented on different text-fields aligned at eyeheight in the background of the scene.

Centered at the participants' hip[2], the task space covered 60 cm from left to right and 55 cm in depth. Corresponding to the data generated by the IGS suit an upper body rig was placed in the scene. It was positioned about 45 cm in front of the spawning position of the cube, slightly behind the the container. Hence, participants could reach both the container as well as the cube comfortably with their right arm. The rig itself was not rendered, only the right hand of the participants appeared in the scene visually.

The multisensory conflict between visual and proprioceptive grip aperture was realized in terms of a visual angular offset on the root joints of the thumb and index finger. They could be rotated either 10° towards each other, or away from each other. To maintain the same aperture, this visual offset had to be compensated by an adjustment of the actual aperture in the opposite direction. To compensate for a visual offset shrinking the grip aperture, the grip aperture had to be wider, while a visual offset extending the grip aperture required a closer grip aperture. In one third of the trials, no manipulation was applied (the different offset conditions are shown in Fig. 1, right panel).

Procedure

Participants received a verbal instruction at the beginning of the experiment regarding the use and function of the applied VR equipment. Then, they were equipped with the inertial motion capture system, consisting of the suit and the glove. If necessary, the finger sensors of the glove were fixated with rubber bands. After aligning the sensors and enabling the data streaming, the vibration motors were fastened underneath the thumb and index finger tip with rubber bands. Participants were then seated comfortably on an arm chair.

After this, participants were asked to hold their right hand over the Leap sensor to scale the virtual hand size according to their actual hand size. The control was then switched to the IGS system and participants put on the HMD to start the training phase. Participants could practice the grasping and carrying of the cube until they felt comfortable with the task. They had to complete at least 15 successful repetitions of the task before they were allowed to proceed. The grasp and carry task is described in detail in the next section.

After completing the training, the experimenter switched manually to the main experiment. The experiment consisted of eight blocks, each composed of 15 trials. The multisensory conflict between seen and felt grip aperture was introduced during the intertrial interval while the screen was blacked out.[3] In each trial participants had to grasp a cube and put it into the target container. After the object interaction, the scene faded out and one of two possible reproduction scenes

[1] http://www.wsi.uni-tuebingen.de/lehrstuehle/cognitive-modeling/staff/staff/johannes-lohmann.html

[2] Based on the inertial data from the IGS suit, it is possible to calculate a kinematic chain with the hips as root. Hence, the position of the hip joint in the virtual scene is the reference point for all body movements.

[3] While most participants remained unaware to the manipulation and attributed the variance in their grip aperture to inaccuracies of the tracking equipment, two participants reported to be aware of the manipulation after the experiment. Seeing that conscious awareness was not critical in this experiment, we did not perform a behavioral manipulation check in terms of a signal detection task to determine whether participants were able to consciously detect the manipulation of the visual grip aperture.

