DOCTOR OF PHILOSOPHY

Visual-haptic integration during tool use

Takahashi, Chie

Award date:
2012

Awarding institution:
Bangor University

Visual-haptic integration during tool use

Chie Takahashi, BSc, MSc

Submitted for the degree of Doctor of Philosophy

2012

School of Psychology

Bangor University

Declarations

This work has not been previously accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Signed ……..... (candidate)

Date ……....

STATEMENT 1

This dissertation is the result of my own independent work/investigations, except where otherwise stated. Other sources are acknowledged by footnotes giving explicit references. A list of references is appended.

Signed ……..... (candidate)

Date ……....

STATEMENT 2

I hereby give consent for my dissertation, if accepted, to be available for photocopying and for inter-library loans, and for the title and summary to be made available to outside organisations.

Signed ……..... (candidate)

Date ……....

Acknowledgements

This would not have been possible without support and guidance from many important people in my life.

I would first and foremost like to thank Dr. Simon Watt, who offered me the opportunity to undertake this PhD. Without his endless patience and excellent supervision (his guidance and expertise) throughout my project, it would not have reached fruition. My research would have been considerably difficult if he had not been available to provide support and direction. I will be eternally grateful for his immense enthusiasm, inspired ideas, and remarkable breadth of knowledge. I was exceptionally lucky to have the benefit of his advice and suggestions all through the course of my research.

I would like to thank my supervisory team – Dr. Simon Watt, Dr. Diedrichsen, and Dr. Robert Ward, for their constant encouragement and guidance. I would also like to thank Prof. Steven Tipper and Dr. James Intriligator for bridging me from Undergraduate to Masters to start this project supervised by Dr. Simon Watt.

I am extremely grateful to the School of Psychology at Bangor University. This exceptional department has provided me with a wealth of information and the most incredible team of administration staff. The department also supported me financially and gave me an opportunity as a teaching assistant.

Thanks to all of my participants in my psychophysics experiments. My research would not have been possible without their willingness to spend hours and hours of their time, using their eyes and fingers in order to respond to repetitive stimuli in virtual environment.

I would like to thank the ORSAS and EPSRC for funding this research project.

Thanks to Dr. Kevin MacKenzie, Ms. Wendy Adams and Ms. Louise Ryan for their precious time and useful comments on my thesis throughout the proof reading process.

Especially, thanks to Kevin for his comments and advice proved invaluable.

Thanks to my lovely lab members, Kevin, Dr. Bruce Keefe, and Louise.

I would like to take this opportunity to express my gratitude to my friends for their love and patience. This project would never have been completed if they had not been there in my life.

I would like to thank my mother and father. Finally, I would particularly like to thank my husband, Kiyoshi, for his support and patience. He has understood my dream and kept encouraging me throughout my studying abroad. He is my soul mate in the world.

My apologies to anyone I have missed out.

Table of contents

Abstract

To integrate visual and haptic information effectively, the brain should only combine information that refers to the same object. Thus, it must solve a 'correspondence problem', to determine if signals relate to the same object or not. This could be achieved by considering the similarity of the two sensory signals in time and space. For example, if two size estimates are spatially separated or conflicting, it is unlikely that they originate from the same object; so sensory integration should not occur. Humans are adept at using tools such as pliers, however, which can systematically change the spatial relationships between (visual) object size and the opening of the hand. Here we investigate whether and how the brain solves this visual-haptic correspondence problem during tool use. In a series of psychophysical experiments we measured object-size discrimination performance, and compared this to statistically optimal predictions, derived from a computational model of sensory integration. We manipulated the spatial offset between seen and felt object positions, and also the relative gain between object size and hand opening. When using a tool, we changed these spatial properties by manipulating tool length and the pivot position (for a pliers-like tool). We found that the brain integrates visual and haptic information near-optimally when using tools (independent of spatial offset and size-conflict between raw sensory signals), but only when the hand opening was appropriately remapped onto the object coordinates by the tool geometry. This suggests that visual-haptic integration is not based on the similarity between raw sensory signals, but instead on the similarity between the distal causes of the visual and haptic estimates. We also showed that perceived size from haptics and the haptic reliability were changed with tool gain. Moreover, cue weights of the same object size were altered by the tool geometry, suggesting that the brain does dynamically take spatial changes into account when using a tool. These findings can be explained within a Bayesian framework of multisensory integration. We conclude that the brain takes into account the dynamics and geometry of tools allowing the visual-haptic correspondence problem to be solved correctly under a range of circumstances. We explore the theoretical implications of this for understanding sensory integration, as well as practical implications for the design of visual-haptic interfaces.

Chapter One

Introduction

1.1. Overview

Humans are adept at using tools such as scissors and pliers. Such "hand tools" have been diverse and well developed over human history. This ability is indispensable in both everyday life and specific situations which require proficient use of tools. In contemporary society, various haptic devices have been rapidly revolutionised by integrating with advanced computer technology and its applications. Modern robotics and virtual technologies enable to widen human reachable space from micro to macro scale. These systems often introduce an unnatural spatial mapping and force hand movements to be rescaled. To manipulate such devices effectively and precisely, it is important to understand the brain mechanisms of human sensory information processing involved in tool use. Understanding the mechanisms would be beneficial to design a new human-machine interface of haptic tools and to generate a "sense of reality" when using a tool in the computer-simulated virtual environment.

A body of neurophysiological and neuropsychological research has investigated the neural mechanisms underlying tool use (Maravita & Iriki, 2004; as a review). A lasting and powerful concept is that tools become incorporated into the body representation, or "body schema". Recent brain imaging studies suggest functionally specialised neural networks for tool use and some studies support the body schema idea (e.g. Johnson-Frey, 2003, 2004; Lewis, 2006). However, the mechanisms are still very much under debate.

From a different perspective of multisensory processing, when grasping an object either with direct touch by hand or using a tool, we "see" and "feel" the object simultaneously. In both cases, estimates of object properties (e.g. size and shape) are available from both vision and haptics. Several studies have shown that in order to perceive the three-dimensional (3-D) world accurately and precisely, the brain integrates information across multiple sensory modalities (Ernst, 2005; Ernst & Bülthoff, 2004; as reviews). Previous research has quantitatively shown that when normally grasping an object with the hand, information from vision and haptics is integrated; exploiting the redundancy in both signals results in better performance than would be possible from either signal alone (Ernst & Banks, 2002). However, the integration reduces when visual and haptic signals are spatially separated (Gepshtein, Burge, Ernst, & Banks, 2005).

These findings reflect how the brain solves a fundamental problem, termed "correspondence problem". For sensory information to be combined effectively the brain should only integrate sensory signals that refer to the same source (and it should not combine them if they come from different sources). Several studies demonstrated that this could be achieved by considering the statistical similarity of sensory signals in terms of coincidence across time, space, and magnitude of signals (Deneve & Pouget, 2004; Ernst, 2005, 2007; Körding, Beierholm, Ma, Quartz, Tenenbaum, & Shams, 2007a). However, tools complicate the correspondence, because they systematically change the spatial relationship between (seen) object size and the opening of the hand, as well as perturbing the spatial locations of each signal.

In addition, anecdotal reports suggest that when we pick up tools we often have the sense that we can feel the acting tip of the tool, even though haptic information is only available at the hand. Converging previous research and such intuition has raised a basic scientific question: "How are our percepts of the 3-D space computed during tool use?" and also an applied question: "How could we design tools to optimise sensory capability for their practical use?" To answer these questions, this thesis explores the brain mechanisms in tool use and discusses it from a general statistical standpoint, within the framework of multisensory integration. This chapter, as an introduction, describes background knowledge with a particular

emphasis on multisensory integration and the correspondence problem, and also provides a review of relevant tool use literature.

1.2. Sensory integration

1.2.1. Combining sensory information from multiple cues

Human sensory organs such as eyes and ears are constantly exposed to various noisy signals from the outside world. The brain simultaneously and continuously receives an enormous amount of sensory input from the environment. Sensory information about the world is generally decomposed into various cues, which indicate the physical entities and statistical structure of the world (Ernst & Bülthoff, 2004; Knill & Pouget, 2004; Angelaki, Gu, & DeAngelis, 2009). When perceiving the 3-D structure in the world, the brain combines different sources of information, for example, binocular disparity, texture, motion, haptics, and audition.

A variety of mechanisms are involved to render a unique interpretation from highly varied sensory signals into the class of perceptual properties (e.g. colour, lightness, orientation, speed, direction, depth, shape and size). However, sensory signals are all corrupted by noise and are often perceptually ambiguous. In other words, the sensory system is persistently confronted with uncertainty due to noisy and ambiguous information. The nervous system therefore has to solve an inverse problem by converting sensory inputs into information about the physical property. To generate a unified percept, levels of uncertainty, caused by sensory noise and ambiguity, need to be reduced. One effective way is to combine available redundant cues to minimise the sensory noise.

Von Helmholtz postulated that cues are utilized to probabilistically best-guess the state of the world (Körding et al., 2007a). This is well known as Helmholtz's notion of unconscious inference, which describes making conclusions from noisy data, based on prior experience and assumptions about the world. More specifically, estimates would be inferred from noisy sensory inputs considering previous meaningful perceptual experience interacting with the current environment.

In such a way, the brain can infer the most likely state of the world from given sensory information. (Clark & Yuille, 1990; Ernst, 2005; Ernst & Bülthoff, 2004; Hillis, Ernst, Banks, & Landy, 2002; Knill, 2003; Landy, Maloney, Johnstone, & Young, 1995; Yuille & Bülthoff, 1996).

Because sensory information is statistically structured and the brain can, in principle, perceive the world using statistical inference, the probabilistic estimation can be mathematically predicted by Bayes' theorem. Bayesian approaches have modelled human perception as a process of statistically estimating the most probable states of the world through available sensory information and past sensory experience (Ernst, 2005; Knill & Richards, 1996). Bayesian models incorporate knowledge of the statistics of the world to compute the most statistically likely state. This approach has been applied to simulate many aspects of human perception and sensorimotor behaviour (e.g. Körding & Wolpert, 2006).

Bayesian framework and Maximum Likelihood Estimation

When estimating a physical property (e.g. an object's size) in the world, the most reliable estimate would be obtained by using all available cues. However, sensory cues are noisy signals caused by sensory measurements and are often ambiguous to estimate the property. Therefore, the nervous system likely employs not only current sensory information but also the previous knowledge and experience to estimate the properties (Knill & Richards, 1996). Consider a case when the brain estimates an object's size from the retinal images. If the source is uncertain or unknown, the nervous system might not be able to derive reliable size estimate. For example, when estimating the size of a cup from the retinal image, the estimate could range from a drinking cup size to a spinning vehicle "Teacup" size at an amusement park. However, once the system knows that it is a coffee cup on a table, the estimate could be constrained to indicate it is between 5cm and 10cm. This estimate would be much more reliable than when the source is unknown. As such, knowledge of the sensory information can be crucial to estimate the property, and the nervous system likely employs this strategy to reduce the ambiguity in natural situations.

The Bayesian approach introduces the concept of prior knowledge of the world state. A Bayesian model simulates human perception as a process of statistically estimating the most probable state of the world, employing both available current sensory information and past sensory experience (Ernst, 2005; Knill & Richards, 1996). In general, when observing the world state, Bayes' rule can be formulated as below (Knill & Richards, 1996):

$$P(state|observation) = \frac{P(observation|state)P(state)}{P(observation)}$$

For example, consider a physical property (S) and available sensory information (I) being observed. The likelihood function, $p(I|S)$, gives the probability of a possible value, produced by the current sensory information. The prior function, $p(S)$ provides the probability of a given value, which is inferred from the past sensory data or experience, independent of the current sensory data. The posterior distribution, $p(S|I)$, conveys the probability of all possible values, which is the product of the likelihood and prior.

$$p(S|I) = \frac{p(I|S)p(S)}{p(I)} \qquad (1.1)$$

This formula illustrates a fundamental rule of the Bayesian model; that is, how the information about the physical property is transformed by a sensory signal and past experience. $p(I)$ can be calculated by the sum of $p(I|S)$ and $p(S)$. The sum is generally a constant, denoted as C, and is considered a normalising constant.

$$p(S|I) = \frac{p(I|S)p(S)}{C} \propto p(I|S)p(S) \qquad (1.2)$$

Because sensory information is inevitably noisy, it falls onto a distribution of the possible physical properties consistent with a given cue. The likelihood distribution $p(I|S)$ is generally modelled as a Gaussian probability distribution with a centre located at the most likely value given the current sensory information. If the prior probability distribution is similarly modelled as a Gaussian, its centre represents the most likely value inferred from the past knowledge or experience. If both the likelihood and prior are Gaussian distributions, their product, the posterior probability distribution, would also be Gaussian. The posterior does not only represent a single

point estimate of the property; instead it provides the probabilities of all possible values given the current sensory input and past sensory experience. Thus, the posterior distribution is determined by the relative variances, or reliabilities, of the likelihood and the prior. For instance, when the prior distribution does not change and the variance of the likelihood distribution increases with holding its peak position, the peak of the posterior distribution is shifted toward the prior. Reducing the relative reliability of the likelihood leads to an increase in the weight of the prior, and then the prior contributes more to determine the shape of the posterior distribution. This means that as the current sensory information becomes less reliable, the brain relies more on the past experience, or knowledge, for estimating the physical property. Conversely, when increasing the variance of the prior, the brain relies more on the current sensory information than the past experience.

Note that when the brain estimates the target property some decision needs to be made, based on the probability distribution given by the posterior. One such decision rule in models of human perception, which is termed as the Maximum a Posteriori (MAP) estimate, postulates that the brain is more likely to choose the most probable state from posterior distribution (Ernst & Banks, 2002; Mamassian, Landy, & Maloney, 2002). The MAP estimate can be obtained from the posterior distribution:

$$I_{MAP} = \underset{I}{\operatorname{argmax}} \{p(S|I)\} = \underset{I}{\operatorname{argmax}} \{p(I|S)p(S)\} \qquad (1.3)$$

Bayesian inference is used for calculating the posterior state interacting with the likelihood function of sensory information over possible hidden variables, or parameters, as the prior. That is, the Bayesian model probes the probability of the property, incorporating likelihood distribution with some known prior distribution of the property, into the decision. As such, the advantage of the Bayesian model is to use inferences of the world to obtain the final estimates. In other words, the prior distribution could reduce the observation variability and the uncertainty. Theoretically, the prior knowledge or experience can be shaped through learning processes and can provide some probabilistic bias or inference to the estimate. However, the complete knowledge of the probabilistic structure is rarely obtained. It has been argued that the

prior distribution is hardly determined and there is no evidence whether the estimated prior actually represents the distribution (c.f. Burge, Fowlkes, & Banks, 2010).

When all possibilities of the targeted physical property are equally likely when based on previous knowledge, the Bayesian prior would be described by a uniform distribution. Under such a situation, the contribution of the prior to the final percept can be ignored, so Equation (1.2) can be simplified: $p(S|I) \propto p(I|S)$. The peak of the posterior distribution would be determined only by the likelihood function. The decision would be made on the final percept by choosing the peak. This process is known as maximum likelihood estimation (MLE) and Equation (1.3) can be:

$$I_{MLE} = \underset{I}{\operatorname{argmax}} \{p(I|S)\} \tag{1.4}$$

As such, both Bayesian and MLE are probabilistic approaches that can explain how the brain perceives the world. The difference is that the MLE model assumes the prior distribution is uniform and the estimators derived from different cues are unbiased. The unbiased estimators mean that the average of perceptual estimates across different cues is equal to the true value. The advantage of the MLE model is to simply employ likelihood distributions of sensory cues for the final estimate and it does not contain free parameters in the estimate such as the prior $p(S)$. However, because the MLE finds the best estimate based on only sensory information, it relies more on the observation sampling than the Bayesian estimation; so a different set of observations would result in a different estimate (high sampling variability). Moreover, it is impossible to determine whether the estimators are unbiased and they may have already introduced unknown biases prior to their final estimate.

In addition, when comparing the MLE and Bayesian models, the Bayesian model provides a more conservative estimate than MLE when dealing with noisy sensory data. Note that if the likelihood probability is more reliable than the prior (in other words, considerable amount of observation data are available) or if the prior distribution is uniform (in other words, all states are equally likely), the Bayesian and MLE estimates would become equivalent. From these considerations, the MLE can be considered as a special case of the Bayesian.

Weighted linear cue combination and optimal cue integration

The human perceptual system likely combines different sensory inputs in order to increase the reliability of its final estimates. There are different approaches to explore the mechanisms of multisensory cue integration and the simplest model is a weighted linear cue combination (Trommershäuser, Körding, & Landy, 2011). Based on the above comparison between the Bayesian and the MLE models, unbiased estimators of the MLE model simply allow us to make an optimal criterion by minimising their variances. In other words, the MLE is a promising model of a statistically optimal cue combination (Ernst & Banks, 2002). Several studies have demonstrated optimal integration such that the variance of the final estimate is minimised, using this weighted linear cue combination model.

Here describe the weighted linear cue combination model under the MLE assumption is varied. The MLE model assumes that noises from different sensory modalities are independent and Gaussian distributed. If estimates are on average unbiased, the combined estimate can be obtained by a weighted sum of estimates from each cue, where individual weights are ruled by the reciprocal of their normalised variance (for a review see Oruç, Maloney, & Landy, 2003). We describe this reliability-weighted rule using key equations here. The reliability of each cue is given by the inverse of its variance, where the standard deviation σ (sigma) is the square root of the variance (see Equation (1.5)) and represents a measure of uncertainty or the precision of sensory measurements. Each cue weight is determined by the reliability, which is proportional to the inverse variance of each estimator, and is normalized such that the sum of all weights is 1 (see Equation (1.6)). This normalisation ensures that the combined estimate is also unbiased.

$$r_i = {}^{1}\!/_{\sigma_i^2} \tag{1.5}$$

$$w_i = \frac{{}^{1}\!/_{\sigma_i^2}}{\sum_{j=1} {}^{1}\!/_{\sigma_j^2}}, \qquad \sum_i w_i = 1 \tag{1.6}$$

According to the MLE principle, a final estimate of physical properties in the world is achieved by combining different cue estimates weighted by their reliabilities, in a linear fashion (weighted linear summation):

$$\hat{S} = \sum_i w_i \hat{S}_i \qquad (1.7)$$

where $\hat{S}_i$ is an output of the *i*-th estimator and w_i is the weight of the *i*-th estimator. When multiple cues are available, the most likely property can be calculated from a weighted linear combination of each estimate and cue reliability. Therefore, more reliable, or less variable cues are assigned larger weights in producing the final estimate (Hillis, Watt, Landy, & Banks, 2004; Landy et al., 1995; Oruç et al., 2003). Importantly, if cue integration occurs, the variance of the combined estimate is always less than that of any single estimate:

$$\sigma^2 = \frac{\prod_i \sigma_i^2}{\sum_i \sigma_i^2} \; < \min\,(\,\sigma_i^2) \qquad (1.8)$$

Thus a key advantage of sensory integration is to provide perceptual estimates with greater precision than can be achieved with either signal alone (Clarke & Yuille, 1990; Knill & Pouget, 2004; Landy et al., 1995; Oruç et al., 2003; Yuille & Bülthoff, 1996). This implies that the nervous system should not select one sensory modality and ignore the others, but instead integrate information from multiple signals.

A statistically ideal observer takes an optimal strategy for integrating signals from different sources. To simply explain how an ideal observer performs, here we illustrate a two-cue case. Imagine when two sensory signals (e.g. vision and haptics) are available. (See Figure 1.1). As described above, assume that both estimates ($\hat{S}_V$ from vision and $\hat{S}_H$ from haptics) are unbiased and independent, and are subject to independent Gaussian noise with variances of σ_V^2 and σ_H^2, respectively. $1/\sigma_V^2$ and $1/\sigma_H^2$ refer to the reliabilities of the two cues. Signals with large variance (thus low reliability) are given lower weight to combine different sensory inputs. The combined visual-haptic estimate $\hat{S}_{VH}$ is a weighted sum of two estimates: $\hat{S}_{VH} = w_V\hat{S}_V + w_H\hat{S}_H$. On the above assumption, the variance of the combined estimate is $\sigma_{VH}^2 = \sigma_V^2\sigma_H^2/(\sigma_V^2 + \sigma_H^2)$.

Figure 1.1 illustrates combining visual and haptic likelihood functions, or probability densities of the size estimates, based on the MLE model. The most probable estimate is at the maximum of the combined distribution. When the variance of each single-cue distribution changes, the position of the combined estimate moves toward the more reliable cue, indicating cue weights relatively alter based on their reliabilities.

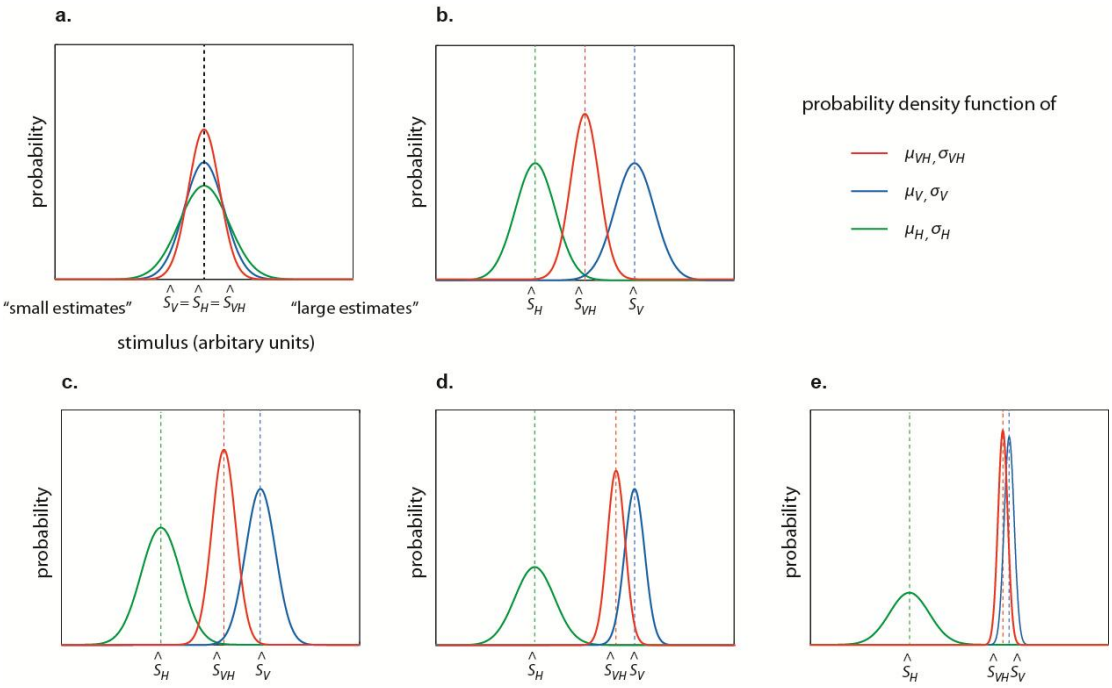


Figure 1.1. Ideal observer's performance based on MLE model. Sensory signals can be described by probability density function. The red line represent combined cue estimates and the blue and green lines are individual cue estimates respectively. Here, the blue line represents vision and the green represents haptics. The variance of combined estimate is calculated by the Equation (1.8) and the mean of combined estimate is calculated by the Equation (1.7) and linearly weighted based on two cue variances. **a.** normal situation $\left(\hat{S}_V \equiv \hat{S}_H\right)$. Different sigma represents a degree of uncertainty or precision. When two sensory cues have equal mean, combined sigma is smaller than either individual cue. **b.** a cue conflict condition ($\hat{S}_V \neq \hat{S}_H$). When two sensory cues have equal variance, cue weights are equal. Therefore the combined estimate is located between two estimates. From **b.** to **e.**, the variance of one cue (here, vision) changes from large (equal to haptics) to small, and the weights change from equal to visual dominance. The combined estimates shift toward visual estimate.

When $\hat{S}_V \equiv \hat{S}_H$, their weight may be different because of the differences in variances, but the combined estimate do not change (Figure 1.1 (a)). In a cue conflict condition ($\hat{S}_V \neq \hat{S}_H$), the mean of combined estimate does change according with the weight

variation. For example, when the variance of vision is equal to the variance of haptics, the combined estimates locates at the middle of both estimates (Figure 1.1 (b)). When the variance of vision becomes relatively smaller than the variance of haptics (from Figure 1.1 (c) to (e)), the combined estimate shifts toward the visual estimate.

A number of studies have shown that the final percept of combined cues can be explained by reliability-based linear cue combination rule (e.g. Knill, 2003; Hillis et al., 2004; Oruç et al., 2003). A statistical approach, or the MLE framework, provides combined estimates of multisensory perception via statistically optimising probabilities. This MLE model predicts an ideal observer's behaviour for a variety of sensory inputs and perceptual tasks. For example, when estimating surface slant, texture and binocular disparity cues are combined with a statistically optimal manner, consistent with the MLE model (Hillis et al., 2004).

1.2.2. Combining information from vision and haptics

When an observer estimates an object's size by looking at and feeling the surface simultaneously, vision and haptic cues are both able to estimate its size. Here we describe important methods and findings in the field of multisensory integration, focusing on these two cues.

Previous research has explored how the brain integrates visual and haptic signals in normal grasping. Rock and Victor (1964) investigated shape and size perception when visual and haptic cues were available. Observers viewed an object through optical lens while exploring the object haptically. Visual shape and size were distorted by the lens and conflicting stimuli between vision and haptics were generated. Such cue-conflict experiments are used for examining each cue's contribution to the combined perception. Rock and Victor found that visual information strongly dominated the percept. This visual dominance, or "visual capture", phenomenon is well-known in multisensory processing, because vision has relatively better spatial resolution and is also more reliable than other modalities (Rock & Victor, 1964; Shams, Kamitani, & Shimojo, 2002; Spence & Driver, 2000; Woods & Recanzone, 2004).

Ernst and Banks (2002) investigated visual and haptic integration from a statistical perspective and re-examined Rock & Victor's study in light of the ideal observer (the MLE model) outlined earlier. They empirically demonstrated that the human brain integrates visual and haptic size information in a statistically optimal way. Before describing their results, it would be useful to briefly review the psychophysical methods used. A two-interval forced choice (2-IFC) task was employed to measure discrimination threshold, or just-noticeable differences (JNDs), for certain physical stimuli under varied conditions. In the Ernst and Banks study (2002), two parallel virtual planes were used for visual and haptic stimuli and these appeared at two intervals. Observers were required to judge the size between the two planes and to respond which interval contained the "larger" size. A psychometric function was produced by plotting the proportion of correct responses against sizes. Thus the obtained psychometric function represents the relationship between the certain physical parameters and the behavioural responses in consequence of the task decisions.

Figure 1.2 illustrates the relationship between the MLE model of cue integration and the behavioural responses. When using both visual and haptic cues to judge the interval in the 2-IFC task, if $\hat{S}_V = \hat{S}_H$, the chance level, or 50% correct response, would be positioned around physical stimuli $S_V = S_H$ in the psychometric function. Thus, the combined estimate $\hat{S}_{VH}$ would be $\hat{S}_{VH} = \hat{S}_V = \hat{S}_H$ in MLE model and the curve slope of the psychometric function would correspond to the degree of sensitivity (Figure 1.2 (a), (c)). If $\hat{S}_V \neq \hat{S}_H$, the chance level would be positioned between S_V and S_H, depending on each cue reliability. If haptic reliability is better than vision, for example, the combined estimate $\hat{S}_{VH}$ would shift toward $\hat{S}_H$ in the MLE model (Figure 1.2 (b), (d)).

Previous studies (e.g. Ernst & Banks, 2002; Knill & Saunders, 2003; Hillis et al., 2004), assumed that the standard deviation (σ) of the psychometric function is proportional to the standard deviation of the underlying size estimator. When JND is defined at the 84% level of the psychometric function, $\sigma = JND/\sqrt{2}$ (Ernst & Banks, 2002). Therefore, the JND is also proportional to the standard deviation of the underlying estimator. Thus, it is possible to estimate sigma for each cue by measuring

JND. Visual and haptic cue weights are relatively changed according to their reliabilities and the relationship with JND can be expressed using the following equation:

$$w_V/w_H = \sigma_H^2/\sigma_V^2 = T_H^2/T_V^2 \quad (1.9)$$

where T_H and T_V denote haptic and visual JNDs. Lower JNDs indicate greater precision in estimating the size and high reliability of the given cue. For example, when a visual cue has small variance ($\sigma_V^2 < \sigma_H^2$), vision gets more weight ($w_V > w_H$) and good discrimination performance ($T_V^2 < T_H^2$).

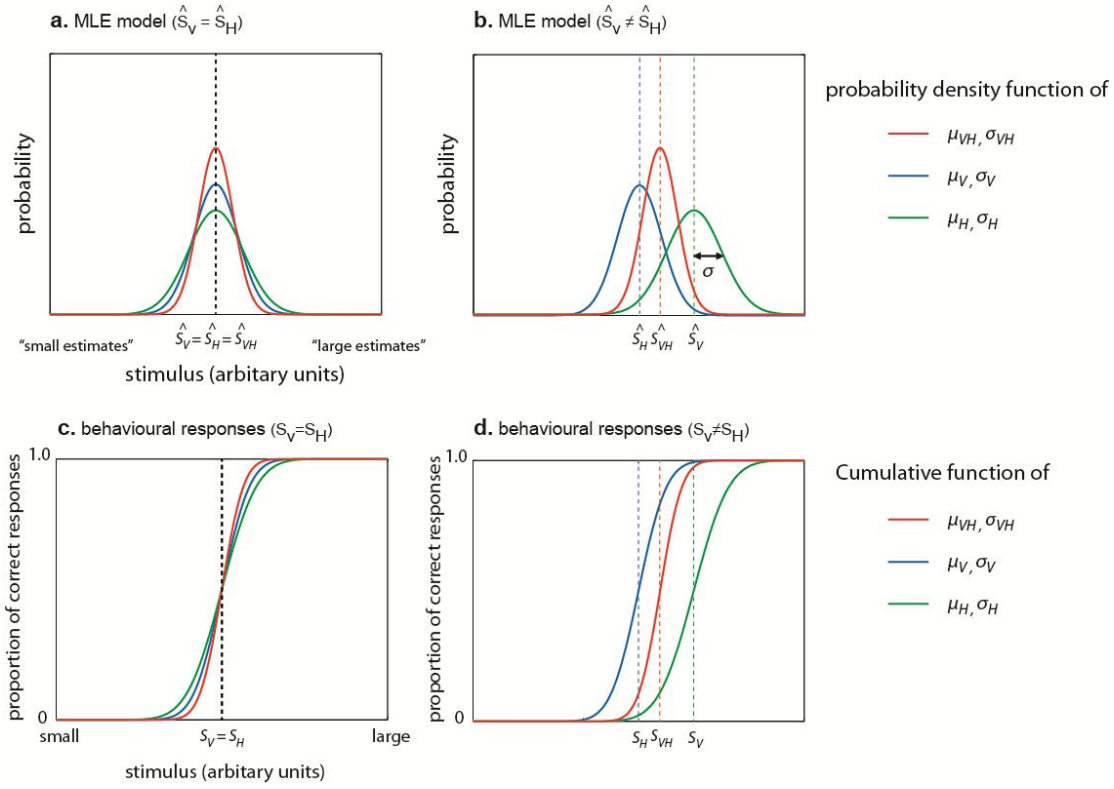


Figure 1.2. Illustrates the relationship between the MLE model of cue integration (top row **a.** and **b.**) and the behavioural responses (bottom row **c.** and **d.**). Generally, such psychometric functions can be fitted to a cumulative Gaussian using a maximum-likelihood criterion, and JND can be determined at the psychometric function. Both are mathematically ruled by parameters (mean and variance). Thus, σ of underlying estimate = σ of psychometric function divided by $\sqrt{2}$ (a and c). When physical stimuli are equal between two cues (vision and haptics), the mean of visual-plus-haptic performance does not change (b and d). When stimuli are conflicting, the mean of visual-plus-haptic estimate shifts toward the more reliable cue.

Manipulation of noise levels of each cue enables us to examine how the relative reliabilities of these cues change across different conditions and how human size perception was affected by cue reliabilities. That is, when each single-cue estimate and sensitivity could be observed, the MLE model predicts two-cue estimate using a linear cue combination of the individual cue estimates, weighted in relation to their reliabilities (see Equation (1.7)). The MLE model also predicts two-cue sensitivity changes (see Equation (1.8)) and this prediction can be used as the criterion of optimal cue integration. This method is referred to as cue-perturbation and is a powerful technique for exploring how estimates from the two senses are combined (Ernst & Banks, 2002; Knill, 1998; Young, Landy, & Maloney, 1993). As described above, one key advantage of the MLE model is that it does not consider prior knowledge and no free parameters are used to make predictions.

Ernst and Banks (2002) conducted two experiments to measure (i) single-cue to estimate weights for individuals and (ii) cue conflict to measure the actual cue weights the observers used. They measured size-discrimination thresholds (JNDs) of single-cue "vision-alone" and "haptics-alone" first. They added random displacement noise (uncertainty) to visual stimuli and varied the reliability of available visual information (binocular disparity) about the object size. Using single-cue data, they predicted bias and sensitivity when both cues were available. They then created cue conflict stimuli and attempted to manipulate cue weights by changing degree of visual noise level. They measured the Point-of-Subjective-Equality (PSE) of size estimates under the several cue conflict conditions. The PSE was determined at 50% threshold, where an observer perceived two stimuli to be the same. The observed two-cue performance was compared with the predictions of the MLE model.

Ernst and Banks (2002) showed that the observed cue weights were well predicted by the single cue reliabilities and estimates of object size were consistent with a weighted average of visual and haptics cues. They demonstrated the modality dominance phenomena (from haptic dominance to visual dominance) by changing the weights. "Visual capture" occurred when vision was highly reliable, whereas "haptic capture" occurred when vision was unreliable. Dominance altered between vision and haptics can be explained by the MLE cue combination rule. Therefore, the visual

dominance phenomenon observed by Rock and Victor (1964) can be considered as a special case, not the general case. In other words, the brain combines signals based on their relative reliabilities and the combined performance changes from single cues (modality dominance) to optimal depending on their weights.

Ernst and Banks' (2002) results also demonstrated that combined thresholds (vision-plus-haptics) were lower than the individual visual-alone or haptic-alone thresholds across the varied noise levels in size-discrimination performance. This means that the two-cue estimate was more reliable than the single-cue estimation. A reduction in the JND when both cues were available is evidence for visual-haptic integration. Other studies have confirmed that the combined estimate was better than either cue on its own (Gepshtein & Banks, 2003, Gepshtein et al., 2005). They showed that the two-cue reliability was near-optimal and this can be considered as the signature of cue integration. Furthermore, predicted combined performance, which was calculated from observed single-cue data using the MLE model, matched the empirical performance. In contrast to measuring JNDs, measuring PSEs often uses conflicting stimuli and analyses the perceived properties such as shape and sizes, assuming the visual and haptic estimates are on average unbiased. However, observers might have a conscious bias for the conflicting stimuli, and this possibility is unable to be eliminated completely.

1.2.3. Optimal cue integration with other senses

Several studies have tested the MLE model's ability to explain cue integration in a similar way to the Ernst and Banks' study. Statistically near-optimal performance has been reported across other sensory modalities. As shown in the previous section, visual dominance can be seen with other modalities and this can also be explained by the MLE model. The ventriloquist effect is well known as one example of the visual dominance phenomena, where visual information can capture auditory perception. It can be seen in a natural situation such as watching a puppet performance or films in a theatre. Humans perceive the sound to be coming from the visual location, even though the sound is actually from a different location.

Alais and Burr (2004) investigated the ventriloquist effect on direction judgement, changing the displacement between visual and auditory cues. They demonstrated that perceived direction was moved from the position of visual stimuli (light) toward the position of auditory stimuli (click) when visual stimuli was getting blurred. That is, the judged direction shifted from visual dominant to auditory dominant with increasing noise of the visual stimuli. The results also showed that single-modal thresholds (visual-alone and auditory-alone) were higher than a cross-modal threshold (audition-plus-vision). The observed data were well matched to the MLE prediction. Although visual dominance has often been reported between cross-modal interactions, other modalities can affect visual perception and even become dominant. Therefore these findings indicate that the MLE model to explain cross-modal integration across different sensory modalities.

Furthermore the MLE principle also provides a good quantitative account of cue combination between sensory inputs within modality; for example, depth perception from disparity and texture cues using a slant-discrimination task (Knill & Saunders, 2003; Hillis et al., 2004) and a location judgement task of texture edges (Landy & Kojima, 2001). As depicted above, when estimating a physical property in the world, the brain uses multiple sources of information through sensory systems; some are within sensory modality (e.g. texture and disparity in vision), and some are across (e.g. vision and haptics, vision and audition). Thus, the nervous system seems to exploit the same principle for combining cues across different sensory modalities and also within modality. Exploiting the redundancy of multiple noisy sensory signals is beneficial because it results in better performance than would be possible from either single cue (e.g. Ernst & Banks, 2002; Alais & Burr, 2004). These findings suggest that, statistically optimal integration is used to combine different cues with respect to their relative reliabilities.

However, previous studies have shown data that does not fit the MLE cue combination model and have claimed that a statistically optimal combination is one way to combine different cues but does not account for all cue combinations (Rosas, Wagemans, Ernst, & Wichman, 2005; Rosas, Wichmann, & Wagemans, 2007). Rosas et al. (2005) examined slant discrimination performance when using visual and haptic

texture information and showed that although each cue weight was sensitive to the individual reliability, the combined performance was not consistent with the statistically optimal rule. Moreover, Rosas et al. (2007) reported that when combining texture and motion cues in the slant discrimination task, motion cue weight was relatively consistent with the reliability but texture cue weight was not. These studies suggest that when two cues are strongly correlated, the linear weighted model would fail.

There are several models to explain multisensory integration when combining cues between and/or within modalities, such as weak fusion model, strong fusion model, and modified weak fusion (MWF) model (Landy et al., 1995). The weak fusion model allows linear interaction between cues, and the strong fusion model allows unconstrained nonlinear interaction between cues. The MWF model considers nonlinear interaction between cues, but the interaction is constrained in a similar manner to the weak fusion model. Although there are such variations in models, according to Hillis et al. (2002), when combining cues within the same sensory modality (e.g. disparity and texture in vision), sensory fusion occurred obligatory and the original information became lost in the final percept, suggesting non-linear interaction between cues. In contrast, when combining cues between modalities (e.g. visual and haptic size cues), the original information was relatively preserved in the final percept. Moreover, cue combination between modalities can be explained using MLE rule (Hillis, et al., 2002).

1.3. Correspondence problem in multisensory integration

1.3.1. Cue combination and correspondence problem

As described above, the key advantage of multi-sensory integration is that using combined cues to estimate properties of the physical world is more precise than any single cue (Ernst, 2007; Körding et al., 2007a; Roach, Heron, & McGraw, 2006). However, combining signals is not always optimal. If the brain simply combines different sensory information, the combined estimate might result in misleading

estimates. Consider a natural situation: when we look at a car outside a window while grasping a coffee cup in our hand, we should not confuse these two sizes. In contrast, when a person looks at a coffee cup while grasping it, the brain should combine size information from vision and haptics. In these cases, how does the brain decide whether or not to combine different sensory signals? In order to combine related sensory information appropriately and avoid combining unrelated information, the brain has to solve a general correspondence problem across sensory systems (and even within modalities).

How does the brain solve this correspondence problem? Clearly, it only makes sense for the brain to integrate sensory signals if and only if each cue provides information referring to the same object. If sensory signals are strongly related, it is highly probable that these signals would have shared the same origin. If a statistical strategy is employed when perceiving the object, the brain could decide which of the activations in the given sensory signals belong together or not. That is, the decision would be made, based on the probabilities of the correspondence.

1.3.2. Influence of spatial and temporal discrepancies on cue integration

In the real world, when different sensory signals share similar properties, in terms of spatial proximity, magnitude, and timing, it is likely that they originate from the same source (Ernst, 2005, 2007; Helbig & Ernst, 2007b). When there is a large spatial separation between two cues, it is unlikely that they originated from the same source, so the integration should not occur. Previous studies have shown that the correct response to the sound location decreased with increasing the spatial discrepancy between visual and auditory stimuli (Jackson, 1953; Witkin, Wapner, & Leventhal, 1952). Even though the task was judging a sound location, the performance was affected by the location of additive visual cue

More recently, Gepshtein et al. (2005) investigated the influence of spatial proximity on optimal cue integration, using a similar method to Ernst and Banks (2002). Gepshtein et al. (2005) measured size-discrimination thresholds and manipulated the spatial separation between visual and haptic stimuli. In visual-alone and haptic-alone conditions, single-modal performance showed no statistically significant differences in size-discrimination when stimuli were spatially dislocated

from the centre of the workspace. In contrast to the single-cue condition, size-discrimination performance was strongly affected by the spatial separation in the vision-plus-haptics condition. The visual-haptic performance worsened even when there was a small separation between two stimuli of just a few centimetres. The combination of vision and touch depended on the degree of their coincidence in space. That is, visual and haptic size estimates are integrated in a statistically optimal fashion only if the two signals come from the same location, whereas the amount of integration reduces with increasing spatial offset between the two sensory signals. This shows that the brain is sensitive to whether signals were caused by the same object.

Such a spatial correspondence problem can be seen in other modalities, including auditory-visual spatial localisation judgements (e.g. Körding et al., 2007a). Körding et al. (2007a) showed that with an increasing spatial separation between visual and auditory signals, auditory weight on response performance decreased. Observers were directly asked their perception of unity (whether auditory and visual signals were from the same source or different sources) on each trial. When the spatial separation between cues increased, the reports of unity perception decreased. In addition, research of a correspondence problem has examined human response behaviour in visual-tactile event judgement (e.g. Bresciani, Dammeier, & Ernst, 2005; Ley, Haggard, & Yarrow, 2009). Bresciani et al. (2006) showed that visual-tactile integration occurred in a task which asked observers to count either the number of flashes or taps. The integration only occurred when visual and tactile stimuli were presented simultaneously, even though one signal was irrelevant to the task. Moreover, speech perception effects such as the McGurk effect, where voice recognition interacts with visual information (shape of a lip), has been investigated the temporal aspects of multisensory integration. The research reported that auditory-visual integration only occurred within a certain temporal duration, or window, between two sensory stimuli (e.g. van Wassenhove, Grant, & Poeppel, 2007). These findings can be considered as evidence that the brain deals with statistical regularities (both spatially and temporally) in multisensory integration (Ernst, 2007). As such the correspondence problem is a general problem.

1.3.3. Bayesian inference and solving the correspondence problem

For multisensory cue integration to be effective, the brain needs to solve such a general correspondence problem across sensory modalities. Consider the case in which the brain should sometimes combine signals from different sensory modalities and should not in other cases. Here, using Bayesian inference, which pertains to the prior knowledge about the statistical world, as the model is the key concept to solve the correspondence problem. Several related models have successfully characterised multisensory integration using a Bayesian framework, introducing ideas of a coupling prior (Ernst, 2007) and causal inference (Körding et al., 2007a; Sato, Toyoizumi, & Aihara, 2007). These ideas incorporate knowledge of the statistics of the world and provide a good account for multisensory cue combination and solving the correspondence problem (e.g. Battaglia, Di Luca, Ernst, Schrater, Machulla, & Kersten, 2010; Bresciani et al., 2006). Both suggest that the correspondence problem could be solved by considering the statistical similarity between sensory signals in terms of coincidence across time, space and magnitude (Deneve & Pouget, 2004; Ernst, 2005, 2007; Körding et al., 2007a). That is, the prior probability provides secondary, or "auxiliary" information, and constrains the interaction between sensory inputs; this contributes to generate a certain perceptual interpretation (Battaglia et al., 2010).

To solve the correspondence problem in multisensory processing, these models consider an interaction between the joint likelihood function of different sensory stimuli and a coupling prior (Bresciani et al., 2006; Ernst, 2005; Roach et al., 2006). Here, the coupling prior refers to the joint distribution of the perceptual knowledge about the stimulus' statistical relationship between sensory cues. That is, it describes the prior expectation of how tightly the stimulus properties are coupled from different sensory modalities. For example, the brain may know that a certain visual size estimate of an object reliably covaries with a certain state of muscles and golgi-tendon organs that indicate a certain grip size. In which case the coupling prior would reflect the fact that these visual and haptic size estimates are likely to occur together, and all other combinations are unlikely. By combining the likelihoods from the sensory estimates with this prior, according to Bayes' rule, one can estimate the

probability that the two sensory estimates refer to the same object. If the sensory estimates are likely to occur together, the joint estimate of a common source will have a high probability, and if they differ substantially the joint estimate will have a low probability; it is more probable that the signals came from separate objects (Ernst, 2005, 2007; for different formulations of this approach see Knill, 2007a; Körding et al., 2007a; Roach et al., 2006; Sato et al., 2007; Shams & Beierholm, 2010; Shams, Ma, & Beierholm, 2005). This Bayesian approach to solve the correspondence problem will be described in more depth in Discussion of Chapter 3.

1.3.4. The correspondence problem when using a tool

As shown in Gepshtein et al.'s study (2005), visual and haptic signals were optimally integrated only when the estimates of hand and object positions coincided, and the integration was reduced with an increase in the spatial separation of the cues. Thus, an invariant spatial rule could be used, when grasping an object with the hand. However, there are common situations in which this rule for determining correspondence between visual and haptic signals would fail. Tool use is one example. Humans are adept at using different tools in daily life and tools manipulate the spatial relationship between felt and "seen" stimuli. For example, when using a hammer, we can hit an object that is spatially separated from the hand location. When using barbeque tongs, we can flexibly pick up different sizes of meats located far from the hand. Even though the invariant spatial rule cannot apply to these cases, intuition suggests signals can be combined. In this thesis we investigate whether this is the case.

Consider normal grasping and using a tool such as a pair of pliers (Figure 1.3). When an object is grasped with this tool, its size can be estimated from the available visual cues as well as from the "felt" opening of the hand. Visual information about the object is available at the tool-tip and haptic information from the hand opening. Does the brain understand the spatial remapping that has occurred when using the tool? And does it still combine signals appropriately?