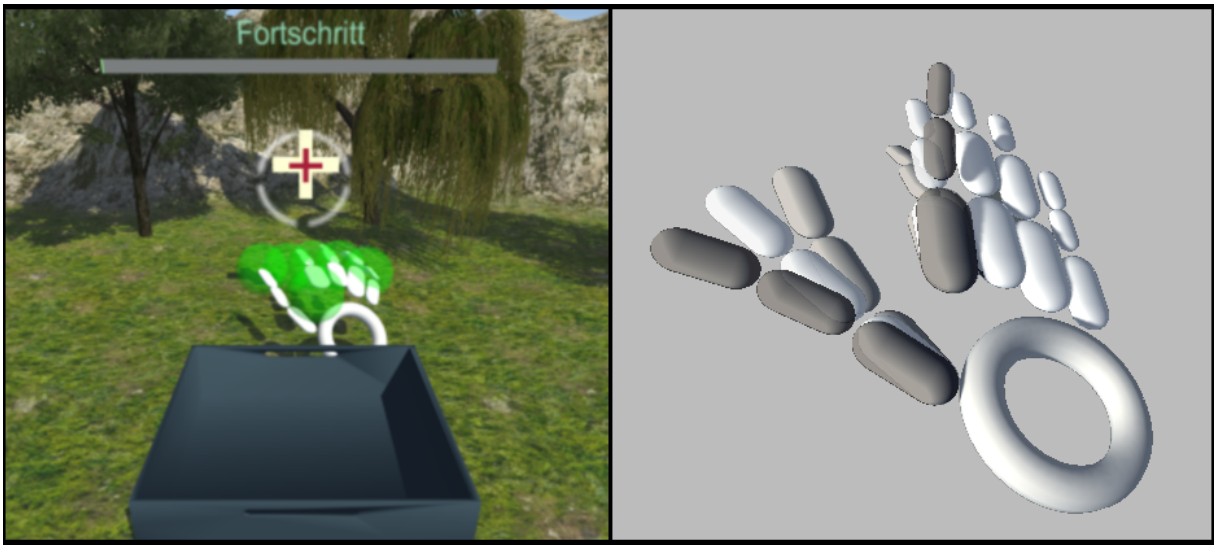

Figure 1: The left panel shows the VR scene and the initial position and fixation checks before the presentation of the target cube. Participants had to maintain a stable fixation on the fixation cross, the green spheres represent the starting position. The right panel shows the different offset conditions. Inward offsets are indicated by the light gray joints, dark gray joints indicate the outward offset condition.

appeared. This was independent of the success in the object interaction, the reproduction scene was also shown in case of error trials. In these scenes participants had to reproduce the size of the cube they interacted with either visually or by indicating the size in terms of a grip aperture. After each block, there was a break of at least ten seconds, after the fourth block, a longer break of at least two minutes was administered. Participants were allowed to put off the HMD during the breaks. After the experiment, participants were asked to complete a presence questionnaire (IPQ, Schubert, Friedmann, & Regenbrecht, 2001). The whole procedure took 90 to 120 minutes, including the preparation and the practice trials.

Grasp and Transportation Task At the beginning of each trial, participants had to move their right hand into a designated starting position, consisting of red, transparent spheres indicating the required positions of the fingers and the palm. The spheres turned green when the respective joints were in position. Furthermore, participants had to maintain a stable looking direction on a fixation cross (see Fig. 1, left panel). When both requirements were met, the fixation cross as well as the visible markers of the initial position disappeared and the target cube appeared. Participants were instructed to grasp the cube with a pinch grasp and to move it into the target container. A successful pinch required the tips of the thumb and the index finger to be placed on opposite sites of the cube and to maintain a stable grip aperture. Participants received vibrotactile feedback whenever touching the cube. The feedback scaled with the depth of penetration, becoming more intense the deeper the fingers were moved into the cube. The task was successfully completed by placing or dropping the cube into the container. Success was indicated by the cube bursting into an explosion of smaller green cubes. Interactions were canceled if the cube was penetrated overly strongly, dropped outside the container, moved outside the reachable space (e.g. by throwing it), or in case the interaction took more than 20 seconds. If one of the conditions was met, participants received error feedback and the trial progressed with the reproduction task.

After completing or failing the interaction, the markers for the initial position reappeared and participants had to move their hands back into the initial position. Then a visual mask was applied, accompanied by random vibrations on the finger tips. The visual and tactile masking commenced for one second. After the masking the scene faded to black and after one second, one of the two reproduction scenes appeared. The offset manipulation was removed during the blank interval.

Size Estimation In both versions of the size estimation task, participants had to reproduce the cube size. For the visual reproduction, the scene was similar to the one in which the interaction took place. However, the ground textures were replaced and different tree models were used to avoid possible comparisons between the cube size and external landmarks. A cube was placed at the center of the scene, at the same position where the cube during the interaction phase appeared. Above the cube, a slider was displayed, which allowed the participants to scale the cube by dragging the slider button with their fingertips. The slider spanned approximately 20 cm from left to right. The initial position of the slider button and thus the initial size of the visual reference cube was determined by the cube size during the interaction phase. For the smaller three sizes the slider started out at 10% and for

the two larger sizes it started out at 90% of the sliding range.

For the proprioceptive reproduction, all visuals were deactivated (including the hand model), only the horizon as well as small white sparks in the center of the scene remained active to remind the participants that the experiment was still running. Participants were instructed to indicate the size of the cube they interacted with by means of the grip aperture between thumb and index finger. To confirm their estimate, participants were requested to say the German word for "continue" or "done" ("weiter" or "fertig"). The voice control identified these commands and ended the trial, recording either the slider position - indicating the visual edge length of the cube - or the grip aperture as the size estimate.

Factors

We varied three factors across trials. First, the edge length of the cube, which had to be interacted with and which size had to be estimated, was either 7 cm, 7.35 cm, 7.7 cm, 8.05 cm, or 8.4 cm. Second, the visual grip aperture was either shrunk, or extended by 10°, or corresponded with the felt grip aperture. In the following, we will refer to visual offsets shrinking the aperture as inward offsets, conversely, we will refer to offsets extending the aperture as outward offsets. Third, we varied the reproduction modality, which could either be visual or proprioceptive. Hence, the experiment followed a 5 × 3 × 2 within-subject design. Each of the 30 conditions was repeated four times, resulting in 120 trials. The trial order was randomized.

Dependent Measures

Besides the size estimates in the two different reproduction conditions, we obtained several time measures. Movement onset was determined as the time between the end of the fixation until leaving the starting position. Contact time refers to the time between movement onset and successful grasp. Interaction time refers to the time interval between the grasp and reaching the container.

Results

Data was aggregated according to the 5 × 3 × 2 within-subject design. Seeing that the size estimation had to be performed after error trials as well, there are no missing data with respect to the size estimates. For the duration measures, only correct trials were considered. The overall error rate was high (nearly 30%), due to the task complexity. In case of missing time data, the respective cell mean was interpolated within participants by the mean over all conditions with the same offset type. For all dependent measures, values differing more than two times of the standard deviation from the mean were excluded, which was the case for 2% of all data points.[4]

Size estimates, time measures, and error rates were analyzed with repeated measures ANOVAs using R (R Core Team, 2016) and the *ez* package (Lawrence, 2015). All post-hoc t-tests were adjusted for multiple comparisons by the method proposed by Holm (Holm, 1979). Results from the presence questionnaire were compared with the reference data from the online database.[5] There were no significant differences.

Table 1: ANOVA table for the analysis of the **size estimates**. The assumption of sphericity was violated for the cube size factor and the interaction between offset and reproduction condition, the according p-values were subjected to a Greenhouse-Geisser adjustment.

factor	df	F	p	η_p^2
size	4	34.84	$< .001^*$	.69
offset	2	17.55	$< .001^*$	.52
repro. type	1	0.48	.50	.03
size× repro. type	4	2.94	$.027^*$	.16
offset × repro. type	2	3.95	$.045^*$	.20
size × offset	8	1.03	.42	.06
size × offset × repro. type	8	2.35	$.022^*$	.13

Size Estimates

Data were analyzed with a 5 (cube size) × 3 (offset) × 2 (reproduction type) factors repeated measures ANOVA. Results are shown in Tab. 1. The analysis yielded significant main effects for cube size and offset. The main effect for cube size matches the actual cube size: larger cubes were estimated larger and smaller cubes were estimated smaller. To check if the estimates were veridical, we tested whether the estimated cube sizes differed from the actual cube sizes. None of the respective comparisons yielded significant results.

With respect to the main effect of offset, participants overestimated the cube size in case of inward offsets, compared to conditions with no offset ($t(16) = 3.45$, $p = .007$). For outward offsets participants underestimated the cube size, compared to conditions with no offset ($t(16) = 2.98$, $p = .009$). Finally participants provided larger estimates in case of inward, compared to outward offsets ($t(16) = 5.23$, $p < .001$).

Both, cube size and offset interacted with the reproduction condition. The interaction between cube size and reproduction type is due to a systematic overestimation of the larger cubes in case of the visual reproduction. In both cases, the estimates are significantly larger than the actual sizes of 8.05 cm ($t(16) = 4.26$, $p = .003$), and 8.4 cm ($t(16) = 3.21$, $p = .022$), respectively.[6]

The interaction between reproduction condition and offset was further analyzed with post-hoc t-tests. Estimates in case of outward offsets were significantly smaller than in case of

[4] Please note that the data pattern remains nearly unaffected if the data is not filtered. Removing the size estimates from error trials only reduces the effect size of the three-way interaction.

[5] Available at http://www.igroup.org/pq/ipq/index.php

[6] The considerable overestimation might be partially due to the initial slider position in the visual reproduction, starting at 90% of the sliding range for larger cubes.