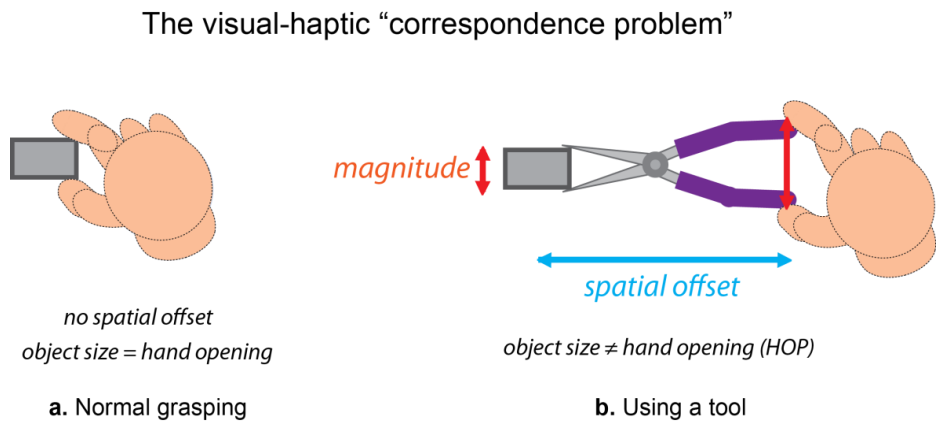


Figure 1.3. This cartoon illustrates the visual-haptic correspondence problem, in comparison to **a.** normal grasping and **b.** using a tool, when estimating an object size. In both cases, haptic size is felt at finger and thumb. When grasping normally (a), there is no spatial offset between visual and haptic signals. The (seen) object size and the (felt) size at hand are matched. In contrast, when using a pair of pliers (b), the (seen) object and the (felt) object are spatially separated via the tool length. The (seen) size and (felt) size are different via the magnitude between the tool-tip and the handle.

In normal grasping, the extent to which haptic and visual signals are spatially coincident is directly related to the likelihood that they refer to the same object (Figure 1.3 (a)). However, a pair of pliers changes this normal spatial relationship between the tool-tip and the handle (Figure 1.3 (b)). In other words, tools can systematically perturb the spatial offset and magnitude between object and hand. In this manner, tools complicate the simple correspondence rule. Even though two raw sensory signals are spatially separated or conflict in size, because the two signals are originating from the same object in tool use, a statistically optimal tool user would integrate these signals. Does the brain understand this spatial and size remapping when using a tool? How does the brain deal with sensory information under such size remapping in tool use?

There are several related studies that investigated these types of conflicts in the magnitude of sensory signals. Recently, Girshick and Banks (2009) investigated slant discrimination performance with conflicts between slants specified by binocular disparity and texture cues. They reported that the weighted averaging model was

preserved with small conflicts between two cues. However, when large conflicts were presented, the brain seemed to robustly estimate the slant, which was deviated from the weighted averaging. These results are consistent with coupling prior model (Bresciani et al., 2006; Ernst, 2005; Roach et al., 2006) and causal inference model (Körding et al., 2007a; Sato et al., 2007). Even though the estimate with large conflict was provided by only one cue, observers were not aware of single-cue estimates. Therefore, two cues could be fused in the robust percept regardless of cue conflicts situations.

Moreover, the natural and intuitive manner of tool use suggests that the brain employs a more sophisticated solution to the correspondence problem during cross-modal integration. In Chapter 3, we will examine human performance during tool use and discuss whether tool use can be explained within a Bayesian framework.

1.4. Human tool-use research

Tool use has a special place in human history and is arguably of fundamental evolutionary importance. It greatly has widened human capabilities. Tools expand our reachable space and provide effective manipulation, interacting with objects in ways not possible with the hand alone. Informally, many of us have experienced the sensation that, when we pick up a tool, we can "feel" the tip of the tool, even though the haptic signals originate at the hand. Tool use might change our sense of where our bodies are in space. Here, we review tool use research, mainly focusing on studies in relation to spatial remapping. Other studies are discussed in later chapters; for example, neural mechanisms of multisensory processing in tool use, visuo-motor control (e.g. Ingram, Howard, Flanagan, & Wolpert, 2010; Scott & Gray, 2010) and transformation rules (e.g. Beisert, Massen, & Prinz, 2010; Massen & Prinz, 2007).

1.4.1. Dynamic changes in peripersonal space and body schema with tool

Previous studies have reported that active tool use dynamically affects body-related spatial encoding systems – so-called "peri-personal space" or "body schema" – which assimilate the body space and the space surrounding the body (for a review see Maravita & Iriki, 2004; see also Farnè & Làdavas, 2000; Maravita, Spence, & Driver, 2003). Different approaches have been used to investigate the brain mechanism of spatial representation interacting with tool use: neurophysiological (e.g. Iriki, Tanaka, & Iwamura, 1996), neuropsychological (e.g. Berti & Frassinetti, 2000; Bonifazi, Farnè, Rinaldesi, & Làdavas, 2007; Farnè, Iriki, & Làdavas, 2005b; Farnè & Làdavas, 2000), and behavioural (e.g. Holmes, Calvert, & Spence, 2004b; Massen & Prinz, 2007) approaches.

Neurophysiological research: changes in receptive fields with tool use

Bimodal neurons, which respond to both visual and somatosensory signals, play a vital role in shaping representation of peri-personal space, within visual processing (e.g. Fogassi, Gallese, Fadiga, Luppino, Matelli, & Rizzolati, 1996; Iriki et al., 1996; Obayashi, Suhara, Kawabe, Okauchi, Maeda, Akine, Onoe, & Iriki, 2001). In a widely cited study, Iriki and colleagues (1996) found that when a monkey used a tool, the bimodal receptive fields of cells responding to locations around the hand enlarged to include the region surrounding the tool. A monkey collected a food pellet using a hand or a rake, and neural activities were measured before and after active tool use, using single unit recording techniques measurements of hand trajectories and eye movements were also made. The monkey's bimodal neurons were selected from in the anterior inferior premotor cortex (IPS), where somatosensory and visual inputs converged. Immediately after tool use, bimodal neurons' activities increased and the visual receptive fields (RF) enlarged in accordance with the expanded accessible space. Interestingly, while the somatosensory RF was unchanged, the visual RF was enlarged mostly along the tool axis. Therefore tool use may affect or modify spatial representation mainly through visual processing. The findings suggested that the neural activities may be specialised for, or adapted by tool use. The results provide

evidence that tool use dynamically changed neural coding of peripersonal space. These neurophysiological studies provided important insights into the basic nature of visuo-somatosensory integration in tool use.

Neuropsychological research: cross-modal extinction

Dynamic effects of tool use on humans' peripersonal space have also been shown in neuropsychological research, mainly investigating changes in cross-modal extinction before and after tool use (e.g., Berti & Frassinetti, 2000; Bonifazi et al., 2007; Farnè, Bonifazi, & Làdavas, 2005a; Farnè et al., 2005b; Farnè & Làdavas, 2000; Farnè, Serino, & Làdavas, 2007; Maravita, Clarke, Husain, & Driver, 2002a; Maravita, Husain, Clarke, & Driver, 2001). Cross-modal extinction can be seen in right hemisphere brain-damaged patients, who fail to report tactile stimuli on the left (contralesional) hand when competing visual stimuli are presented near the right (ipsilesional) hand. Studies have shown that before tool use, the extinction was severe at near hand space, but after tool use the extinction was more severe at the far space (reachable by the tool) than near hand space. This pattern of spatial interaction with tools was also shown in normal healthy humans by measuring tactile judgement with a visual distractor (Maravita, Spence, Kennett, & Driver, 2002b).

A series of experiments of visual-tactile extinction measured patients' left contralesional tactile detection under different conditions manipulating the spatial locations of bilateral cross-modal stimuli after different types of tool use (Farnè et al., 2005a, 2005b). The research showed that the visual-tactile extinction was affected by only actual tool use and was not affected by passively observed tool use. The extinction was also affected by the tool's practical and functional length rather than the total length. Further experiments examined cross-modal extinction at different locations along the tool axis and showed that after the tool use the extinction occurred not only at the tool-tip but also at the middle of the tool (Farnè et al., 2007; Bonifazi et al., 2007). They claimed that the peri-hand space was continuously enlarged from the hand space towards the tool-tip. A recent study explored the relationship between peri-personal space and tool use by combining clinical and computational approaches (Magosso, Ursino, di Pellegrino, Làdavas, & Serino, 2010). Based on an idea of

neural network, Magosso et al. proposed a model that the visual receptive field (RF) of bimodal neurons enlarged along a tool. The model predicted the increase of behavioural responses to the place where the tool-tip was located before. The model also provided a prediction of behavioural changes in visual-tactile extinction before and after tool use. By merging the neurophysiological and neuropsychological findings, a series of cross-modal extinction studies claimed that neural representation of perceptive space could be remapped via active tool use.

Behavioural research: spatial attention and spatial encoding

Spatial attention involves the construct of spatial representation in cross-modal processing (Driver & Spence, 1998, 1999) and studies of visual spatial attention in tool use have suggested that tools are incorporated into the body schema (Holmes & Spence, 2004a; Maravita & Iriki, 2004). Evidence for remapping has also been shown in visual-tactile congruency effect, in which reaction time was faster and fewer discrimination errors occurred when vision and tactile locations were matched. Holmes and colleagues (2004b) examined the congruency effects between tactile stimuli at hand and visual stimuli at three different locations of a handheld stick (at the handle, at the middle and at the tip) in normal healthy controls. They showed that when visual stimuli were at the tool-tip and tool-end, response times were similar and faster than that at the middle of the tool. The response time to the stimuli showed that the cross-modal congruency effect seemed to shift, or project, from hand position to the tip of the tool, rather than expanding from the hand position. Recently, Holmes, Calvert and Spence (2007a) investigated spatial and temporal effects of tool use using the same cross-modal congruency paradigm and compared the effects between predictable and unpredictable sequences of tool use. They demonstrated that preparing for tool use seemed to enhance multisensory integration between visual and tactile modalities and the spatial interactions could be seen only near the hand and at the tool-tip. This may suggest that some knowledge about the tool at the pre-attentive level could be involved in integration.

Several studies have argued the relationship between active tool use and the peripersonal space, compared with behavioural responses and neurophysiological

evidence of bimodal receptive fields. There are two intuitive hypotheses concerning how tool use computes the space: (i) shift of the peripersonal space from the hand position to the tool-tip, or (ii) continuous enlargement of the peripersonal space (Bonifazi et al., 2007; Holmes et al., 2004b); in other words, "body-extension" or "body-incorporation" (De Preester & Tsakiris, 2009). The former concept is that even though spatial attention shifts at the tool-tip and the receptive field of sensory neurons enlarge toward the tool-tip; the sense of body is persistent. While the latter concept is that a tool becomes a part of the body and the sense of body can be changed.

A brain imaging study of healthy participants showed that specific regions (e.g. right hemisphere lingual gyrus) were enhanced during tactile stimulation in tool use only when the visual stimuli were matched at the tool-tip (Holmes, Spence, Hansen, Mackay, & Calvert, 2008). Holmes and his colleagues suggested that shifting spatial attention via visual stimuli to the area where a tool was acting assimilated the tool-tip and the hand space. Through a series of attention studies, they claimed that spatial attention seems to play an important role in multisensory integration in tool use and supported the shift idea of peripersonal space (e.g. Holmes et al., 2004b; Holmes, Sanabria, Calvert, & Spence, 2007b). These studies have explored the spatial encoding system by arguing the question: whether active tool use shift the peripersonal space to the acting tool-tip or enlarges the peripersonal space along the tool. These approaches, combining brain imaging and behavioural experiments, have contributed to reveal the neural mechanisms underlying tool use.

Mounting evidence from tool use research shows that active tool use affects human perception and spatial remapping seems to occur when using a tool. However, the neural mechanisms underlying tool use are not yet fully understood. To our knowledge, previous research has not explored these mechanisms from a statistical framework, in terms of the correspondence problem. As described before, if the brain treats information statistically even when using a tool, the phenomena of human tool-use performance might be ruled by more general principles and would involve similar neural processes of solving the correspondence problems.

1.4.2. Virtual tools

Overall, tool research is important not only to understand the underlying neural mechanism of human tool use but also to apply this knowledge to practical usage. Modern technology has been developed to enable us to widen human reachable space, in particular interacting with the computer-simulated virtual environments. Virtual reality (VR) and tele-presence systems have become more prevalent in contemporary society, including robotic surgical devices and tele-rehabilitation (e.g. Bello, Coles, Gould, Hughes, John, Vidal, & Watt, 2010; Murescaux, Leroy, Gagner, Rubino, Mutter, Vix, Butner, & Smith, 2001; Popescu, Burdea, Bouzit, & Hentz, 2000; Robles-De-La-Torre, 2006). Moreover, the recent technologies have enabled doctors to connect veins of a patient at a different hospital via a tele-surgical operation. These systems introduce a new spatial mapping, such as minifying/magnifying the normal scale and sharing the dimensions even over long distance. The systems force hand movements to be rescaled; for example, a surgeon may move a small amount with his hand in order to manipulate a much larger or smaller surgical instrument located in another part of the room. That is, information from touch and the position of the hand-held tool (haptic information) must be integrated with information from vision about the position of the haptic device itself, even when the spatial origins of the haptic and visual information are different. At the same time, the systems are required to achieve natural senses or a "sense of reality" for controlling an object precisely in the virtual environment. This would also help to minimise risks to the patient by treating the body delicately and safely by a surgeon in the above medical situations.

For human hands to manipulate such haptic devices effectively and precisely, the brain must solve the correspondence problem. Understanding the mechanism of the visual-haptic correspondence problem and the integration in a computer-simulated new spatial mapping has become progressively important. We will discuss how the correspondence problem is a fundamental issue in multisensory integration. To learn how the brain deals with the correspondence problem when using such modern tools, and also to know when it cannot solve it, would be useful for designing interfaces.

1.5. Thesis outline

This thesis explores human performance when using tools. Although sensory integration and tool use have been examined from different aspects as described earlier, little is known about cue integration and the spatial correspondence problem in tool use. We will examine these underlying mechanisms in human tool use, specifically whether visual-haptic integration occurs during tool use. We employ techniques from sensory integration literature, and statistical modelling. We have used stereoscopic 3-D displays and (PHANToM) force-feedback devices in order to manipulate visual and haptic stimuli independently in the virtual environment. These techniques allow us an insight into the fundamental questions of how the brain integrates multisensory information and solves the correspondence problem during tool use.

The reminder of this thesis consists of four empirical chapters and a general discussion and conclusion as a whole. Chapter 2 explores whether visual-haptic integration occurs in tool use, and whether the brain appropriately solves the correspondence problem when tools introduce a spatial offset. Chapter 3 examines whether the brain correctly solves the correspondence problem when tools change the relationship between magnitude of felt size and visual object size, and outlines a modelling of tool use. Chapter 4 explores a key question raised by the modelling: whether haptic size estimates are rescaled to account for tool geometry, to allow accurate size estimates. Finally, Chapter 5 explores whether cue weights are altered optimally, when haptic reliability is changed by using a tool.

Chapter Two

Effects of spatial offset on visual-haptic integration in tool use[1]

2.1. Introduction

Here we explore whether the brain appropriately solves the correspondence problem during tool use. To make the problem simple, consider using a simple tool consisting of two sticks attached rigidly to the thumb and index finger (Figure 2.1). When an object is grasped with this tool its size can be estimated from the available visual cues, as well as from the "felt" opening of the hand. As introduced in Chapter 1, a statistically optimal tool user would integrate these signals, because they relate to the same object.

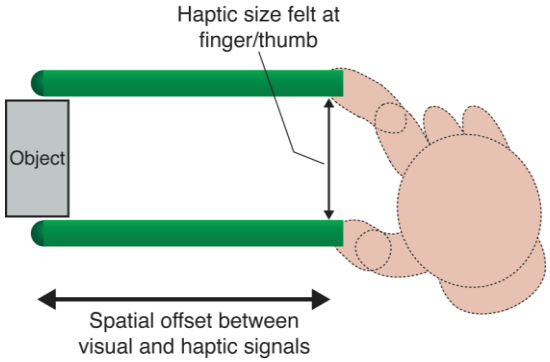


Figure 2.1. A cartoon illustrating the spatial offset between visual and haptic signals to object size when using a simple tool, comprised of two sticks rigidly attached to the finger and thumb, respectively. We used a "virtual" tool of this kind in our experiments.

[1] The experiments in this Chapter have been published in: Takahashi, Diedrichsen, & Watt. (2009). Integration of vision and haptics during tool use. Journal of Vision, 9 (6): 3, 1-13.

Yet the tool systematically perturbs the 'normal' spatial mapping between visual and haptic signals – in this case they are spatially offset – and a spatial proximity rule may cause them not be combined.

To investigate this research question, we examined visual-haptic size discrimination under a variety of spatial offset and tool-use conditions. We first replicated Gepshtein et al.'s (2005) study to confirm that, when grasping with the finger and thumb, cross-modal integration is optimal when the signals originate for the same location, but reduces with increasing spatial offset between visual and haptic signals. To do this, we measured single-modality (vision-alone and haptics-alone) discrimination performance and used this to predict optimal performance when both cues were available (Experiment 1: normal grasping). We then determined whether cross-modal integration could be restored despite a spatial offset between visual and haptic signals, when using a simple tool, the length of which varied with spatial offset such that it always reached the visual object (Experiment 2: zero tool-object offset). Finally, we examined whether spatial offset between the tool-tip and the object resulted in similar reductions in cross-modal integration as when the hand is offset in natural grasping (Experiment 3: variable tool-object offset). Such pattern of results would suggest that during tool use the haptic signal is treated as originating from the tool-tip.

2.2. Experiment 1: normal grasping

Experiment 1 examined the influence of spatial offset on visual-haptic integration in normal grasping. We measured size-discrimination performance for single cues (visual-alone and haptic-alone) first, and then used this to predict optimal performance when both cues were available, using the MLE model. Then, we measured two-cue performance by manipulating the spatial offsets between visual and haptic signals. In this vision-plus-haptic condition, observers could see and feel the object at the same time.

2.2.1. Methods

Observers

Seven right-handed observers (aged 21-44 years) participated in this experiment. All observers had normal or corrected to normal vision, normal stereo-acuity, and no known motor deficits. Six of the observers were naïve to the purpose of the experiment and one (CT, the author) was aware of the purpose.

Apparatus

Figure 2.2 shows the apparatus. Observers viewed 3-D visual stimuli in a conventional Wheatstone mirror stereoscope, consisting of a separate TFT monitor (refresh rate 60 Hz) and mirror for each eye. Observers were positioned in front of the mirror so that its centre came to their body centre. Their head position was stabilised using chin and forehead rests. The haptic stimuli were generated by two PHANToM 3.0 force-feedback devices (SenseAble Technologies, Inc.), one for the index finger and another for the thumb of the observer's right hand. Observers could not see their own hand because it was occluded by the mirror. Touching the virtual object with the finger or thumb resulted in an appropriate opposing force. The 3-D positions of the tips of the finger and thumb were monitored by the PHANToM to simulate the presence of haptic surfaces in the 3-D space. Here, the visual and haptic "workspaces" are set as coincident.

Note that all of our experiments in this thesis were designed using PHANToM devices, which introduced a computer-simulated unnatural mapping. We also generated a small visual sphere representing a finger position and a finger sled connecting with a robot arm in the virtual environment; therefore, the sets of experiments can be considered as using a "type of tool", regardless of visually defined no-tool or tool representation.

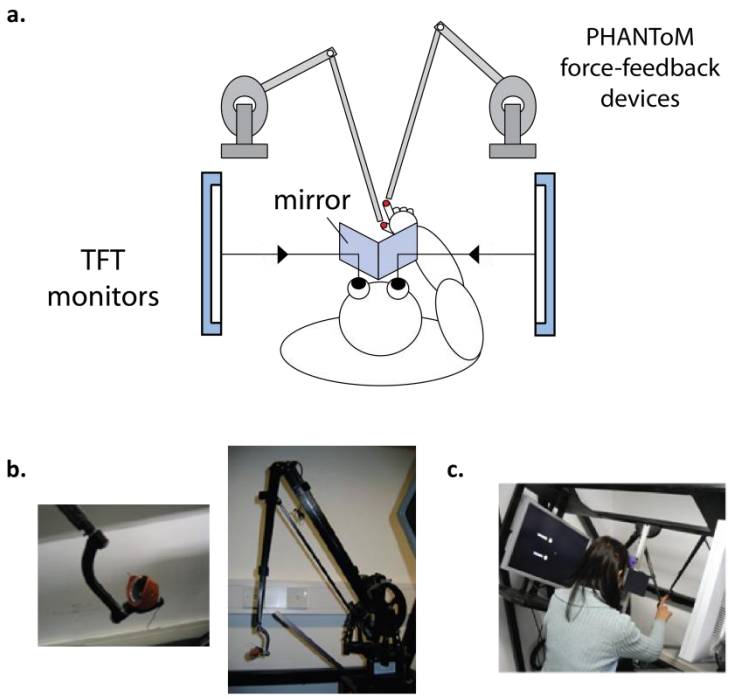


Figure 2.2. The apparatus used to present visual and haptic stimuli. **a.** A schematic diagram of the apparatus settings. An observer viewed the 3-D virtual stimuli in a mirror, where two images on the left and right side monitor screens were fused. The observer felt the virtual object with finger and thumb attached to the PHANToM force-feedback device respectively. **b.** Photographs: one of the PHANToM robot's arm (right), and a rubber sled which a finger is put into (left). **c.** An observer in the experiment setup.

Stimuli

Following Gepshtein et al. (2005), the visual and haptic stimuli were two parallel rectangular planes (Figure 2.3).

The visual stimuli were random-dot stereograms depicting two transparent parallel planes, separated along their surface normals (c.f. Gepshtein & Banks, 2003). Each plane consisted of uniformly distributed rectangular 'dots' with an average width and height of 2mm (Figure 2.3 (a)). A random jitter was added to the width and height of each dot (±1 mm, uniform distribution) to disrupt the use of dot size as a cue to the planes' separation. The dots covered approximately 8% of the total surface area

of each plane. A new stereogram was generated for each plane on each stimulus presentation. The average width and height of each plane was 50mm. A random variation (±10 mm, uniform distribution) was added to the height and width of each plane, so that the degree of visible overlap between the two planes was not a reliable cue to their separation. The distance from the cyclopean eye to the (visual and haptic) stimulus, in the mid-sagittal plane, was varied randomly in the range 460 – 530 mm so that the distance to one plane could not be used to judge the planes' separation. We used anti-aliasing to achieve sub-pixel accuracy of dot positions.

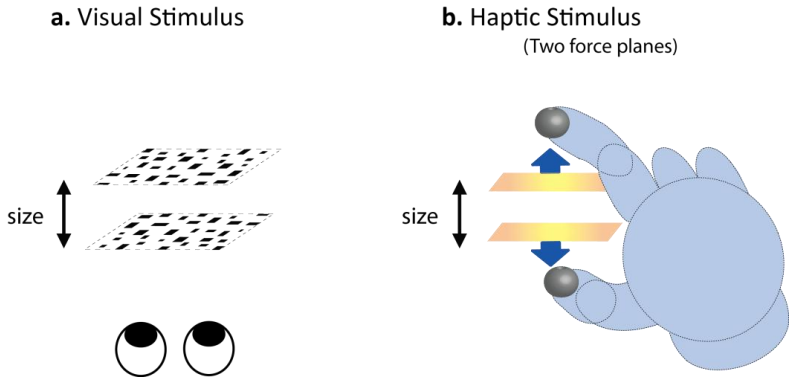


Figure 2.3. Illustrations of the stimuli used. **a.** Visual stimulus. Two transparent parallel planes were separated with the certain distance (size). Each plane was generated by random dots stereogram. **b.** Haptic stimulus. Two parallel force planes were also separated with the certain distance (size). Blue arrows represent the opposing force when touching the virtual surfaces with the finger and thumb.

The haptic stimuli were also two rectangular planes, the dimensions and position of which matched those of the visual planes save for a variable horizontal spatial offset (Figure 2.3 (b)). The haptic stimuli were rendered by attaching the finger and thumb of the observer's right hand to the PHANToM force-feedback devices. Touching the virtual surfaces with the finger or thumb resulted in an opposing force generated, simulating contact with a real plane. The rigidness of the

surfaces was 0.8 [N/mm] in all conditions; so, observers felt a slightly spongy 3-D object. In addition, two visual spheres indicated the positions of finger and thumb (Figure 2.3 (b)), and whether they were visible or not was controlled by the computer program.

Procedure

In all conditions, we measured size-discrimination performance using a two-interval forced-choice (2-IFC) procedure. Visual and/or haptic size (defined as the separation between the planes) was varied according to a method of constant stimuli, and observers indicated which interval contained the larger size. The standard size was 50mm, and the eight comparison sizes were ±1, 3, 6, 9 mm from the standard size. The presentation order of standard and comparison stimuli was randomised and each interval the stimuli were presented for 1 sec. In each condition observers completed 30 repetitions of each stimulus level. No feedback was given at any stage during the experiment.

Single-cue experiment (vision-alone/haptics-alone condition)

We measured size JNDs in the separation of the two planes for vision- and haptics- alone in order to predict performance when both cues were available. According to Equation (1.6), the maximum improvement in discrimination performance that results from cross-modal integration occurs when the two signals have equal variance. Therefore it is important to experimentally match the variance of estimates from each cue, for each observer, so that the effects of cue integration are most clearly evident (Gepshtein et al., 2005).

In both single-cue conditions, we varied the orientation of the stimulus to alter cue reliability for individual observers. We measured size JNDs as a function of stimulus orientation (Figure 2.4). Varying the orientation of the stimulus relative to the line of sight has a large effect on the variance of size estimates from vision, but leaves estimates from haptics relatively unaffected (Gepshtein & Banks, 2003). We

set the orientation range of stimuli individually according to the observer's visual and haptic performance. For example, we measured size JNDs for the observer 'BK' at stimulus orientations (40, 50, 60, 70, and 80 deg) for vision, and because haptic size JNDs vary little with orientation, we used fewer stimulus orientations (40, 60, and 80 deg). Vision-alone and haptic-alone blocks were completed separately.

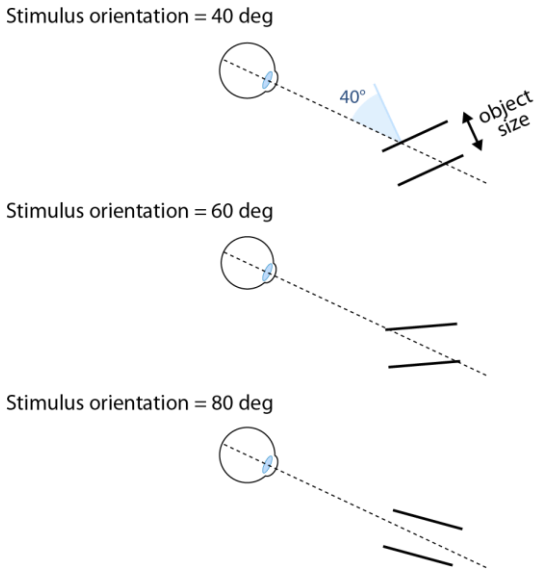


Figure 2.4. Stimulus orientation was defined as the surface slant (relative to the line of sight). As stimulus orientation approaches 0 deg (fronto-parallel) visual performance is increasingly dependent on a depth estimate from binocular disparity, and thresholds increase (Gepshtein & Banks, 2003).

In the vision-alone conditions, a fixation-cross appeared at the beginning of each trial, indicating the position and orientation of the upcoming stimulus. The two intervals were then presented for 1 sec each, separated by a 1.6 sec inter-stimulus interval (this time was determined based on the typical inter-stimulus interval when

haptic information was available, see below). Observers then indicated which interval contained the larger size by pressing one of two virtual (visual and haptic) buttons.

In the haptics-alone condition, the observers grasped the stimulus with the index finger and thumb. Two visual spheres appeared which indicated the position and orientation (but not any relevant size information) of the upcoming stimulus. When observers inserted their finger and thumb these 'start zones' changed colour from yellow to green indicating that they should begin to grasp the stimulus. The start zones could not be felt. All visual information disappeared immediately when the observers moved their digits inward from the start zone positions. Observers were trained to grasp the stimulus for ~1 sec in each interval and then release it. If the time both digits touched the surface was less than 1000 msec, or more than 1500 msec, the message "too fast" or "too slow" appeared on the screen, and that trial was discarded. In practice this was a small proportion of trials. This process was repeated for the second interval, and the observer then judged which interval contained the larger size, as described above.

Two-cue experiment (vision-plus-haptics condition)

In the two-cue condition, visual and haptic stimuli were presented together. The procedure was similar to the haptics-alone condition, with the following exceptions. First, the visual stimulus was presented only when both haptic planes were touched simultaneously, to match presentation time in each modality. Second, the visual and haptic stimuli were presented with a variety of horizontal spatial offsets between the finger/thumb and the visual object (see Figure 2.5). We refer to this spatial offset as the *hand-object offset*. Visual and haptic stimuli were always offset by an equal and opposite amount on either side of the body midline. Hand-object offsets of 0, ±50 and ±100 mm were presented. A positive offset means the hand position was to the right of the visual object. The offset was chosen randomly on each trial, because a constant offset could lead to visuo-motor adaptation.

The visual start zones were shown on all stimulus intervals and were offset to indicate where the observer should grasp the haptic object. A fixation-cross was

presented before each interval, at the location of the visual object. The fixation-cross and feedback about finger position disappeared before the visual-haptic stimuli were presented. They were extinguished when the finger and thumb were moved from the start zones, toward the haptic object. During stimulus presentation only the random-dot planes were visible.

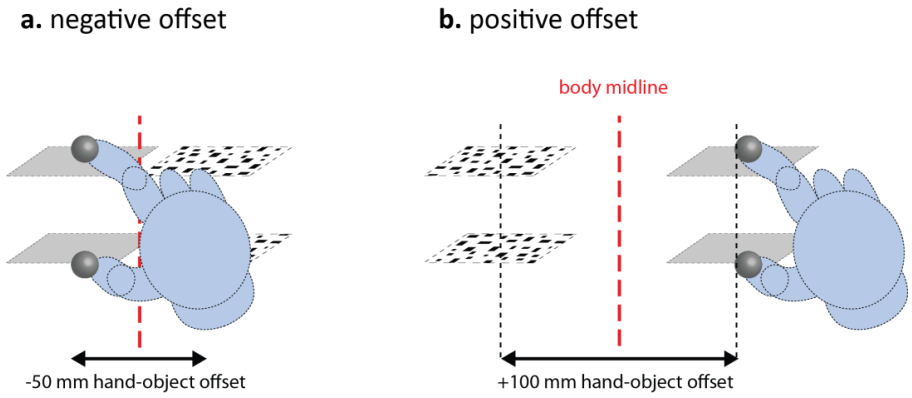


Figure 2.5. A schematic of the stimuli. The textured rectangles show the visual object, defined by two random-dot-stereogram planes. The grey rectangles show the haptic stimulus (not visible to observers). The grey spheres indicated the position of the finger and thumb. The hand was not visible. We could not simulate rotation forces on the digits because the experimental design required that visual and haptic stimuli were exactly matched across tool and no-tool experiments. **a.** negative and **b.** positive spatial offset between visual and haptic stimuli, respectively (referred to as hand-object offset). The finger-position spheres were extinguished before presentation of the haptic and visual objects (see Procedure).

Observers completed Experiments 1 and 2 (see next section) during the same time period. Trials were blocked by Experiment. The hand-object offsets were randomised within each experiment.

2.2.2. Data Analysis

JNDs were defined as the sigma parameters for the best-fitting cumulative Gaussian to the psychometric data divided by $\sqrt{2}$, using a maximum-likelihood criterion. As shown in previous studies (e.g. Ernst & Banks, 2002; Hillis et al., 2004; Knill & Saunders, 2003), JNDs are proportional to the standard deviation (σ) of the underlying size estimators.

We used several customised Matlab scripts/functions [2] for analysing behavioural data. The best-fitting parameters and estimates of the Gaussian (mean μ, standard deviation σ (sigma), standard error of mean (SEM), and standard error of sigma) were obtained by maximising the matrix of second partial derivatives of the log-likelihood. We also used "psignifit" which has been widely used to fit psychometric functions and to test hypotheses using a bootstrap method (Wichmann & Hill, 2001a, 2001b).

2.2.3. Results

Single-cue performance: Matching the variance of visual and haptic estimates

Figure 2.6 plots size-discrimination performance for visual and haptic modalities as a function of stimulus orientation for one example observer (BK). Similar to Gepshtein et al.'s (2005) results, JNDs for vision depended strongly on object orientation. In contrast, JNDs for haptics were insensitive to stimulus orientation. These patterns were similar for each observer. (See Appendix 1 for individual data). As noted, the observed trends for haptics were also consistent with previous data using real objects of different orientations (e.g. Lawson, 2009).

[2] Those scripts were provided by Michael S. Landy.

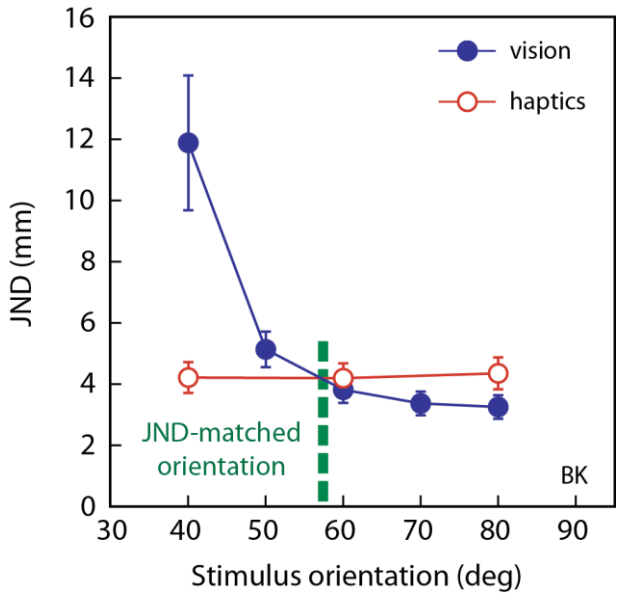


Figure 2.6. Discrimination performance (JND) of one example observer in vision- (blue closed circles) and haptic-alone (red open circles) conditions plotted as a function of stimulus orientation. The dashed line denotes the orientation that provided a close match between the precision of size estimates from the two cues. Each observer completed the cross-modal conditions using his or her 'JND-matched' stimulus orientation. Error bars denote ± 1SEM.

We could find a matching point in Figure 2.6, where JNDs and therefore variances of visual and haptic estimates, or their cue reliabilities, were equal for each individual. We determined the stimulus orientation at which visual and haptic JNDs were approximately equal. These vision- and haptics-alone JNDs were used to compute the improvement in discrimination performance (reduction in JND) under statistically optimal cross-modal integration (Equation (1.4)). Each observer completed all of the cross-modal experiments using his or her 'JND-matched' stimulus orientation.

Two-cue performance: Effect of hand-object offset

Figure 2.7 plots visual-haptic size-discrimination performance (JNDs), averaged across all observers (individual data are shown in Appendix 2). The solid horizontal lines show observed single-modality JNDs for vision-alone (black) and haptics-alone (grey). The dashed line represents the statistically-optimal prediction if both signals are fully integrated (Equation (1.8)).

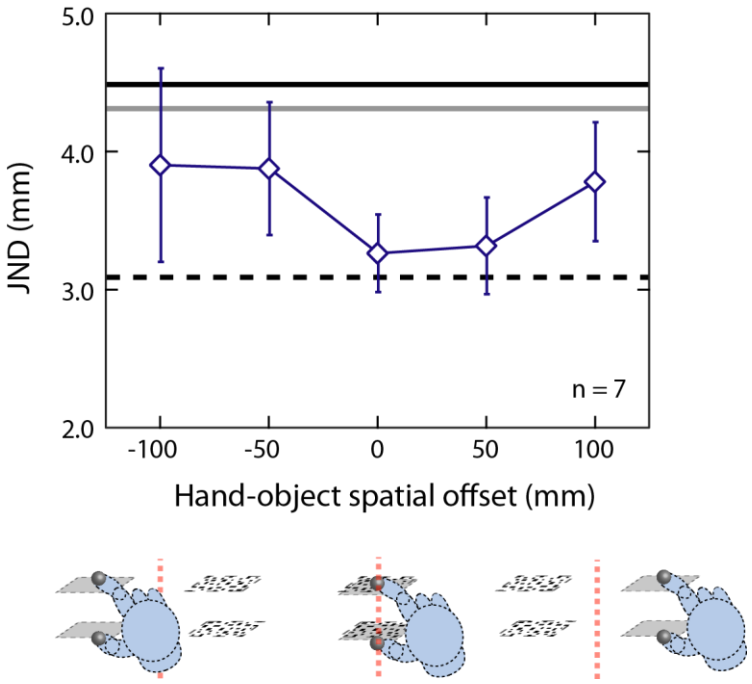


Figure 2.7. The effects of hand-object offset on discrimination performance in Experiment 1 (no tool). Mean size-discrimination performance (JND) is plotted as a function of spatial offset between the visual and haptic stimuli. The black and grey horizontal lines show mean discrimination performance from vision- and haptics-alone, respectively. The dashed line shows predicted performance if the signals are integrated optimally, calculated from the single-modality JNDs using Equation 2.1. Error bars denote ±1 SEM.

In general the results of our no-tool experiment replicated those of Gepshtein et al. (2005), even though Figure 2.7 showed that size discrimination performance was slightly asymmetrical from the zero offset. Discrimination performance was nearly optimal when the hand-object offset was zero, but was increasingly poor with increasing spatial offset. One-tailed *t*-tests showed that, compared to zero offset, the increase in JNDs did not quite reach statistical significance at |50| mm offset ($t(6) = 1.73$, $p = .068$), but was significant at |100| mm ($t(6) = 2.15$, $p < .05$). It shows both that the brain does integrate information from vision and haptics near-optimally when the signals originate from the same object (Ernst & Banks, 2002; Gepshtein & Banks, 2003) and, crucially, that integration reduces systematically when visual and haptic signals are separated, and so are likely to have come from different objects (Gepshtein et al., 2005).

2.3. Experiment 2: Grasping an object with a tool

As in Chapter 1, the natural and intuitive way that we use tools suggests that the brain employs a more sophisticated solution to the correspondence problem during cross-modal integration. Experiment 2 examined whether, during tool use, visual-haptic integration occurred, independent of spatial offset.

To do this, we used a simple tool and manipulated the tool length (spatial offset) between visual object and hand position in vision-plus-haptics condition with the same manner as Experiment 1. Thus both conditions (Experiment 1: 'no-tool' and Experiment 2: 'tool-use') are spatially identical. We measured the effect of spatial separation on discrimination performance when observers felt an object with a simple tool.

2.3.1. Methods

Observers

The same seven right-handed observers participated in Experiment 1 and 2.

Stimuli

In Experiment 2, visual and haptic stimuli were again presented simultaneously, but instead of only seeing spheres indicating the finger and thumb positions, observers now saw a 'virtual tool', consisting of cylindrical sticks, rigidly attached to each sphere (Figure 2.8), and so appearing to be attached to, and moving with, the finger and thumb.

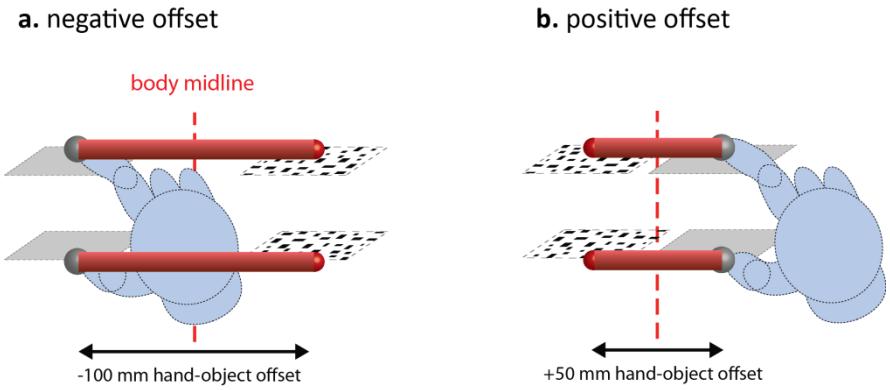


Figure 2.8. A schematic of the experimental stimuli. The visual and haptic stimuli were the same as Experiment 1 (See Figure 2.6). The tool was shown visually as a pair of parallel "sticks" rigidly attached to the finger and thumb, and was constrained to always lie horizontal. Grasping with the tool was simulated by spatially offsetting the visual and haptic stimuli horizontally. Thus when the tool-tip touched the visual object, force was generated at the finger and thumb. **a.** show negative and **b.** positive spatial offsets between visual and haptic stimuli, respectively (referred to as hand-object offset). The tool and finger-position spheres were extinguished before presentation of the haptic and visual objects (see Procedure). The information available to make size-discrimination judgements was therefore identical in both Experiment 1 and 2.

The tool freely translated with the finger/thumb in the x, y and z dimensions but rotation was prevented so that it was always oriented horizontally. Stimuli were presented with the same hand-object spatial offsets as in Experiment 1 (0, ±50 and ±100 mm). The length of the tool varied with hand-object offset such that the tips of the tool always reached the location of the visual object (Figure 2.8), indicating to observers that they could feel the remote visual object. As in Experiment 1, the hand-object offset (here, the tool length) was chosen randomly on each trial, to minimise the chance of visuo-motor adaptation (See 2.6. Control Experiment 1). We refer to the spatial offset between the tool-tip and the visual object as the tool-offset. Therefore in Experiment 2 there was always "zero" tool-offset.

Procedure

The procedure was similar to Experiment 1 except at the start of each trial observers placed the tool-tips into the start zones (instead of the finger and thumb), which indicated the location of the upcoming visual object. Once again all visual information, including the tool, disappeared before the visual and haptic stimuli were presented (see above). For a given hand-object offset, therefore, the information available to perform the discrimination task was identical to Experiment 1 (no tool). The tool was somewhat unintuitive at negative hand-object offsets, but these were necessary to prevent visuo-motor adaptation to an offset in a constant direction. Observers were given the opportunity to practice using the tool before the experiment began, and all found it straightforward to use.

2.3.2. Results

Observers learned to use the tool and the varied tool length in the practice session in the similar manner to the no-tool condition. Figure 2.9 shows the effect of hand-object offset on size discrimination in Experiment 2, when the tool-tips always reached the location of the visual object (zero tool-object offset, red circles; see Appendix 2 for individual data). The results of Experiment 1 (blue diamonds) are re-plotted from Figure 2.7. Here near-optimal visual-haptic integration was largely restored, despite the hand-object spatial offset.

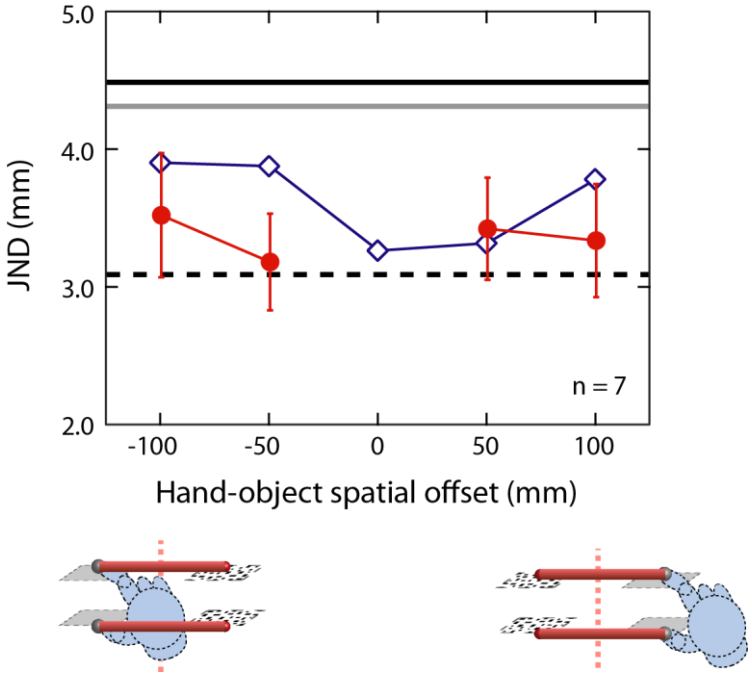


Figure 2.9. The effects of hand-object offset on discrimination performance in Experiment 2 (zero tool-object offset, red circles). Mean JNDs are plotted as a function of spatial offset between visual and haptic stimuli. The blue diamonds show the results of the no-tool condition (Experiment 1), re-plotted from Figure 2.7. (error bars removed for clarity). The horizontal lines again show mean single-modality discrimination performance and predicted optimal performance. Error bars denote ±1 SEM.

One tailed *t*-tests showed that discrimination performance with the tool was not significantly different from the near-optimal performance observed in Experiment 1, at zero offset, at |50| mm ($t(6) = 0.36$, $p > .05$) or |100| mm ($t(6) = 0.93$, $p > .05$) hand-object offsets. Moreover, JNDs were significantly lower with the tool than without the tool, both at |50| mm ($t(6) = 2.89$, $p < .05$) and |100| mm ($t(6) = 2.78$, $p < .05$) hand-object offsets. One possible exception to this overall pattern was the performance at minus 100mm offset. Although still showing significant cross-modal integration, discrimination performance at this offset was quite far from the optimal prediction. Interestingly, observers reported that this condition, in which a long tool came out of the "back" of their hand (see Figure 2.8 (a)) felt unnatural, suggesting that if the tool is unintuitive to use, cross-modal integration may be compromised.

Overall, however, the difference we observed between the pattern of results with and without the tool indicates that humans do integrate spatially offset visual and haptic signals when using a simple tool. Thus, visual-haptic correspondence problem is solved regardless of the spatial separation so that visual-haptic integration would take place "correctly" in tool use.

2.4. Experiment 3: Variable tool-object offset

In Experiment 3, we examined the effect on visual-haptic integration of a spatial offset between the tool-tip and the visual offset. If, when solving the problem of which signals to integrate, the brain takes into account the geometry of specific tools, we would expect to see a fall-off in cross-modal integration as the tool-tip is offset from the visual object. This would presumably be similar to that observed in Experiment 1 (no tool), above, when the hand was spatially offset, because the likelihood that the visual and haptic signals came from the same object would be similarly reduced. Alternatively, it is possible that observers in Experiment 2 adopted a simple "strategy" of always combining signals when using the tool. If so, we would again see cross-modal integration at all tool offsets. To examine this we again measured visual-haptic integration during tool use, and measured the effect of

offsetting the tip of the tool from the object by varying the offset, using a fixed-length tool (variable tool-object offset).

2.4.1. Methods

Observers

Six of the same observers participated in Experiment 3 with the continuation of Experiment 1 and 2. One exception was observer (BK), who was unable to continue.

Stimuli & Procedure

We measured size-discrimination performance in the same manner as Experiment 2, but using a fixed-length tool (50mm), the tips of which were spatially offset from the visual object by ±0, 50 or 100 mm (Figure 2.10). The different tool-offset conditions were randomly interleaved. All other details were as before.