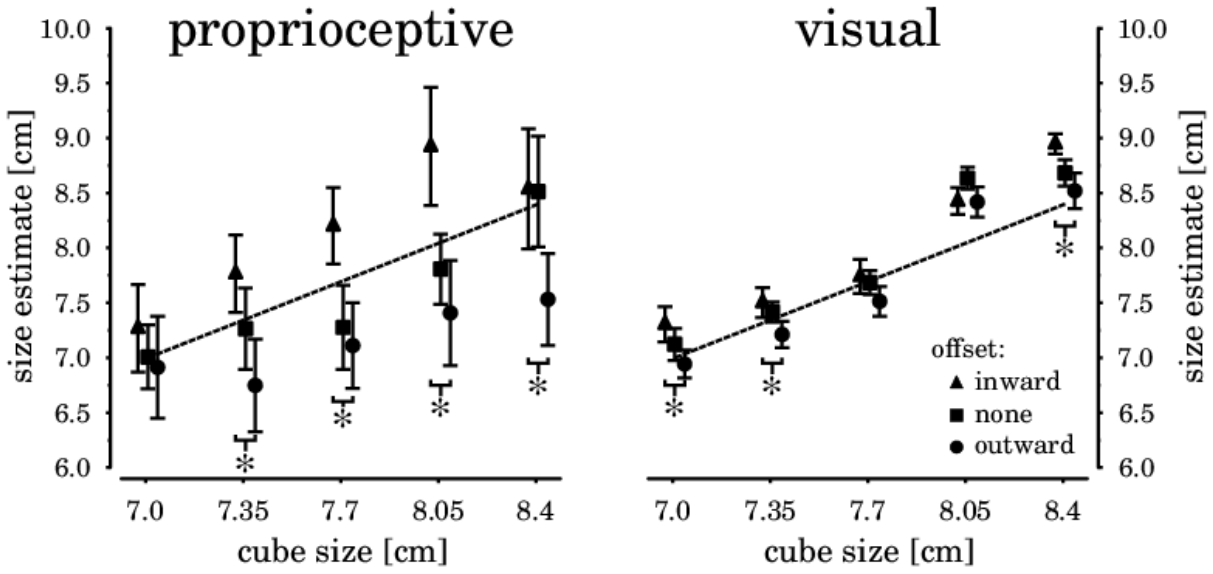

Figure 2: Three-way interaction between reproduction condition, cube size and offset. Significant differences with $p < .05$ between estimates in case of inward and outward offsets are indicated by an asterisk. The respective t-tests were one-sided (inward $>$ outward) and were adjusted for multiple comparisons. The dashed line indicates the actual cube size.

inward offsets, both, for visual ($t(16)$ = -2.21, $p = .021$), as well as for proprioceptive ($t(16)$ = -5.48, $p = .002$) reproduction. However, the differences between the offset conditions were much more pronounced in case of proprioceptive reproduction, resulting in the observed two-way interaction.

This pattern of results was modified by a three-way interaction between cube size, offset and reproduction condition. Separate ANOVAs for the different cube sizes showed that the interaction between reproduction condition and offset was only present for cubes of intermediate (7.7 cm) and large size (8.05 cm). For these two conditions, there were no significant differences between the offset conditions in case of visual reproduction. The differences for proprioceptive reproduction remained significant. The main effect of offset, however, remained significant for all of these separate analyses.

With respect to our hypotheses, the difference between inward and outward offsets is most relevant. To check whether inward offsets always result in larger estimates than outward offsets, we checked whether the respective difference is significant for the five different cube sizes, separately for the two reproduction conditions. In case of proprioceptive reproduction, the difference is significant for all cube sizes, except the smallest one of 7 cm. For visual reproduction the differences reached significance for all cube sizes, except the intermediate (7.7 cm) and large size (8.05 cm). The results are shown in Fig. 2.

Time Measures

Data were analyzed with a 5 (cube size) $\times$ 3 (offset) factors repeated measures ANOVA. No significant effects were found for the movement onset times. The analysis of object contact times yielded a significant main effect for offset ($F(2,32)$ = 76.57, $p < .001$, η_p^2 = .83). Slowest contact times were observed for outward offsets, while inward offsets yielded the fastest response times. All of the respective pairwise comparisons yielded significant results. The analysis of the interaction times yielded a significant main effect for offset as well ($F(2,32)$ = 4.90, $p < .014$, η_p^2 = .23). Participants were slower in transporting the cube in case of outward offsets. Post-hoc t-tests showed that the interaction times were significantly elevated in case of outward offsets, both compared to inward offsets ($t(16)$ = 2.39, $p = .042$), as well as to trials without offset ($t(16)$ = 2.42, $p = .042$).

Error Rates

The analysis of the error rates yielded significant main effects for cube size ($F(4,64)$ = 4.27, p = .004, η_p^2 = .21) and offset ($F(2,32)$ = 12.22, $p < .001$, η_p^2 = .43). In general, participants made fewer errors during interactions with larger cubes. Furthermore, error rates were higher in case of inward offsets. Post-hoc t-tests showed that error rates increased for inward offsets, when compared to both outward offsets ($t(16)$ = -3.67, $p = .004$), and no offsets ($t(16)$ = -4.56, $p < .001$).

General Discussion

Previous studies on multisensory integration have shown a dominance of visual information in the perception of object size (e.g. Ernst & Banks, 2002). To investigate whether task demands, which require to focus on another modality, can reduce this dominance, we let participants perform a grasp-and-carry task under multisensory conflict between vision and proprioception. In order to do so, we manipulated the mapping between seen and felt grip aperture. After the ob-

ject interaction we let participants estimate the size of the object they interacted with – either visually or by providing a proprioceptive estimate via grip aperture. Our results show a systematic bias in the size estimates due to the introduced offset between seen and felt grip aperture. A wider grip aperture resulted in object size overestimations, while a smaller aperture yielded underestimations. This was true for both, visual and proprioceptive size estimates. Hence, the adaptation of the size estimation followed the proprioceptive adaptation, which was necessary to compensate for the visual offset.

While the offset manipulation led to different actual grip apertures for cubes of the same size, the visual impression of both the cube size and the grasp of the virtual hand remained the same. Thus, if the size estimate was dominated by the visual impression, there should have been no effect of the offset condition in the visual reproduction trials. In contrast, our results show a clear influence of proprioceptive information on the size estimates in both modalities. However, this influence was much more pronounced in the case of the proprioceptive reproduction. Apparently, proprioceptive information dominated the resulting percept, even if proprioception was much noisier than vision, indicated by the comparatively large variance in the proprioceptive size estimates.

The combination of VR with motion capturing enabled us to dissociate vision and proprioception in an interactive setup. Compared to previous studies, which investigated the effects of mismatching sensory information regarding an object, the applied setup allows to manipulate the own body perception without affecting the visual impression of the external, virtual world. Some issues with respect to the experimental setup remain. The high error rates imply that even with the vibrotactile augmentation, the object interaction remained difficult for the participants. Especially in case of outward offsets, participants took quite long to grasp and carry the cube. The error rates were elevated for inward offsets, which were associated with the fastest grasping and interaction times, implying a speed accuracy trade-off. Furthermore, our setup did not comprise a control condition without grasping. Including trials which only require touching the object will clarify whether the mere presence of a graspable object yields a bias towards proprioceptive information, or if performing the actual interaction is necessary to induce the bias.

Despite these issues, the results allow us to draw the following two conclusions. First, visual and proprioceptive information regarding the object size seem to be stored separately, but are able to affect each other. If there was only a single percept reflecting the cube size across modalities, then the reproduced size should be independent of the reproduction modality. This is clearly not the case, given the huge difference in the variance of the visual and proprioceptive estimates and the stronger bias in proprioceptive compared to visual reproduction. This conclusion dovetails with results reported by (Ernst & Banks, 2002), who showed that sensory data are stored separately, when they originate from different modalities. Second, the integration process that produces a visual or a proprioceptive estimate is influenced by the type of reproduction. The considerable difference between the effect sizes implies a different weighting of the modality-specific encodings in the two reproduction conditions.

References

Baddeley, A. D., & Hitch, G. (1974). Working memory. *Psychology of learning and motivation*, *8*, 47–89.

Butz, M. V., Kutter, E. F., & Lorenz, C. (2014). Rubber hand illusion affects joint angle perception. *PloS One*, *9*(3), e92854.

Cowan, N. (2001). The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behavioral and brain sciences*, *24*(1), 87–114.

Ehrenfeld, S., Herbort, O., & Butz, M. V. (2013). Modular neuron-based body estimation: maintaining consistency over different limbs, modalities, and frames of reference. *Frontiers in Computational Neuroscience*, *7*(Article UNSP 148).

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*(6870), 429–433.

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

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–138.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 65–70.

Lawrence, M. A. (2015). ez: Easy analysis and visualization of factorial experiments [Computer software manual]. Retrieved from `https://CRAN.R-project.org/package=ez` (R package version 4.3)

Morey, C. C., & Cowan, N. (2005). When do visual and verbal memories conflict? the importance of working-memory load and retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(4), 703.

Quak, M., London, R. E., & Talsma, D. (2015). A multisensory perspective of working memory. *Frontiers in human neuroscience*, *9*.

R Core Team. (2016). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `https://www.R-project.org/`

Schubert, T., Friedmann, F., & Regenbrecht, H. (2001). The experience of presence: Factor analytic insights. *Presence*, *10*(3), 266–281.

Thelen, A., Talsma, D., & Murray, M. M. (2015). Single-trial multisensory memories affect later auditory and visual object discrimination. *Cognition*, *138*, 148–160.