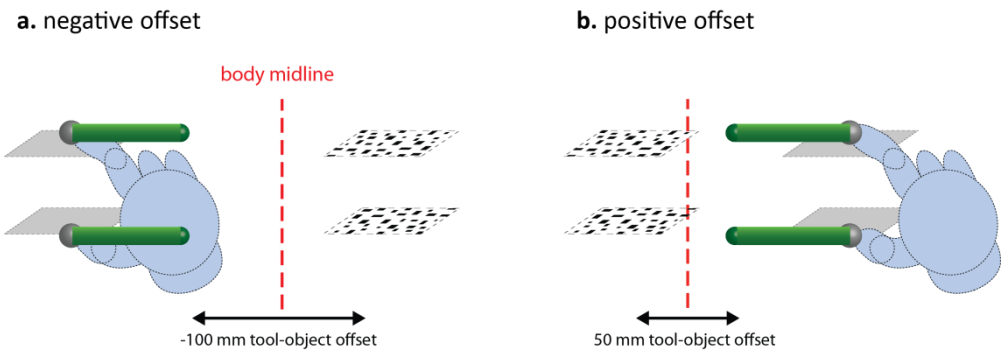


Figure 2.10. A schematic of the stimuli in Experiment 3. Here the length of the tool was always 50 mm, and the tips of the tool were spatially offset from the visual stimulus position by ±0, 50 or 100 mm (variable tool-object offset). It was possible to have zero tool-object offset with the hand to the left or right of the visual stimulus, so we collected data for both configurations. As before the tool and finger/thumb spheres were extinguished before the to-be-judged stimuli were presented.

2.4.2. Results

Figure 2.11 shows the results of Experiment 3 (variable tool-object offset, green squares/triangles) plotted with the earlier data from Experiment 1 (no tool). (See Appendix 2 for individual data). The data from Experiment 3 are plotted as a function of the tool-object offset. The data from Experiment 1 are plotted as earlier, as a function of the hand-object offset. Overall size discrimination performance was slightly poorer with the tool than without it, suggesting that tool use might increase sensory noise and uncertainty.

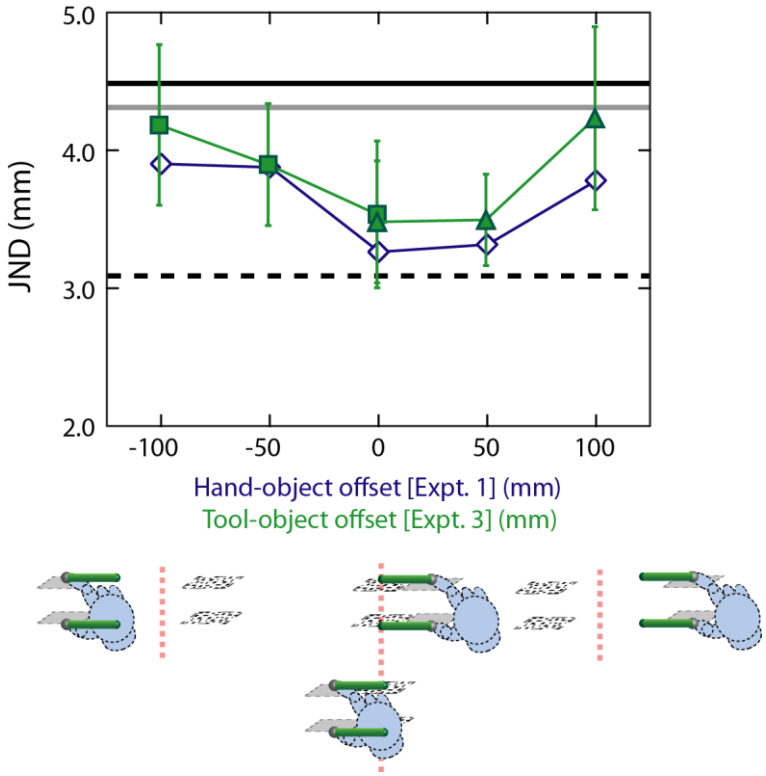


Figure 2.11. Mean size-discrimination performance (JND) for the tool with variable tip-offset (Experiment 3). Negative offsets are denoted by green squares, and positive offsets by triangles. The data from the no-tool condition (Experiment 1, blue diamonds) are plotted for comparison purposes. The data for the tool experiment are plotted as a function of the tool-object offset (the offset between the tool-tip and the visual stimulus). The data for the no-tool experiment are plotted as a function of the hand-object offset (the offset between the visual and haptic stimuli). The horizontal lines show the single cue JNDs for vision (black solid line) and haptics (grey solid line), and the predicted two-cue performance (dashed line). Error bars denote ±1 SEM.

The Figure shows that the changes in cross-modal discrimination performance with spatial offset were strikingly similar in Experiments 1 and 3. When the tool reached the visual object, discrimination performance was again close to that predicted by optimal integration. However, when the tool-tips were offset from the visual object, discrimination performance became systematically poorer, approaching single-modality levels. The pattern of effects observed with the tool use were similar to the results reported above in Experiment 1. Compared to the zero tool-object offset condition, JNDs were significantly larger with a tool-object offset of |50| mm ($t(5) = 3.20$, $p < .05$), and they were at |100| mm ($t(5) = 3.07$, $p < .05$).

Moreover, the similarity between the results in the two conditions (Experiment 1 and 3) was also reflected by statistical analyses. There were no statistically significant differences between JNDs in hand-object offset in Experiment 1 and JNDs in tool-object offset in Experiment 3, suggesting that observed data in Experiment 3 were better fitted by tool-object offset than hand-object offset. Repeated-measures ANOVA analysis between two conditions showed that there were statistically significant differences on the main effect of spatial offsets (0, |50|, |100|): $F(2,10) = 6.60$, $p < .05$, but importantly, not statistically significant on the main effect of the conditions (normal vs. variable tool-object offset): $F(1,5) = 2.05$, $p = .21$. The interaction was also non-significant (offsets × conditions): $F(2,10) = 2.78$, $p = .11$.

In addition, Figure 2.11 showed the similar asymmetry patterns on the spatial offsets in both the no-tool and tool-use conditions. This seemed to not only be due to unnaturalness of the tool extending backwards. As objects often exist beyond the hand space in the natural situation, it might be highly likely that haptics size estimates are combined with the visual size estimates at the space extended beyond the hand.

Offsetting the tool-tip, or the finger/thumb when grasping without the tool, therefore had a very similar effect on the extent of visual-haptic integration that was observed. These results are consistent with the idea that the brain can take into account the geometry and dynamics of tools, and integrates spatially offset visual and haptic signals only when it is appropriate to do so.

2.5. Discussion

2.5.1. Summary of results

For integration of information from vision and haptics to be effective, the brain should integrate signals if and only if each modality provides information about the same object. This means that the brain has to solve a correspondence problem: which of the activations in the given sensory modalities belong together? It has previously been reported that a spatial offset between where an object is seen and where an object is felt significantly reduces sensory integration (Gepshtein et al., 2005; see also Jackson, 1953; Warren & Cleaves, 1971; Witkin et al., 1952). Here we observed nearly optimal integration of size information from vision and haptics, when the spatial offset between the fingers and the object was bridged by a tool. In contrast, performance approached single-modality levels when the tool did not reach the object. These results cannot be explained by a fixed cue-integration rule that considers spatial coincidence of visual signals and an estimate of hand position. Instead we argue that the system is more flexible, making use of other information to determine when the "felt" and "seen" perceptual estimates refer to the same object.

2.5.2. Solving the spatial correspondence problem

As described in Chapter 1, for sensory integration to be effective, the brain needs to solve a general correspondence problem across sensory modalities. Several related models have successfully characterized to solve the correspondence problem within a Bayesian framework (e.g. Bresciani et al., 2006; Ernst, 2005, 2007; Knill, 2007a; Körding et al., 2007a; Roach et al., 2006; Shames et al., 2005). If sensory estimates are likely to occur together, the joint (common source) estimate will have a high probability, and if they differ substantially the joint estimate will have a low probability: it is more probable that the signals came from separate objects.

It should be highlighted that solving the correspondence problem does not result in a binary output (integration or no integration), but provides an estimate of the probability that the signals have a common source. This can then be used to

determine the extent to which signals should be integrated. As such, these types of models give a good account of empirical data showing a gradual transition from complete integration, through partial integration, to no integration of signals, as they become increasingly dissimilar (e.g. Körding et al, 2007a; Roach et al., 2006).

These studies have generally modelled the relationship between two estimates of the to-be-judged stimulus property (e.g. perceived direction). In principle, however, this computation can be carried out over any number of relevant dimensions (Ernst, 2005), including time of occurrence, and the parameter of interest, namely spatial location. Such a model could therefore provide a good account of the progressive reduction in cross-modal integration with increasing spatial offset observed by Gepshtein et al. (2005) and in our Experiment 1 (no-tool). The coupling prior would describe how visual and haptic signals from the same location are most likely to originate from the same object, and signals from different locations are highly unlikely to originate from the same object. Therefore as the spatial offset between stimuli increases, the estimate of the probability that the signals have a common source decreases, and the signals are integrated less. If the uncertainty in the spatial offset is increased, then the uncertainty of the estimate of the common source and the variance of the coupling prior would increase.

2.5.3. The correspondence problem in tool use

The approach described above provides a plausible framework for understanding how the correspondence problem in visual-haptic integration might be solved for normal grasping. On its own, however, it cannot explain the finding that, during tool-use, observers showed near-optimal integration of spatially offset visual and haptic signals. How might this be achieved?

In visuo-motor control perspectives, an important concept is that of a forward or internal model: the visuo-motor system uses a copy of the motor commands to predict the future state, for example the position and velocity of a moving limb (Wolpert, Ghahramani, & Jordan, 1995). Such an approach can naturally be extended to include a representation of tool position, allowing the system

to predict which visual objects will touch the tool. Therefore, in the moment the mechanoreceptors in the fingertip signal make contact with an object, the brain has already predicted this stimulus and assigns the proximal sensory stimulus as arising from a spatial location that is the actual source of the signal: the tool-tip.

When some outputs of the sensorymotor system are used for inputs of the forward model to compute further predictions, the system could adjust the model and modify the output. If the forward model effectively predicts the tool-tip position from the current state and motor commands, interacting with the sensory information and the movements of the tool, it could reduce the error between the prediction and actual positions of the tool-tip. Assuming that visual and haptic object size estimates are combined at the object space, or at the tool-tip, the well-controlled tool movements produced by the forward model could reduce sensory noises when grasping and also enhance the correspondence between two sensory signals both spatially and temporally. Moreover, an online feedback process would be a useful strategy for correcting the initial programming errors of the tool movements and controlling the tool-tip to make appropriate contact with the object. These processes might also improve the performance of cue integration. The relationship between the forward model and cue integration might be not only direct feed-forward relationship but also feed-back processing. In the natural situation, the nervous system likely receives feedback of behavioural outcomes and evaluates it to improve the subsequent movement. However, note that in our experiment, behavioural outcomes as cue integration were less informative to the forward model because a tool was turned off during grasping an object and judging the object size, and feedback was less available for the behavioural outcomes.

This forward model idea is consistent with evidence from several domains. Informally, many of us have experienced the sensation that, when we pick up a tool, we can "feel" with its tip, even though haptic signals originate at the hand. Results of a number of empirical studies also support this idea, as described in Chapter 1. Studies of visual spatial attention have suggested that tools are incorporated into our sense of where our bodies are in space. Thus we hypothesize that the correspondence problem in multi-sensory integration is not solved on a spatial map of the locations of

the sensory stimuli, but is instead based on a spatial map of the inferred sources of those stimuli. This is a process that requires knowledge of the viewing geometry and the dynamics of the manipulated object and tools.

2.5.4. Alternative explanations

While forward models provide an appealing explanation of our results, we need to consider simpler mechanisms that could underlie the observed behaviour. First, the decision whether to integrate or not may be made based on visual information alone. As we have noticed, two previous studies have reported integration of visual and haptic signals despite a large spatial discrepancy. Di Luca, Ernst and Adams (2008) found that visual-haptic integration was induced when spatially separated objects were made to appear as two parts of the same object, by covering the gap between them with an occluder, where a phenomenon known as "amodal completion". Also, Helbig and Ernst (2007b), in a study much more closely related to ours, found near-optimal integration of visual and haptic shape information when a visual object was seen straight ahead in a mirror, but the haptic signal was in a different location, provided that the hand was seen touching the object.

One account of these findings is that visual information provided compelling evidence that the two signals referred to the same object, causing any spatial discrepancy between the signals to be overridden. This can be accomplished by knowledge or experience base. In terms of the above models, this can be implemented by making the coupling prior for spatial location flat. That is, all combinations of the spatial locations of the two sensory estimates are considered equally likely. The estimate of the probability that the signals have a common source will be unaffected by spatial offset, and will be made on the basis of other relevant dimensions (size/shape, time of occurrence, etc.)

Such a strategy – in effect a "mode" in which any spatial discrepancy is simply ignored – could account for the results of Experiment 2, in which visual-haptic integration occurred independent of spatial offset. It cannot, however, explain the results of Experiment 3, in which we found a systematic reduction in the degree of

cross-modal integration with spatial offset of the tool-tip from the visual object. Instead, our findings suggest that the correspondence problem was solved appropriately, trial-by-trial. Helbig and Ernst's (2007b) result suggests that merely seeing the moving hand apparently in contact with the object is sufficient to determine that visual and haptic signals should be integrated. Could only vision of the finger cursors and the tool have provided this information in our experiments? This seems unlikely for two reasons. First, our procedure differed from Helbig and Ernst's (2007b) in a key respect. Their observers could see their hand while they actively explored the object, whereas in our experiments, the hand was never visible, and visual feedback about the tool was extinguished prior to contact with the haptic object. Therefore in our task, observers needed to predict the position of the tool-tip based on the motor commands to solve the correspondence problem appropriately. Second, and more generally, it is unclear that vision-alone is sufficient to identify a hand (or tool) as being one's own. As Helbig and Ernst (2007a) point out, this process, too, presumably depends on a comparison of the expected and actual visible movements of the hand, given the motor commands.

2.5.5. No visuo-motor adaptation during tool use

A second alternative explanation of these results is that the changes in integration are caused not by the brain predicting the positions of the tool-tips, instead by visuo-motor adaptation of the estimated hand position (Ernst, 2005). In our experiment, according to this explanation, the estimated hand position is shifted towards the tip of the tool. The fixed rule of comparing hand and object position can then still be applied to solve the correspondence problem.

Prism adaptation studies have demonstrated that the mapping between visual and motor space can readily be changed (Redding & Wallace, 1997). In our tool conditions, therefore, the magnitudes and signs of the spatial offsets were randomly interleaved on a trial-by-trial basis (with a mean of zero), to try to prevent a consistent "error signal" required for adaptation to occur. Nonetheless, it remains possible that adaptation occurred rapidly, on single trials, in the period during which the tool was

visible. To explore this we ran a control experiment in which we examined whether the felt position of the fingers adapted while using our tool (for details see the following 2.6. Control Experiment 1). Observers were asked to make grasping movements similar to those in Experiments 1–3, under two conditions: (i) a no-tool condition, in which the visible cursors indicating finger position were offset horizontally from the actual finger position (this was expected to induce adaptation), and (ii) a tool condition, similar to Experiments 2 and 3, using tool lengths that matched the offsets in the no-tool condition. Trials were blocked by condition, to provide a good opportunity for adaptation to occur. At regular intervals we probed adaptation by asking observers to place their unseen fingers at the location of two visual crosses.

As expected, in the no-tool condition observers showed a clear adaptation effect of between 50 and 60% of the spatial offset between the cursors and actual finger position. However, there was no adaptation in the felt position of the fingers in the tool condition, although the offset between the tool-tip and hand was identical to the no-tool condition. Furthermore, we carefully manipulated spatial offset randomly ordered in the main experiments, so prism situation was unlikely to be involved. Thus, these results argue strongly that no visual-haptic remapping of the felt finger position occurred during tool use, and that the changed integration observed in Experiments 2 and 3 must be due to a process that is dissociable from normal visuo-motor adaptation.

2.5.6. Causal inference in sensory integration

While the mechanisms underlying visual-haptic integration during tool-use remain to be determined, our data suggest that the visuo-motor system is able to correctly decide when it is appropriate to combine visual and haptic signals under the conditions of spatial transformation imposed by tool use. These findings are consistent with the broader idea that the perceptual system routinely makes inferences about the causal structure of sensory signals (Körding et al., 2007a). That is, in this case, the decision to integrate or not is made on the distal causes of the sensory stimuli, rather than simply on their spatial proximity.

The results of Gepshtein et al.'s (2005) study, and of our no-tool experiment, also suggest that solving the correspondence problem may be obligatory. In both studies, the visual and haptic stimuli were the only signals presented to observers in an otherwise impoverished sensory environment. Moreover, the stimuli were perfectly correlated on every stimulus interval in terms of magnitude, and time of occurrence, differing only in terms of spatial offset. An ideal observer might therefore be expected to integrate these signals because both were informative about which interval contained the larger object. Yet even under these circumstances a (relatively small) spatial offset between the felt and seen locations of an object reliably reduced sensory integration in both studies. This suggests that the solution to the correspondence problem is provided by an automatic and robust process, which cannot simply be "switched off", even if the task performed would favour such a strategy.

2.6. Control Experiment 1: Does the felt position of the fingers adapt during tool use?

As discussed above, one alternative explanation of our results is visuo-motor adaptation of estimated hand position. We therefore ran an adaptation experiment to examine whether, in our tool conditions, the felt position of the fingers was remapped onto the location of the tool-tips.

2.6.1. Methods

Observers

Four right-handed observers (aged 21-44 years) participated in this experiment. All observers had normal or corrected to normal vision, normal stereo-acuity, and no known motor deficits. Two (PA, AW) of four observers completed all the previous experiments, one (BK) completed Experiment 1 and 2, and one (BF) was new. All the observers were naïve to the purpose of the experiment.

Stimuli

The visual object was a plain white rectangular object, of a constant size (50mm × 50mm × 20mm). We used the force-feedback devices to limit movement to a fronto-parallel plane (see Figure 2.4), to simplify data analysis. The visual stimulus was presented alternately on the left and right side of the workspace.

Procedure

An experimental trial was similar to one interval of Experiment 1 or 2. The visual stimulus was presented alternately on the left and right side of the workspace. The exact x and y position of the stimulus was selected at random within a 100mm × 100mm region, centred either 75mm to the left or right of the body midline. This ensured that observers made significant movements on consecutive trials to reach the stimulus. The start zones operated in the similar manner as before (Experiments 1-3), and the tool and/or cursors were again extinguished when the observer started to move their finger toward the object. In this experiment, observers were not required to make any judgement.

Along with a baseline condition, in which the cursors indicated the true position of the finger and thumb, observers repeatedly grasped the stimulus under (i) two cursor-offset conditions, in which the cursors were offset to the left of the physical position of the finger/thumb by either 50 or 100 mm but no tool was presented, and (ii), tool conditions, in which the tool length was either 50 or 100 mm. Periodically, observers completed a test trial, in which they were required to place their finger and thumb at the position of two crosses (whose position was determined in the same way as the start zones). In this test period, only two crosses were provided on the screen and there were no other visual stimuli such as cursors or tool representations. Thus, observers received no visual feedback in the test trials, and the position of each digit was recorded. The test period was identical in the all conditions.

The experimental protocol (Figure 2.12) started with 10 baseline trials, with a test trial after every five trials. Each tool/offset condition consisted of 36 adaptation

trials, with a test trial after every six trials. After each adaptation condition there was an additional 18 trial baseline period, with a test trial after every six trials. This was a "wash-out" period, designed to eliminate any adaptation from the previous condition. Aiming to clearly distinguish the baseline between the initial and wash-out periods, here we called only the initial 10 trials with two tests as "baseline", and the 18 trials with three tests succeeding each adaptation condition as wash-out. Each observer completed two complete blocks (72 trials in each tool/offset condition).

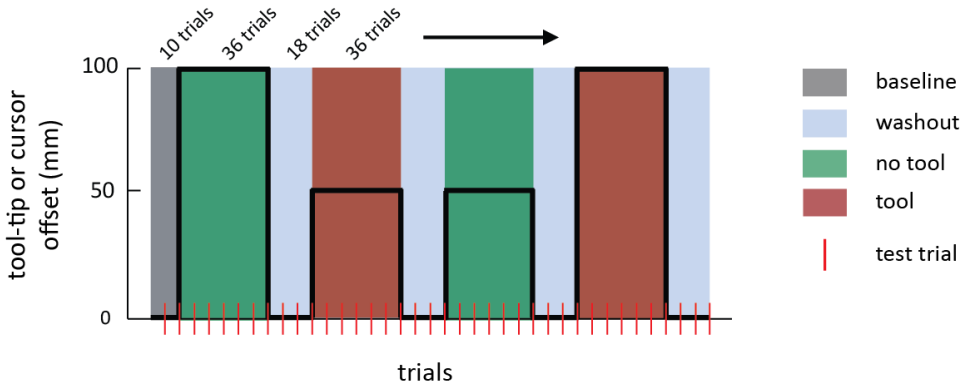


Figure 2.12. Example of one 'block' of the experiment. The tool (50 or 100 mm) and finger/thumb offset (50 or 100 mm) conditions were randomly ordered within each block. Each observer completed two such blocks. The cursors were always offset to the left of the true hand position, and the tool also extended to the left of the hand. Test trials, in which we measured the amount of adaptation in the felt position of the fingers, occurred after every five trials in the initial baseline period, and after every six trials thereafter. One block consisted of 226 trials with 38 tests (See text for details).

2.6.2. Results

Figure 2.13 plots the average error (in x-position) of the thumb/finger positions across all test trials, averaged across four observers (there were no statistically significant changes in adaptation within conditions, so we averaged across all trials). Positive errors mean the fingers were positioned to the right of the test targets. The gray area is the baseline condition, which indicates averaged performance of the initial periods of two blocks.

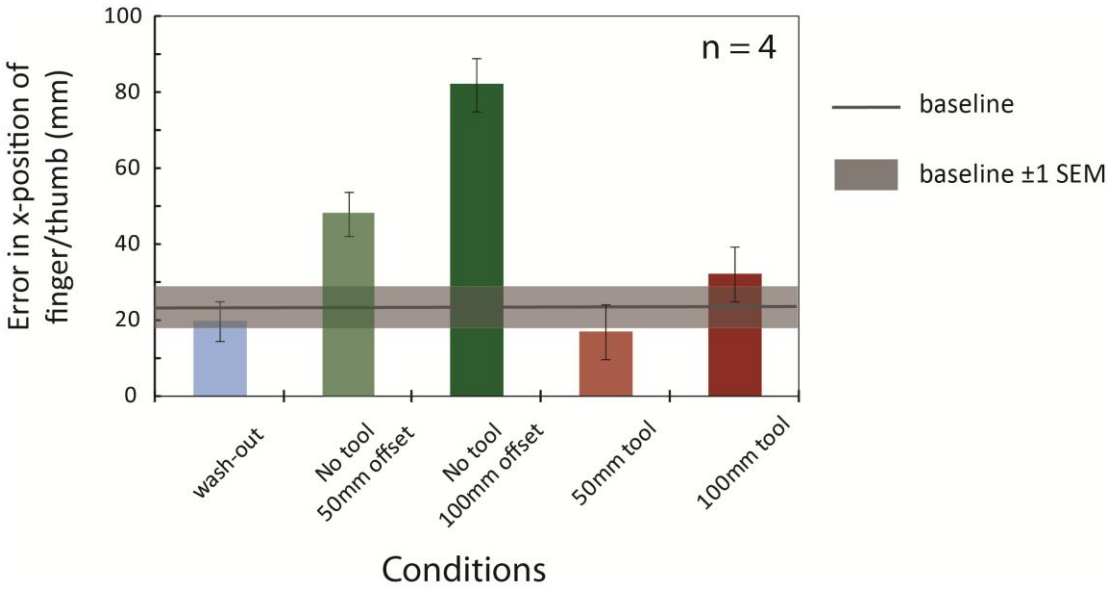


Figure 2.13. Average error in the x-position of the finger and thumb on test trials, for all observers. The gray area represents the baseline condition to allow comparison with each adaptation condition. Error bars denote ±1 SEM.

Figure 2.13 suggests that there was a small overall bias in observers' responses; observers seemed to position their fingers approximately 20mm to the right of the test target in the baseline condition. To examine this further, we compared the baseline performances between the first and the second blocks. The first block average (*M* = 24.5mm, *SD* = 7.3mm) and the second block average (*M* = 22.4mm, *SD* = 28.6mm) were quite similar around 20mm. The second block baseline performance was relatively varied than the first block. The variable in the second block might have

been influenced by the previous block or might have involved other factors such as tiredness; however, the reason was hardly identified in our analyses.

Here, further examination of the wash-out period was conducted to investigate whether the error in x positions of finger/thumb was affected by the previous conditions: no-tool/tool and degrees of offsets. Repeated-measures t-tests were used for comparing between the wash-out periods after the four different conditions and the baseline performance across two blocks and four observers. There were not statistically significant differences for all wash-out periods: after 50mm offset, $t(7) = 0.14$, $p = .89$; after 100mm offset: $t(7) = -0.47$, $p = .65$; after 50mm tool: $t(7) = 0.85$, $p = .42$; after 100mm tool: $t(7) = 1.07$, $p = .32$. Thus, during the wash-up period, observers seemed to return their position to the baseline, regardless of the previous conditions.

Although individual differences were relatively large (see Appendix 3) and the number of observers was limited, overall it seemed to be 20mm bias across conditions. However, it is unclear whether it was some generalised adaptation or not; the bias might be caused by mis-calibrations of the perceived centre of the workspace and/or observation errors of the finger position representing by a small sphere at the centre of PHANToM robot attachment (finger sledge).

Under the above consideration, we analysed the obtained data in the four adaptation conditions with the baseline performance. Compared with the no-tool and the baseline, Figure 2.13 shows that offsetting the cursors to the left of the true finger position caused observers' to make significant positive errors on the test trials, by approximately 50-60% of the cursor offset. This indicates adaptation of the felt position of the fingers, as was to be expected from conventional visuo-motor adaptation studies. Repeated-measures t-tests across four observers and two blocks show that there were statistically significant differences between baseline and no-tool conditions for 50mm offset, $t(7) = 2.86$, $p < .05$, and for 100mm offset, $t(7) = 6.77$, $p < .01$, two-tailed. In contrast, the errors made in the tool conditions were indistinguishable from the baseline condition. The repeated-measures t-tests show that there were no statistically significant differences between baseline and tool conditions: for 50mm tool, $t(7) = -0.75$, $p = .48$, and for 100mm tool, $t(7) = 0.84$, p

= .43. In addition, a repeated-measures ANOVA test shows the statistically significant differences between the four conditions; the main effect for no-tool/tool was significant, $F(1,7) = 72.75$, $p < .01$, and a main effect for offsets 50mm/100mm was significant, $F(1,7) = 105.21$, $p < .01$ and the interaction was also significant $F(1,7) = 717.54$, $p < .01$.

These indicate that there was no adaptation of the felt position of the fingers when using the tool, even though the same tool was used repeatedly on consecutive trials. We therefore conclude that it is very unlikely that our main tool effects can be attributed to visuo-motor adaptation of the felt hand position (see 2.5 Discussion).

2.7. Chapter conclusion

We have found that humans integrate size information from vision and haptics in a near-optimal fashion when using a simple tool that introduces a spatial offset between the spatial locations of the two signals. Moreover, we observed a fall-off in cross-modal integration with increasing offset of the tool-tip from the visual object. This is consistent with the brain deciding that the haptic signal originated at the tool-tip, rather than the hand. We suggest that the brain therefore infers appropriately the causal structure of the two signals and integrates them only when it is appropriate to do so. Whether this is achieved by constructing a forward model of the tool, or a more straightforward "strategy", remains to be determined. Nonetheless, it appears the cross-modal integration process is more sophisticated than previously thought, and can take into account the dynamics and geometry of tools.

Chapter Three

Effects of magnitude change on visual-haptic integration in tool use

3.1. Introduction

3.1.1. Remapping of haptic signals in tool use

We previously showed that when using a simple tool to estimate an object's size, visual and haptic information was integrated in a near-optimal fashion, even with a large spatial offset between the two raw sensory signals (Experiment 2). The integration was reduced however when the tool-tip was away from the object (Experiment 3). As discussed in Chapter 2, these results suggest that during tool use the haptic signals are treated as originating from the tool-tip, not at the hand. One explanation for this is that the brain takes account of the tool geometry, and haptic signals are spatially remapped accordingly. This would allow the correspondence problem to be solved in "object space", where the tool-tip is actually acting. However, tools change spatial relationship not only in terms of offset between object and hand positions, but also the magnitude between the object size and hand opening.

In this chapter, we will examine how the brain solves the correspondence problem when tools change the magnitude between the tool-tip and the handle, followed by a review of related studies in respect to combining conflicting stimuli.

3.1.2. Conflict studies and visual-haptic integration

Normally, visual and haptic signals indicating different sizes would be strong evidence that these two cues do not share a common cause. However, when grasping an object using a tool such as pliers, there are inherently differences in size between the tool-tip and the handle. As described in Chapter 1, from a statistical point of view, it can be hypothesised that if the brain understands the dynamics and geometry of a tool, the cue integration could occur even though two raw sensory signals are "conflicting" in size. However, to combine conflicting stimuli, rescaling of cross-modal bias would be essential to generate a unified percept (De Gelder & Bertelson, 2003; Ernst & Bülthoff, 2004). Previous studies showed that the weighted linear summation model is still applicable for even large conflicts (e.g. Oruç et al., 2003; Roach et al., 2006), while others have postulated that the final estimate would be mandatorily combined in order to avoid perceptual contradiction (e.g. Hillis et al., 2002).

The effect of conflict in magnitude has not been well examined for visual-haptic integration, but it has been explored in other modalities (Girshick & Banks, 2009; Hogervorst & Brenner, 2004). For example, when binocular disparities and vergence are conflicting in scale, these cues are rescaled and combined for perceiving depth and distance of 3-D scene appropriately (Landy & Brenner, 2001). Recently Schindel and Arnold (2010) examined the relationship between visual sensitivity and scaling by measuring orientation discrimination thresholds. They showed that the sensitivity affected spatial scaling process, by manipulating viewing distance. Therefore, we could anticipate that visual and haptic sensitivities also play an important role in perceiving an object's size when a tool rescales the sizes.

In addition, Adams, Banks, and van Ee (2001) measured human slant perception by manipulating the relationship between disparity and texture cues. They used glasses, which produced horizontal distortion and changed image magnification, so as to introduce a cue conflict in slant. They showed that cue weights given to disparity were not changed by the glasses, suggesting that the relationship between sensory cues was not changed; rather, the relationship between retinal disparity and slant perception was changed. The cue conflict between disparity and texture was not

large when wearing glasses; thus spatial remapping might not be necessary to combine them in order to perceive slant in this study.

Recently, Girshick and Banks (2009) measured slant discrimination thresholds manipulating magnitude of conflict between disparity and texture cues. They showed that when there was a small conflict of magnitude between two cues, slant estimation followed the weight averaging model. In contrast, when there was a large conflict, the two cues were robustly combined to estimate slant. Their results were consistent with Bayesian models such as the coupling prior and the causal inference. The brain likely employs robust estimations when a spatially large conflict exists between information sources (Ernst, 2005; Ernst & Bülthoff, 2004; Girshick & Banks, 2009). These related studies also provide practical implications for modelling tool use in multisensory integration discussed at the end of this chapter.

3.2. Experiment 4: Size-discrimination performance

Here we asked whether the brain solves the correspondence problem correctly when a tool dynamically rescales object size estimation from haptics. To do this, we simulated a pair of pliers, which linearly changed the relationship in size between the tool-tip and the handle. We use the term gain to explain the scalar change in size. We then measured size-discrimination thresholds using three tools with different gains.

3.2.1. Experimental predictions and design

Imagine a tool, which changes the relationship between the tool-tip and the handle with a gain. When using such a tool to estimate an object's size, visual signals from the object (at the tool-tip, or we call the space "object coordinates") and raw haptic signals from the hand opening (at the handle or hand coordinates). The relationship between the seen object size and hand opening are changed with the tool gain. Thus, even when object and tool-tip size are matched, the object size and hand

opening are different because the gain ≠ 1. However, both refer to the same object. Thus, to estimate the object size precisely, the brain needs to combine visual and haptic size information appropriately.

We designed four different conditions to examine the visual-haptic correspondence problem during tool use (See Figure 3.1). We used a pair of pliers and manipulated the pivot position to change the ratio (or gain) of the opening of the tool-tip and the handle. The pliers smoothly followed the hand movements in a natural manner; that is, when closing the hand opening, the tool-tip size was getting small, and when widening the hand opening, the tool-tip size was getting bigger. We generated two different gain tools; one preserved the normal 1:1 relationship between the tool-tip and the handle, and another magnified it. To show them distinctively, the two tools were coloured blue and red respectively. When the tool-tips touched the haptic surfaces, an appropriate force was rendered at the hand.

Figure 3.1 illustrates the four combinations between visual object size and haptic object size in the object coordinates, with the two gain tools. As shown in the Figure, two object sizes were used and the hand opening size was all the same. We assumed that haptic size would be appropriately rescaled from hand coordinates to object coordinates according with the tool gain. When visual object size and haptic object size were matched, even when the object size and hand opening were conflicting, we called this a "no-conflict" condition. In contrast, when visual and haptic sizes were conflicting, even when raw sensory signals were matched, we called this a "conflict" condition.

Aiming to clearly explain the four conditions, we depict the actual values (of gain) used in Experiment 4. The blue tool was the same size at the tool-tip and at the handle (visual size: hand opening = 1:1), so we call this the 1:1 tool. The red tool was 1.6 times bigger at the tool-tip than at the handle (visual size: hand opening = 1.6:1), so we call this the 1.6:1 tool. Here, the hand opening remained constant at 50mm. When using 1:1 tool, haptic size in the object coordinates was 50mm; in contrast, when using 1.6:1 tool, haptic size was 80mm.

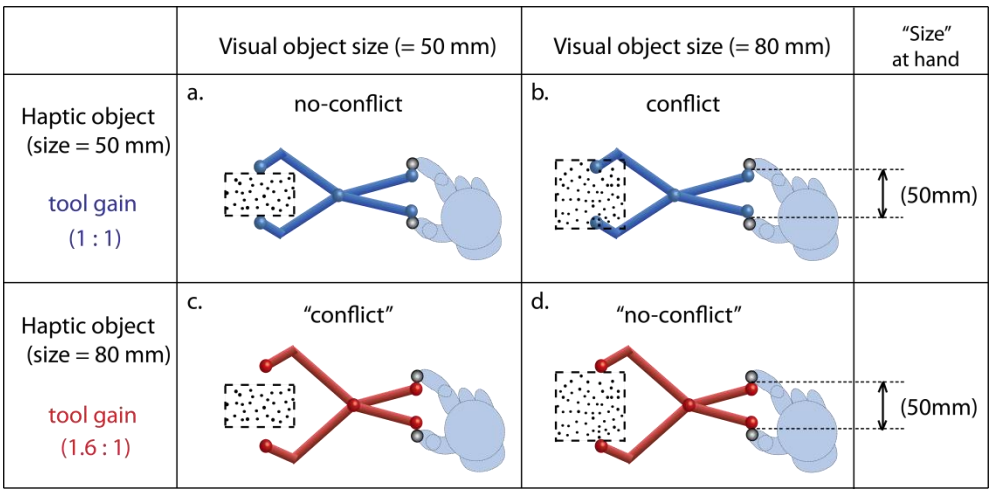


Figure 3.1. The four experimental conditions: hand opening is, on average, the same size in each condition (50 mm). **a.** no-conflict condition using the 1:1 tool. Visual object size is equal to haptic object size (no-conflict). **b.** conflict condition using the 1:1 tool. Visual object size is not equal to haptic object size (conflict). **c.** "conflict" condition using the 1.6:1 tool. Visual object size is not equal to haptic object size, although object size is matched to the hand opening. **d.** "no-conflict" condition using the 1.6:1 tool. Visual object size is equal to haptic object size, although object size does not match to the hand opening.

The 1:1 tool case (a and b) was a simple relationship, because the tool-tip and the hand opening were the same size. In the conflict condition, the tool-tip unrealistically moved into the object (b). In contrast, the 1.6:1 tool case (c and d) was more complicated. In the conflict condition, the 1.6:1 tool did not reach the object (c). When visual object sizes were the same, raw sensory signals (vision and hand opening) were quantitatively the same regardless of the tool gain (a and c, b and d). We could expect that the visual-haptic integration only occurred in the "no-conflict" conditions (a and d); if so, this would be strong evidence that spatial mapping is involved when using a tool. Alternatively, it is possible that observers adopted a simple "strategy" of always combining signals when using the tool; if so, we would see visual-haptic integration in all conditions.

3.2.2. Methods

Observers

Seven right-handed observers (aged 18-34 years, 2 males) took part in this experiment. All observers had normal or corrected to normal vision with normal stereo acuity, and no known motor deficits. All observers were naïve to the purpose of the experiment.

Apparatus & Stimuli

The same apparatus was used as in Chapter 2. The visual and haptic stimuli represented a raised bar (rectangular object, or surface) which was made by parallel rectangular planes. It was similar to those used by Ernst and Banks (2002).

Figure 3.2 shows the visual stimuli made by random-dot stereograms depicting two parallel planes, which represented an object front plane and a ground plane. The front plane was located above the ground plane (80mm width, 200mm height), separated with the object depth (20mm). The object width matched the ground plane (80mm) and the object height was varied. These planes were perpendicular to the line of sight (33.415 degree in the mid-sagittal plane) and located at the left side of the workspace. Each plane consisted of uniformly distributed dots. The dot size was 4.0 ± 1.0 mm (diameter), and the density was 0.20 dots per mm^2. We used anti-aliasing to achieve sub-pixel accuracy of dot positions. Because the borders between the two surfaces were essential to judge the object size, we randomly chose 3% of the total number of dots and moved them to the edge. Therefore, the dot density at the border is slightly higher. The viewing distance to the object ground plane was 500mm. The dots subtended 14 arcmin at the viewing distance. To vary visual reliabilities, we added random displacements of dots in the stereogram. (See Procedure for more details).

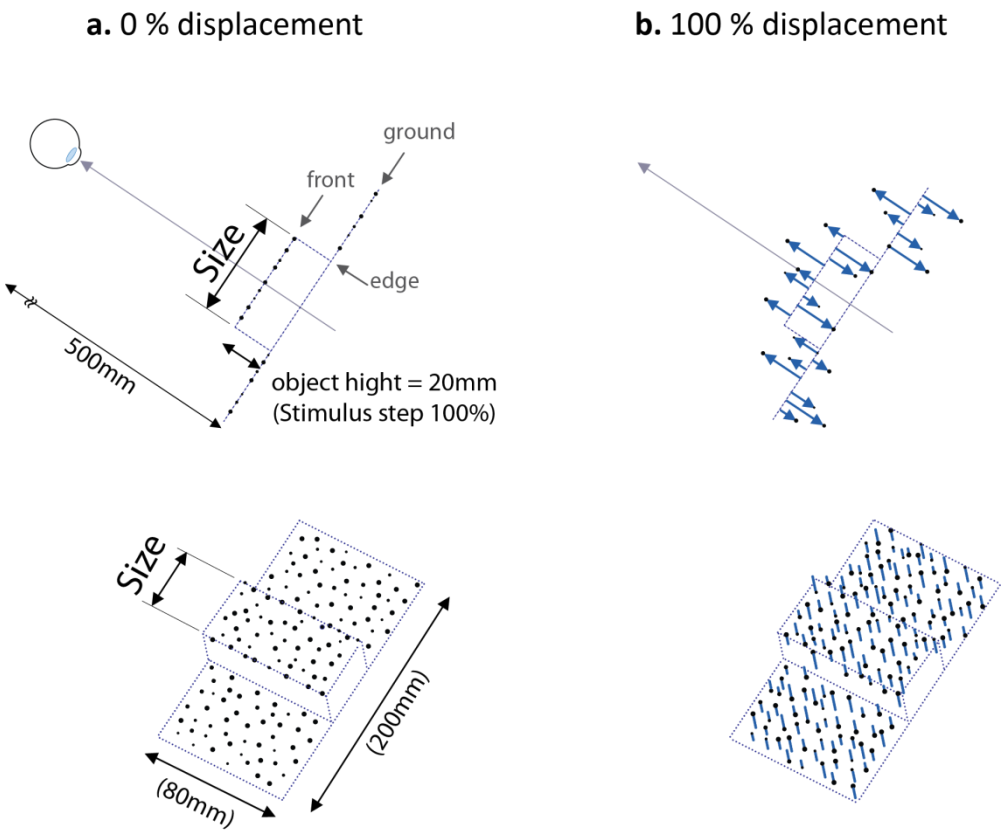


Figure 3.2. Visual stimuli. Random-dot stereogram appeared at the left side of the workspace. The object front plane located at 20mm high from the ground plane. Displacement noises were generated from a random uniform distribution with 2 cm depth step. **a.** 0% displacement. Dots were on the planes. **b.** 100% displacement. Dots were randomly moved from the minimum (0cm) to the maximum (2cm) depth with 100 steps (per 0.02 cm) with equal probability.

Figure 3.3 shows the haptic stimuli, illustrating the relationship between the two tools and the visual stimuli. The haptic stimuli were two parallel force planes at hand. We made two explicitly different gain tools, which visually simulated a pair of pliers attached to the finger and thumb and changed the relationship between tool-tip and handle by manipulating the pivot position. These tools smoothly follow hand movements and only when the tool-tip touched the virtual haptic surfaces, and even when visual and haptic sizes were mismatched in the object coordinates (b), the opposite force was generated when the tool-tip touched the object surface, according to the tool geometry.

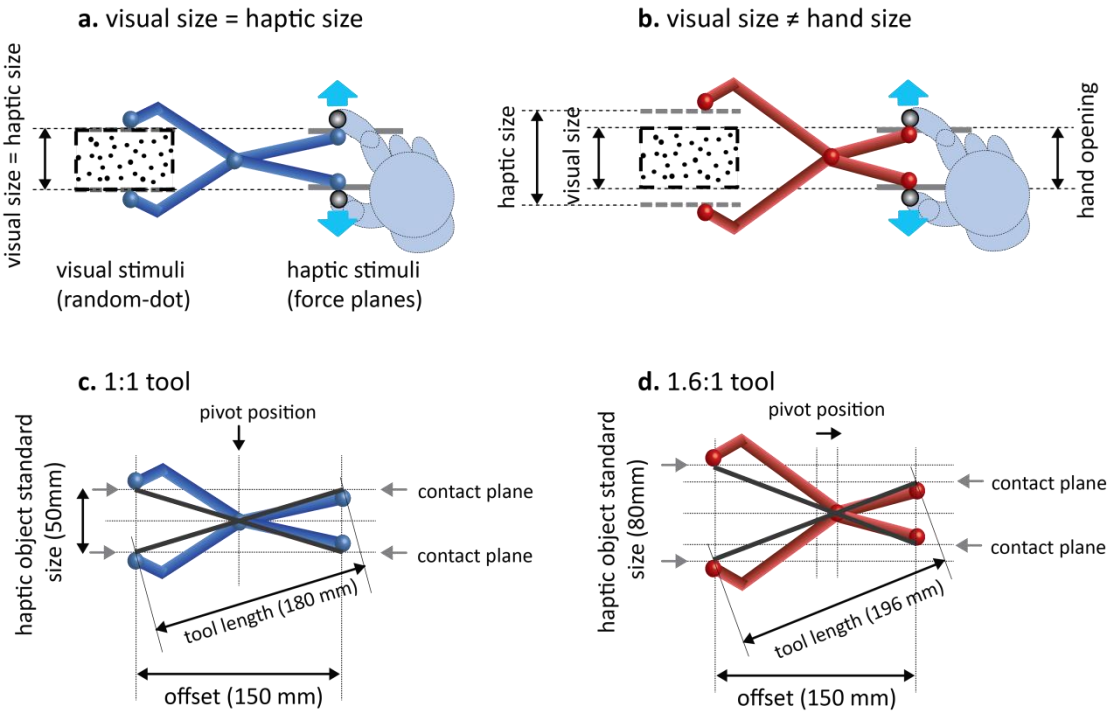


Figure 3.3. Haptic stimuli and geometry of the virtual tools. Force planes appeared at the hand when touching the virtual surfaces with the tool-tips. The light blue arrows represent the opposing force at the finger and thumb. Examples of the relationship with visual stimuli: **a.** visual size = haptic size when using the 1:1 tool. **b.** visual size ≠ haptic size when using the 1.6:1 tool. Tool dimensions: **c.** the 1:1 tool and **d.** the 1.6:1 tool.

The tool consisted of three parts (tip, handle, and pivot). Aiming to avoid unrealistic movements, the tool-tip was formed from two cylinders (short and long), so as to establish contact from the outside of the object. The cylinder sticks were 2.5mm radius. In order to set a constant offset between visual size and hand opening for the standard stimuli, the tool length differed between the two tools (180mm for the 1:1 tool and 198mm for the 1.6:1 tool). The tools dynamically followed hand movements and the two cylinders rotated around the pivot position. The tool movements were visually constrained on the plane which was parallel to the object's ground plane (0 degree to the line of sight), in order to simplify the generation of

forces. Thus, the tool freely translated with the finger and thumb in the x and y dimensions.

Procedure

Our procedure was similar to Ernst & Banks (2002) and the previous experiments reported in Chapter 2. We measured single-modality (vision-alone and haptics-alone) discrimination performance first and used this to predict optimal integration when visual and haptic reliabilities were matched. We then measured the cross-modal (vision-plus-haptics) performance under the four different conditions. In all conditions, visual and/or haptic size (defined as the separation between the planes) was varied according to a method of constant stimuli, and observers were required to judge object size in a 2-IFC task. Observers indicated which interval contained the larger size. In each condition observers completed 30 repetitions of each stimulus level. No feedback was given at any stage during the experiment.

As described in Chapter 2, based on theories of sensory integration, the statistically optimal strategy for integrating signals from vision and haptics can be calculated using Equation (1.8)). If cue integration occurs, the maximum improvement in discrimination would appear when the two signals have equal variance. It is therefore important experimentally to match the variance of estimates from each cue, so that the effects of cue integration are clearly evident. In the same way Ernst and Banks (2002) did, we matched cue reliabilities between vision and haptics. This procedure enabled us to evaluate visual-haptic integration, when using a tool under the four different conditions.

Single-modality experiment

We measured JNDs for two standard object sizes: 50mm and 80mm.

In the vision-alone condition, the object size was varied with the standard and eight comparison sizes. When the standard size was 50mm, the comparisons were 50mm +/- 1, 3, 6, and 9 mm, so 41, 44, 47, 49, 51, 53, 56, 59 mm sizes were used.

When the standard size was 80mm, the comparisons were 65.6, 70.4, 75.2, 78.4, 81.6, 84.8, 89.6, 94.4 mm. These were determined by multiplying the comparisons at 50mm by the size ratio (80/50 = 1.6), assuming that visual reliabilities would approximately follow Weber's law.

We measured visual size-discrimination threshold as a function of visual noise. The noise levels of binocular disparity were manipulated by adding random displacement on a given trial. The displacement direction was parallel to line of sight along object depth (See Figure 3.2). The percentage of noise was normalised by the object depth (20mm). Each dot was drawn from the object surface with a certain displacement, which was calculated by a random uniform distribution. Thus, 100% noise means that the displacement was equally distributed with a 1% step from zero (object surface) to the maximum (object depth) with equal probability. We set the range of the noise percentages (0 – 400 %) based on the discrimination performance of the individual observers. The noise range was well covered enough to find out the points where vision and haptic reliabilities were matched at two object sizes (50mm and 80mm). A new stereogram was generated for each plane on each stimulus presentation.

In the vision-alone condition, a fixation-cross appeared at the beginning of each trial, indicating the position of the upcoming stimulus (but no relevant size information) was given. In addition, because the visual stimuli were noisy, some observers had difficulties in fusing the 3-D object. Therefore, a white solid rectangular borderline was displayed just inside the screen edge, which aided 3-D fusion. The two intervals were presented for 1 sec each, separated by a 1.6 sec inter-stimulus interval. This time was determined based on the typical inter-stimulus interval when haptic information was available. Observers then indicated which interval contained the larger size by pressing one of two virtual buttons. The presentation order of standard and comparison stimuli was randomised. This is the same way as described in Experiments 1-3.

In the haptic-alone condition, we measured haptic sensitivity without a tool, assuming that there were no significant differences in haptic sensitivity at hand coordinates between no-tool and with-tool representation. This assumption was made

considering the previous findings (Experiment 1-3); specifically, there were no statistically significant differences between normal grasping (Experiment 1) and tool-object offset (Experiment 3). Although Figure 2.11 showed similar trends between two experiments, the data indicates that JNDs were relatively larger in the tool condition than no-tool condition. This would make sense as some noise and uncertainty likely increased when estimating object size when using a tool. Note that these two experiments were visual-plus-haptics conditions. The contribution of tool use was unclear on haptics sensitivity; so haptic performance might be different. This will be discussed in Chapter 5.

The object standard was 50mm, with the same comparisons as the vision-alone condition: 41, 44, 47, 49, 51, 53, 56, 59 mm, at the hand. The procedure in the haptics-alone condition was also similar to Experiment 1. Two gray small markers indicated the positions of their finger and thumb. Two visual spheres ("start zones") appeared, indicating the position of the upcoming stimulus (but no relevant information of the object size). When observers inserted their finger and thumb in to the start zones, the spheres' colour changed from yellow to green, indicating ready to start grasping the stimuli. Observers then moved their index finger and thumb inward to grasp the stimulus. When the observers moved their digits inward from the start zone positions, all visual information disappeared immediately. Observers were trained to grasp the stimulus for approximately 1 sec in each interval and then release it, aiming to match the vision-alone presentation timing. If the time both digits touched the surface was less than 900 msec, or more than 1200 msec, the message "too fast" or "too slow" appeared on the screen, and that trial was discarded. This process was repeated for the second interval, and the observer then indicated which interval contained the larger size, in the same way as the vision-alone condition.

Both vision-alone and haptics-alone blocks were completed separately. Observers learned the experimental procedure during a practise session at the beginning of the test block. As in Chapter 2, the observed data in each condition were fitted using a psychometric function (Cumulative Gaussians) and maximum-likelihood criterion to estimate JNDs.

Cross-modality experiment

We measured size-discrimination performance (vision-plus-haptics) in four conditions with matching reliabilities by adding visual noise respectively. Instead of only seeing markers indicating the finger and thumb positions in the haptic-alone condition, observers now saw a visual 'virtual tool', rigidly attached to the markers.

We introduced a visual "elastic band" between the current digit position and the tool handle, aiming to indicate the degree of distance between them and to help observers to set their digit appropriately. Because the tool orientation was fixed parallel to the object ground plane, the observers were required to adjust their hand orientation to match the tool, by minimising the length of the band. When fingers detached from the tool handle, this elastic band appeared; observers could adjust their fingers until the elastic band disappeared. The appropriate force was generated, only when the hand orientation was correct, otherwise observers could not feel anything in the virtual space.

The procedure was similar to the haptics-alone condition, with the exception of the start of the trial. Observers placed the tool-tips (instead of the finger and thumb) into the visual start zones, which indicated the upcoming object location (though not size). All visual information, such as the fixation cross, the start zones, the tool, and the markers indicating finger positions, were extinguished when the tool-tips were moved from the start zones towards the haptic object (prior to the visual stimuli being presented). The visual stimulus was presented only when both haptic planes were touched by the tool simultaneously. Thus, during the size judgement, only the random-dot planes were visible.

The four conditions were randomly assigned on a trial-by-trial basis to minimise any effects caused by a certain exposure of a constant stimuli. Observers were given the opportunity to practice using the tool before the experiment began, and all found it straightforward to use.

3.2.3. Results & Discussion

Matching cue reliabilities

Figure 3.4 plots one observer's JNDs for two standard sizes (50mm and 80mm) in both vision-alone and haptics-alone conditions. Data for all observers are shown in Appendix 4. Visual size-discrimination performance is plotted as a function of displacement noises and the two dashed lines represent fitting curves to the observed data (blue for 50mm and red for 80mm). Varying the noise had a large effect on the JNDs of size estimates from vision. This trend was similar to all observers. The two straight lines show haptic size-discrimination performance (blue for 50mm and red for 80mm). We actually measured haptic JND when the object size was 50mm. Then we calculated the JND for 80mm object, assuming that haptic JND to the object size was linearly rescaled form the JND at the hand opening with tool gain. (The validity of this assumption will be checked in Chapter 5.)

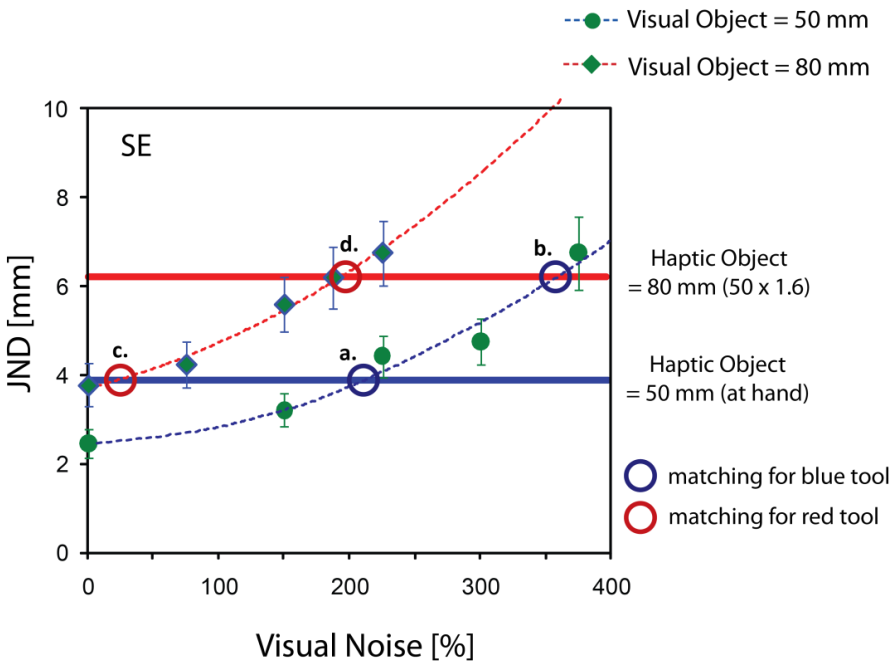


Figure 3.4. Discrimination performance (JND) of one example observer (SE) in vision-alone and haptics-alone conditions as a function of visual added noise. The displacement noise was normalised by the object depth; 100% meant that the maximum displacement was equal to the object depth. At 300%, the maximum displacement was three times the object depth. The degree of displacement of each dot was randomly assigned by the uniform distribution. The dashed fitting lines are the second order polynomial curves. The four points denoted (a, b, c, d respectively in the Figure) are matched to the four conditions in Figure 3.1. Error bars denote ±1 standard error.

Figure 3.4 was used to determine the noise level for the four conditions in the vision-plus-haptics. The combinations of vision and haptic sizes (S_V, S_H) were (50, 50), (50, 80), (80, 50), (80, 80), at the object coordinates. When using a tool, if the brain rescales haptic estimates appropriately from the hand coordinates to the object coordinates, the haptic threshold would be determined by the haptic object size at the object coordinates. Therefore, we expected that haptic JNDs were linearly rescaled according with the tool gain. (We will examine this later in the following chapters). We also assumed that sensitivity of hand opening did not change when using different gain (1:1 or 1.6:1) tools, because the hand opening was the same (50mm) for all conditions. Therefore, haptic JND for 80mm was calculated from the observed JND (at 50mm), by multiplying the gain 1.6. We varied the noise level for individuals, considering their visual and haptic discrimination performance. We then determined four different noise levels for each observer, so as to match reliabilities between vision and haptics. In Figure 3.4, the red circles indicate the matched reliabilities for the 1.6:1 tool, and the blue circles for the 1:1 tool.

Figure 3.5 shows the mean of size-discrimination performance (JNDs) in each tool condition, averaged across seven observers. The light gray bars represent the single-modality performance, derived from Figure 3.4. Note that we matched vision-alone and haptics-alone performance, so there is only one bar for the two single cue JNDs. This single-cue performance was used to compute a predicted optimal integration if both signals were fully integrated using Equation (1.8)). The dark bars represent the predicted optimal integration. The coloured bars represent observed vision-plus-haptics performance; blue bars for when using the 1:1 tool and the red bars for when using the 1.6:1 tool. Individual performance can be seen in Appendix 5.

Note here that we used repeated-measures t-test to analyse the data. In this experiment, factorial analysis was inappropriate because the design was incomplete (see Table 3.1). As described in the method section, we chose four conditions under the restriction of that the size at hand was 50mm and the cue reliabilities between visual and haptic sizes were carefully matched at object coordinates by manipulating visual displacement noises. Thus, we could not properly analyse the interaction

between object sizes and tool gains or between tool gains and no-conflict/conflict conditions.

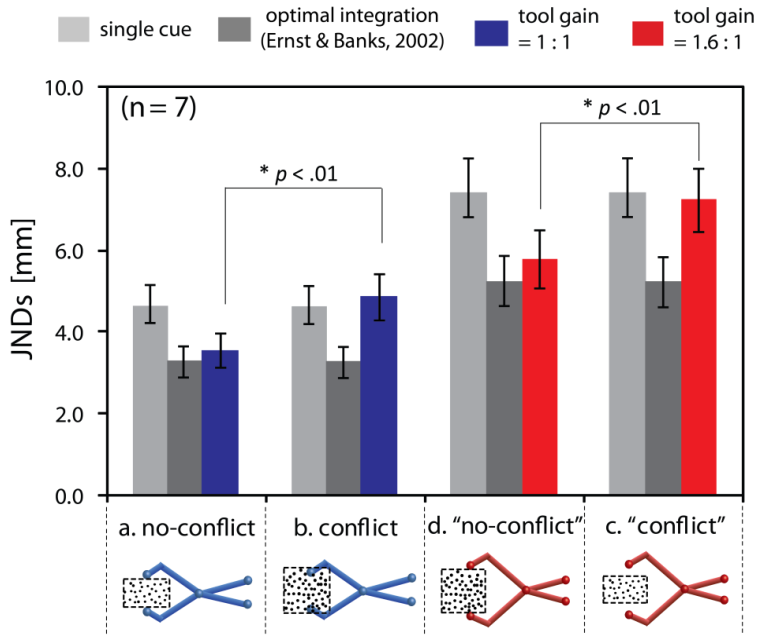


Figure 3.5. Results of discrimination performance in each tool condition in Experiment 5. Mean size-discrimination performance (JND) is plotted, averaged over seven observers. The light grey bar represents single-cue performance, and the dark grey bar represents predicted performance under optimal integration, calculated from the single-cue JND, using Equation (1.8). We used stimuli with matched reliabilities between vision and haptics, so single-cue performance for both vision-alone and haptics-alone is represented by the grey bar. The red and blue bars show the measured JNDs for the 1.6:1 tool and the 1:1 tool respectively. The a, b, c, d, stand for the four conditions in Figure 3.1. Error bars denote ±1 standard error.

Table 3.1. Possible combinations: object (seen) size and haptic size with two different gain tools. We only chose four shaded conditions for this experiment.

Object (seen) size S_V	Haptic object (at tool-tip) S_H	1:1 tool		16:1 tool	
		tip – handle	(S_V - S_H)	tip – handle	(S_V - S_H)
50 mm	50 mm	50 mm – 50 mm	(no-conflict)	50 mm – 32 mm	(no-conflict)
	80 mm	80 mm – 80 mm	(conflict)	80 mm – 50 mm	(conflict)
80 mm	50 mm	50 mm – 50 mm	(conflict)	50 mm – 32 mm	(conflict)
	80 mm	80 mm – 80 mm	(no-conflict)	80 mm – 50 mm	(no-conflict)

When using the 1:1 tool (blue), in the no-conflict condition, observed visual-haptic discrimination performance was better than single-cue performance, and was close to the optimal prediction. In the conflict condition, the performance was at the level of single-modality performance, suggesting that visual and haptic signals were not integrated. These two performances were significantly different ($t(6) = 5.21$, $p < .01$).

When using the 1.6:1 tool (red), discrimination performance showed a similar pattern to the 1:1 tool. In the "no-conflict" condition, the discrimination performance was again close to the prediction of optimal cue integration, even though the raw sensory signals (visual size and hand opening) were different. In the "conflict" condition, the performance was poorer and was close to the single-modality level, even though in this case the raw sensory signals were not in conflict. These two performances were also significantly different ($t(6) = 4.82$, $p < .01$).

Regardless of the types of tool, in the no-conflict conditions, the discrimination performance was close to the optimal cue-integration predictions. When using the 1.6:1 tool, although visual object size and hand opening were conflicting, haptic size was rescaled from hand coordinates to the object coordinates according to the tool gain. Therefore, the rescaled haptic size matched the visual size. In contrast, in the conflict conditions, the performance was at the level of single-cue performance; so visual and haptic signals were not integrated. When using the 1.6:1 tool, although visual object size and hand opening were matched, visual size and rescaled haptic size were conflicting in the object coordinates. Thus, these results suggest that size information from vision and haptics were appropriately combined when they refer to the same object, independent of conflicts in raw sensory signals (the visual size and the hand opening). The brain seems to solve the correspondence problem in the object coordinates, taking into account changes in the relative magnitude of visual and haptic signals introduced by tools.

Individual differences in visual-haptic integration during tool use

The above analysis describes the average results across all observers. We did, however, observe some individual differences in the pattern of the integration performance, including one observer who integrated cues across all conditions. Figure 3.6 shows the performance of three example observers. One observer (SE) clearly showed visual-haptic integration in no-conflict conditions, and no integration in conflict conditions. Thus, SE's performance showed optimal integration when using an appropriate tool gain. In contrast, another observer (YY) showed near optimal integration for all four conditions, regardless of no-conflict or conflict conditions. The pattern of JB's performance seemed to be between SE and YY.

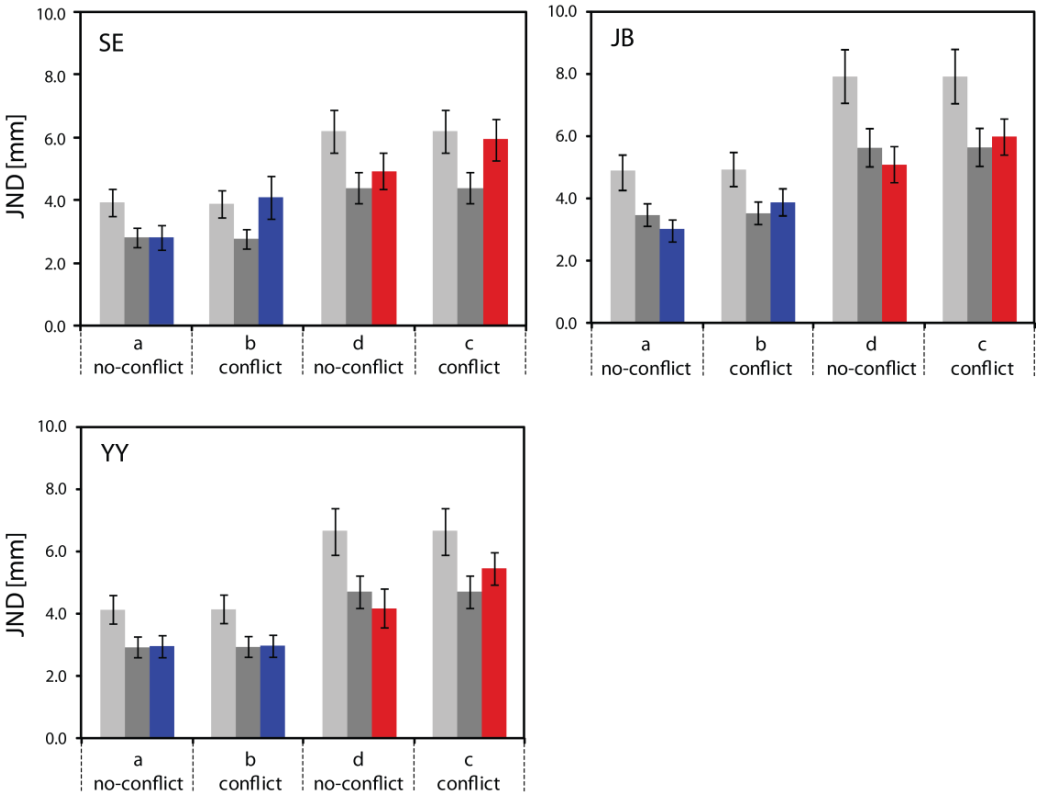


Figure 3.6. Three example performances. Observers (SE, JB, and YY) were chosen from 7 observers. All denotes are similar to Figure 3.5. The bars are presented in the same order as Figure 3.5: size discrimination performance for single-cue (light grey), predicted optimal integration (dark grey), and vision-plus-haptics with tool (colour coded by tool type). The a, b, c, d, stand for the four conditions in Figure 3.1. Error bars denote ±1 standard error.

3.3. Control Experiment 2: Effect of conflict on visual-haptic integration

We showed that visual-haptic integration was consistent with the appropriate tool gain, but there were some individual performance differences (See Figure 3.6 and also Appendix 5). Here we asked whether the pattern was due to individual differences in tool use, or individual differences in integrating cue conflicts in general. To examine this, we selected three observers who showed differences in performance (SE, JB, YY) and measured their vision-plus-haptics performance with increasing size conflicts between vision and haptics. We examined the effects of conflict *per se*, during normal grasping.

3.3.1. Methods

Observers

The three observers from Experiment 4 took part in this control experiment. Their data in tool use were shown in Figure 3.6.

Stimuli & Procedure

We measured size-discrimination performance without a tool by manipulating the conflict ratios between visual and haptic sizes. The stimuli were the same as those used previously in Experiment 4. The haptic size was 50mm, and the visual size was varied as 50, 60, 80, and 100 mm; i.e. size ratio [vision : haptics] ranged 1:1, 1.2:1, 1.6:1 and 2.0:1. Based on the observed individual vision-alone data at 50mm and 80mm in Figure 3.6, we predicted visual noise levels for 60mm and 100mm mm in order to match reliabilities between vision and haptics (50mm), assuming that simple Weber's law is preserved for visual sensitivity at these sizes. The procedure was the same as Experiment 4, with the exception of removing the tool.

3.3.2. Results & Discussion

Figure 3.7.a shows two-cue discrimination performance as a function of the size ratio between visual and haptic stimuli. The three observers showed slightly different patterns of JNDs with increasing conflict ratio. Both SE and JB's performance reduced with increasing conflict ratio. When the conflict ratio was 1.6, SE showed vision-plus-haptics performance close to single cue performance; JB showed that visual-haptic performance between single alone and predicted optimal integration. In contrast, the observer YY integrated visual and haptic size information near optimally for the wide range of the conflict ratios. This was similar to YY's performance across the four tool-use conditions, suggesting that YY's tool-use performance in Experiment 4 reflected a general insensitivity to conflict, rather than being particular to tool use. YY's performance trend might be affected by the experimental setting; for example, visual and haptic stimuli were presented in the virtual space at the same time regardless of the conflicting in size. That is, the presentation timing was perfectly correlated between two signals. YY might have used a different strategy (timing) to solve the correspondence problem between signals rather than the spatial discrepancies. To explore this further, we estimated the relationships between conflict no-tool and tool conditions.

We predicted the performance in four conditions when using a tool, based on the no-tool performance. Figure 3.7 (b) shows the predicted performance, compared with the observed performance with tool-use in Experiment 4. The prediction was calculated from observed no-tool data, by considering that (i) visual and haptic correspondence would be solved in the object coordinates and (ii) sensitivity at hand space would be rescaled to the object space according to the tool gain. Even though performance was slightly worse when using a tool in the previous experiments, here we assumed that tool-use noise did not significantly influence size judgement. As shown in Figure 3.7 (b), similar patterns between predicted and observed data were found in four conditions for all three observers. There were no statistically significant differences between the predicted performance and the observed performances in Figure 3.7 (b).

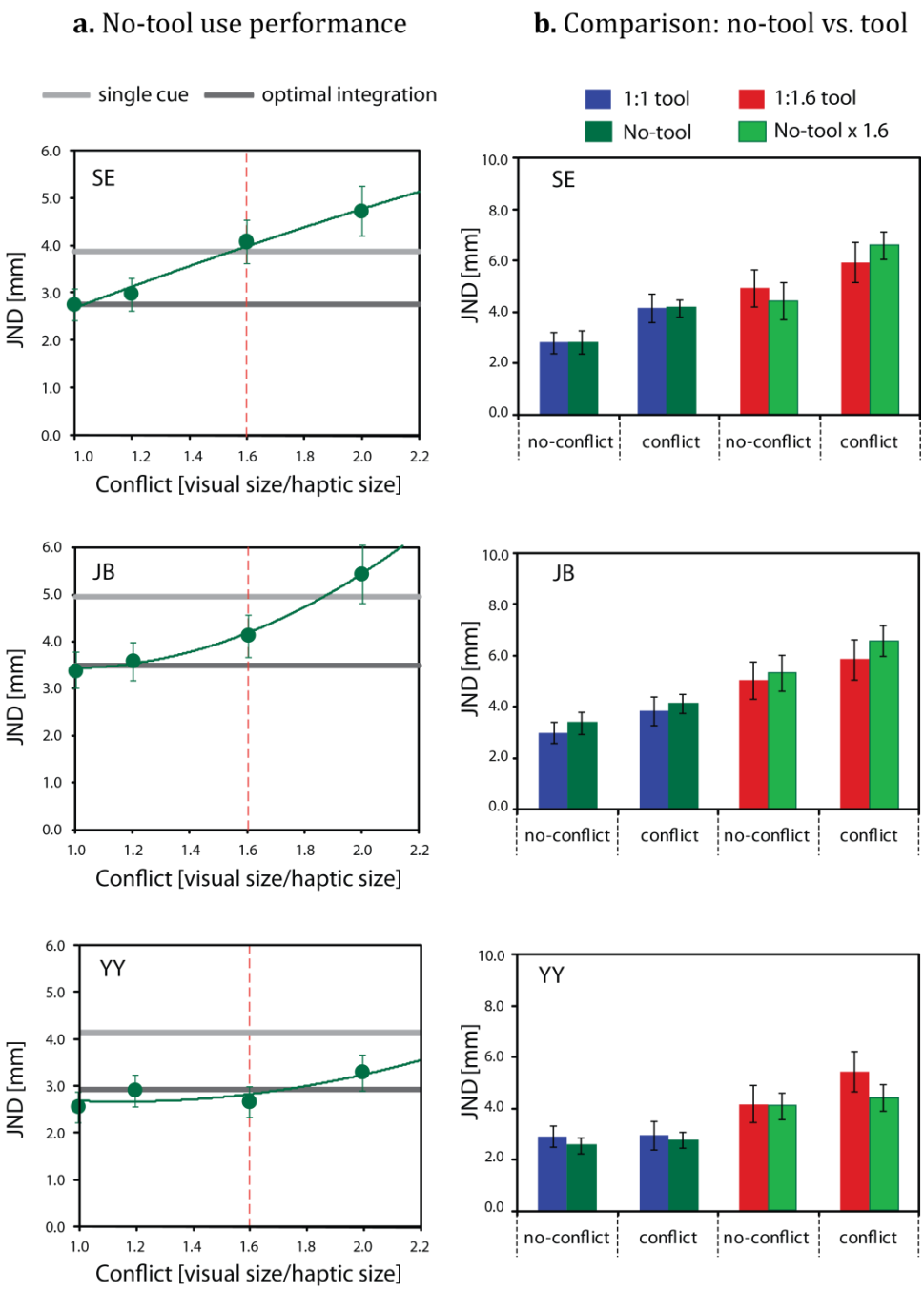


Figure 3.7. The effect of conflict between visual and haptic sizes on cue integration. Individual differences for three observers [SE, JB, YY]. **a.** Size discrimination performance when no-tool use with increasing degree of conflict sizes. The conflict ratio was varied from1:1 to 2:1. **b.** Comparison with the observed performance in Experiment 4 and predicted performance from no-tool use performance. Error bars denote ±1 standard error.

These data show that the individual performance differences in tool use resembled the performance in no-tool use. The sensitivity to conflicting stimuli was similar when combining the size information in both tool-use and no-tool use, suggesting both were dealing with object coordinates. Thus, the correspondence problem may be always solved in object coordinates, regardless of normal grasping and tool-use.

3.4. Discussion

3.4.1. Chapter Summary

We investigated visual-haptic integration when using a pair of pliers which altered the gain between the tool-tip and the hand opening. The results showed that visual and haptic information was integrated near optimally in size-discrimination judgements when the object and tool-tip sizes were spatially coincident. Visual-haptic integration therefore occurred not based on the similarity between the visual size and hand opening, but instead based on the similarity between the visual size and haptic size from the tool-tip. Moreover, individual performance differences in the integration during tool use reflected the effect of conflict between visual and haptic sizes on integration in normal grasping. In contrast to normal grasping, when using a tool, even if the haptic size at the hand does not reflect the true size of the grasped object, a statistically optimal tool user would integrate these signals. Because two sensory estimates refer to the same object, the brain should integrate visual and haptic size estimates even though raw sensory signals are conflicting.

Importantly, our results suggest that haptic estimates were correctly remapped onto object coordinates from hand coordinates, taking into account tool geometry. This remapping idea is consistent with the findings in Experiments 2-3, where a simple tool introduced a spatial remapping of the offset between the locations of visual and haptic signals. In Experiment 4 the pliers introduced not only a spatial remapping of the offset but also a required rescaling of the size between the tool-tip

and the handle. The brain therefore appeared to treat haptic signals as originating from the tool-tip, independent of the spatial separation between object and hand positions, and conflict in object size and hand opening.

Here, we restate our fundamental questions: How does the brain solve the correspondence problem to integrate visual and haptic information during tool use? To investigate the answer, we model our results to explore the mechanism of spatial remapping in tool use.

3.4.2. A computational model for solving the correspondence problem

As introduced in Chapter 1, Bayes' rule deals with uncertainty and the probability of different states of the world. The process of multisensory integration can be described in Bayesian framework (Bresciani et al., 2006; Ernst, 2007; Knill, 2007a, 2007b; Knill & Pouget, 2004; Körding et al., 2007a; Shams et al., 2005). Bayesian estimation is similar to the MLE model, but the key difference is the use of the prior. The final percept can be determined by the interaction between likelihood functions and the prior distributions. The important aspect is that use the prior alters the posterior by according for the knowledge of the world. There are different priors when perceiving the world and the prior information might influence on different levels of sensory processing on cue combinations before obtaining the final estimates (Ernst & Bülthoff, 2004).

Based on Bayesian decision theory, several models that consider the similarity of signals in terms of their spatial coincidence and magnitude in order to solve the correspondence problem have been proposed: the coupling prior model (Bresciani et al., 2006; Ernst, 2005, 2007; Roach et al., 2006), the causal inference model (Körding et al., 2007a; Sato et al., 2007), and the mixture model (Girshick & Banks, 2009; Knill, 2003; Natarajan, Murray, Shams, & Zemel, 2009). A decision in cue integration has to be made on causality of sensory information, rather than simply on raw sensory signals *per se* (Ernst, 2005, 2007; Körding et al., 2007a; Roach et al., 2006; Shames et al., 2005; Beierholm, Körding, Shams, & Ma, 2008). Here, two key

concepts in relation to multisensory cue integration and the correspondence problem should be highlighted: a coupling prior and causal inference.

Coupling prior model

To mathematically explain causal inference, a coupling prior has been introduced by Ernst (2005, 2007). Figure 3.8 illustrates the coupling prior model (Ernst, 2007). When multiple cues are available to provide information about a physical property, the raw sensory signals are distributed around the physical values plus sensory noises. Consider estimating an object's size (s) when visual and haptic cues are available, Equation (1.2) can be extended to

$$p(s|x_V, s_H) \propto p(x_V, x_H|s)p(s) \sim p(x_V|s)p(x_H|s)p(s) \qquad (3.1)$$

As described in the Introduction, x_V represents the visual information and x_H represents the haptic information; both are available to the observer. $p(x_V, x_H|s)$ is called the combined, or "joint", likelihood function. If sensory noises are unbiased and independent for each sense, this can be $p(x_V|s)p(x_H|s)$, which represent a likelihood function of each cue: vision and haptics respectively. Each likelihood distribution provides the probability of object size estimates in relation to the reliability of the current sensory information. The joint likelihood distribution would be a two-dimensional (2-D) Gaussian with the mean (s_V, s_H) and standard deviations (σ_V, σ_H). Here $p(s)$ is the probability of the past sensory experience of the object size and this is independent of both the current sensory cues (Hillis et al., 2004; Mamassian et al., 2002). The posterior distribution $p(s|x_V, x_H)$ provides the probability of the object size, s, given the sensory information available to the observer from vision and haptics interacting with the prior constraints.

When applying the Bayes decision rule, a unified estimate can be determined through a process which chooses the peak position of the posterior distribution using the MAP rule. To simulate this process, a delta gain function can be used for maximising benefit or minimising loss in order to make a decision (Mamassian et al., 2002; Roach et al., 2006; Knill, 2007a; Knill & Richard, 1996). The delta gain

function mathematically has the fundamental properties which is the limit of the sequence of Gaussians and is infinitely high within a narrow window around the correct physical value. Bayesian decision theory can predict optimal outcomes of an estimate by considering such costs and benefit of sensory integration (Körding & Wolpert, 2006; Trommershäuser, Maloney, & Landy, 2003, 2008).

In terms of solving the correspondence problem, the prior can be introduced as a representation of the statistical relationship between sensory signals. When the prior is weak to interact with the likelihood distributions, this is termed as a coupling prior (Ernst, 2005, 2007). In this case, the coupling prior distribution indicates the degree of the statistical similarity, or correlation, of the mapping between visual and haptic sensory estimates. The Bayesian prior is generally considered as independent from the likelihood functions (Beierholm, Quartz, & Shams, 2009; Ernst, 2007; Knill & Pouget, 2004; Shams & Beierholm, 2010), so the coupling prior is as well. In the real situation, the coupling prior would be distributed between no correlation and perfect correlation with uncertainty of the signal causes. The coupling prior interacts with the joint likelihood given by the sensory estimates and produces the posterior distribution; therefore the degree of the correlation affects the degree of sensory combination (See Figure 3.8).

Because the coupling prior distribution represents the brain's knowledge of degree of correlation (σ_p) between visual and haptic estimates, it could indicate how it is likely that signals originate from the same object. When visual and haptic signals originate from two independent causes and there is no correlation between them, the prior can be assumed as a uniform distribution. The left column in Figure 3.8 considers a case, when any combination of two sensory estimates is equally likely. Here, the coupling prior would be a flat, or uniform, distribution and there is no information about whether signals originate from the same source. In this case, the posterior distribution, as a product of the likelihood and the prior, would therefore be unaffected by the prior and there is no interaction between vision and haptics; therefore, this results in no sensory integration.

In contrast, when visual and haptic signals originate from a common cause, two sensory signals should be strongly correlated: a diagonal ridgeline in the 2-D

prior distribution, at the right column in Figure 3.8. In the ideal case, a delta gain function can be used for the prior distribution. The delta function means that probability density is infinite when the visual signal is equal to the haptic signal. When sensory signals are perfectly correlated, the coupling prior would be a delta function, vision and haptics would be completely combined. The MAP estimate can be employed from the posterior distribution. Two sensory signals are integrated optimally, where the MAP rule would be comparable to the MLE rule. In this sense, the MLE model assumes that signals come from the same object.

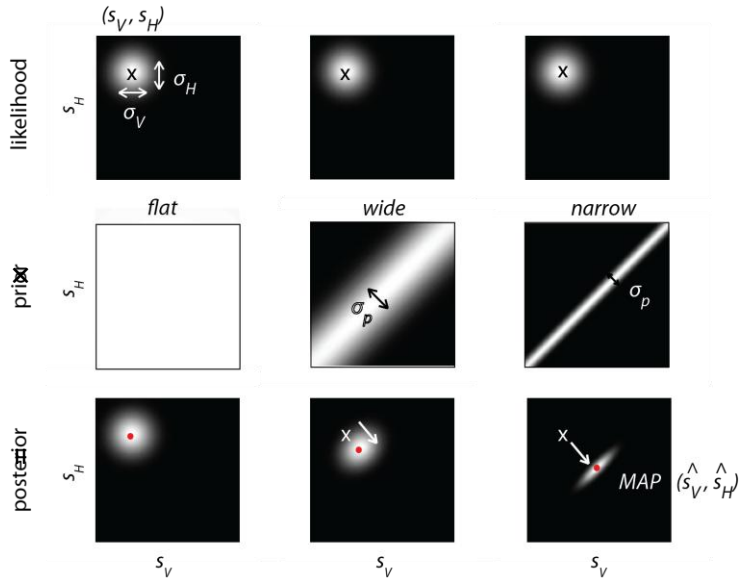


Figure 3.8. A coupling prior model for combining vision and haptics. The figure illustrates how the prior distribution influences the posterior distribution (product of the prior and the likelihood). The joint likelihood distributions are 2-D Gaussian distributions. Here three different priors: flat, wide and narrow are used. When the prior is flat and all combinations of (s_V, s_H) are equally likely, it has no influence on the posterior distribution. When the prior is narrow, the mean of the posterior is shifted from the likelihood toward the prior distribution. The shift is continuous from flat prior (no combination) to a delta function (complete integration). x represents physical stimulus (e.g. size) and · indicates the MAP estimate. The schematic diagram is adapted from Ernst (2007). Notes: colour map indicates relative probability density: white colour is high probability and black colour is low probability.

The degree of cue integration therefore can be derived from the posterior distribution as the product of the likelihood and prior. The posterior distribution could be varied from no-combination to complete integration, according to the correlation between cues. When the variance of the prior is varied from wide to narrow, the max probability of the posterior distribution gradually shifts closer to the peak of the prior (Figure 3.8). This shifting effect depends on the degree of prior knowledge of the correlation between signals. The coupling prior model has provided good predictions of cue combination, consistent with many aspects of multisensory processing. In principle, these computational approaches can be carried out over any number of relevant dimensions (Ernst, 2007), including the time of occurrence of signals and other parameters of interest. Ernst (2007) demonstrated that a new relationship between arbitrary signals such as luminance and stiffness can be learned and explained, in terms of the model, as learning the coupling prior.

Causal inference model

Körding et al. (2007a) investigated audio-visual combinations by manipulating the spatial separation between the two cues. They measured observers' location judgement performance and they showed how automatically the source of the stimuli is identified by either a common cause or two different causes, depending on spatial separation. To deal with the causal inference of two sensory signals more explicitly, Körding et al. (2007a) introduced an extra decision stage to a coupling prior model similar to Ernst's (2007). To do this, they consider two prior distributions, representing "a common cause" or "two independent causes", producing two posterior distributions. The posterior distribution indicates whether 'a common cause' or 'different causes' of the signals were most likely, and the most likely state is used to decide whether to integrate or not, thus solving the correspondence problem (Körding et al, 2007a; Beierholm et al., 2009).

Figure 3.9 illustrates to explain the concept of the causal inference. When sensory estimates are likely to occur together, the brain could infer that they originate from a common cause, so the decision could be made to combine them. However, the

estimates should not be combined if the brain infers that they originate from two independent causes. The prior probability of each cause can be expressed: $p(C = 1)$ and $p(C = 2)$. Considering two cues:(s_V, s_H), $p(C = 1|s_V, s_H) + p(C = 2|s_V, s_H) = 1$. The decision would result in the posterior distribution. Comparing the peaks of two hypotheses to each other; if $p(C = 1|s_V, s_H) > p(C = 2|s_V, s_H)$, it would be a common cause, and if $p(C = 2|s_V, s_H) > p(C = 1|s_V, s_H)$, it would be two independent causes.

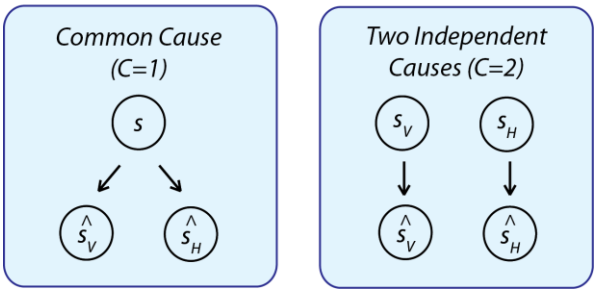


Figure 3.9. A cartoon of a causal inference model in cue integration. When observing the physical property (s) in the world (which gives rise to visual and haptic sensory estimates ($\hat{S}_V, \hat{S}_H$)), a decision of combining cues could be made based on the probability of "Which is most likely: common cause or two independent causes?" If S_V *and* S_H are caused by different objects (*C=2*), the system should not combine them, and if they are caused by the same object (*C=1*), it should. Schematic diagrams are adapted from Körding et al. (2007).

Körding et al. (2007a) successfully predicted the degree of audio-visual integration in direction judgements using this computational model. It should be highlighted again that this process does not result in a binary output (integrate or do not integrate), but the output would vary depending on the probability that the signals have a common cause or two independent causes. This model could account for how the correspondence problem is solved in multisensory processing.

Mixture models

Although coupling prior and causal inference models provide an appealing explanation of how the brain solves the visual-haptic correspondence problem, there are much simpler models. One of the alternative models is a mixture model or heavy-tailed Gaussian model, which considers additive and non-linear interactions between the likelihood and prior distributions (Knill, 2003, 2007a, 2007b). In the relation to the coupling prior model, the mixture model introduce an idea that the prior can be formed as a sum of a uniform distribution (independent causes) and a prior distribution with a diagonal ridge (common cause) with a certain degree of uncertainty (Wei & Körding, 2011). This mixture distribution might be heavy tailed (Knill, 2003; Girshick & Banks, 2009). In this model, the final estimate could be simply decided by choosing the most probable state in the posterior distribution implementing MAP estimates. In contrast to the above coupling prior explanations, the decision phase for the binary choice of the correspondence is not necessary to determine the final percept in the mixture model. As such, the mixture model would be attractive because it could automatically produce the most probable final percept without the explicit decision phase. Therefore this mixed prior distribution would provide more parsimonious and intuitive accounts for multisensory integration.

However, the research showed that in a direction judgement of vision and auditory cues, observers consciously perceive whether signals come from the same source or two different sources (e.g. Körding, et al., 2007a; Shames et al., 2005). The result suggests that the nervous system would somehow solve the correspondence problem with an explicit decision phase. This explicit response to the signal correspondence cannot be explained by a general mixture model. Thus the decision phase might be clearly involved in solving the correspondence problem for not only vision-auditory but also vision-haptic processes. The coupling prior and causal inference models explicitly set a decision phase by introducing a concept of solving a correspondence problem; therefore, the model would be conceptually preferable than the mixture model from this point of view. On the other hand, the mixture model can simply choose the final estimate from the posterior distribution, but this does not mean that there is no explicit process in relation to the correspondence problem in the

model. Further experiments would be useful to clarify this possibility and to explore the appropriate model in tool use.

3.4.3. Modelling of normal grasping and tool use

Normal grasping

Here we discuss visual-haptic integration during normal grasping using the model, as described above. Figure 3.10 shows the process in a similar way to the process outlined in Figure 3.9, adding an extra stage: the decision of whether two signals should be combined or not. The probability distributions of raw sensory signals are plotted in arbitrary coordinates, and the joint likelihood of vision and haptics share common coordinates. To make the decision of cue combination (the correspondence problem), the coupling prior represents two statistical possibilities: independent causes (flat prior) and common cause (correlated). If signals originate from two different objects, a coupling prior would be a flat prior. If they originate from a common object, there would be a strongly correlated distribution between visual and haptic estimates. Compared with the two posterior distributions, the decision would be made by choosing the most likely one.

Figure 3.10 shows the extent to which haptic and visual size estimates are similar relating to the posterior distribution of sensory estimates. The Figures illustrate how the decision to combine cues in the final percept could be made by comparing two hypotheses of the causal structure: the common cause ($C = 1$)or two independent causes ($C = 2$), in a similar manner with Körding et al. (2007a).

Figure 3.10 (a) shows that when visual and haptic size estimates are matched, the posterior distribution derived from a common object would show higher probability than from two different objects. The decision would be made that signals originate from a common object, and so size information would be combined. The outcome would be a combined estimate ($\hat{S}_{VH}$). In contrast, Figure 3.10 (b) shows that when visual and haptic size estimates are conflicting, the posterior distribution derived from two different objects would show higher probability than from a

common object. The decision would be made therefore that the signals came from different objects and size information should not be combined. The outcomes would be two estimates ($\hat{S}_V, \hat{S}_H$).

These processes would state that sensory signals giving rise to similar estimates of the some property (we explained about "size" as an example, in Figure 3.10 but this could also apply to spatial location) are most likely to originate from the same object, and signals giving rise to different estimates of a property are highly unlikely to come from the same object. When the probability of a common cause decreases, signals are less integrated. As such, this model demonstrates a good account of a gradual transition from complete integration, through partial integration, to no integration of signals, as sensory estimates become increasingly dissimilar (e.g. Körding et al., 2007a; Roach et al., 2006). Such a progressive reduction in cross-modal integration with increasing spatial offset between two sensory stimuli was empirically observed in visual-haptic integration (Gepshtein et al., 2005, Experiment 1 in our studies) and visual-auditory integration (Körding et al., 2007a; Meyer, Wuerger, Röhrbein, & Zetzsche, 2005).

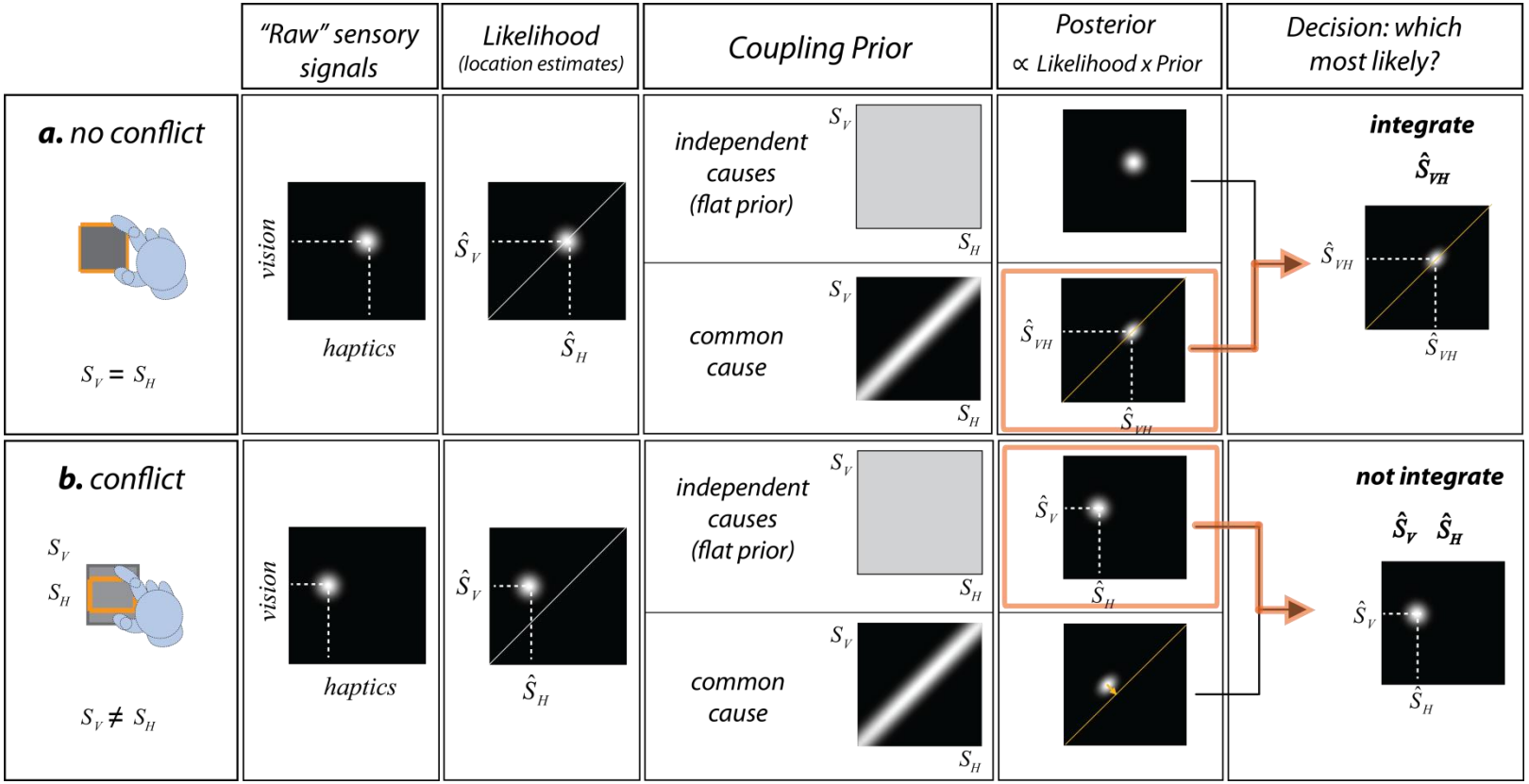


Figure 3.10. Bayesian model for visual-haptic correspondence problem (for size estimates). The raw sensory signals are normally on arbitrary coordinates but the joint likelihood share coordinates. The coupling prior represents two possibilities: independent causes (flat prior) and common cause (correlated). The posterior distribution is the product of the joint likelihood and the coupling prior. The decision would be made based on the most likely state in the posterior distribution (integrate or not integrate). **a.** normal grasping (no tool). **b.** normal grasping with a conflict between visual and haptic size estimates.

Tool use

Our results showed that the correspondence problem in visual-haptic integration during tool use appeared not to be solved on a spatial mapping of raw sensory stimuli, but was instead solved on a spatial remapping of the inferred source of those stimuli. The results can be explained by a probabilistic idea that the brain infers the causal structure of the two signals and integrates them only when it is appropriate to do so, as described above. It only makes sense to integrate these cues, if the brain knows the relationship between the sensory signals. We can therefore hypothesize that the mechanism is flexible enough to compute that, when using a tool, different signals can refer to the same object even when visual and raw haptic signals are conflicting, for example, in size. The decision to integrate or not would be made by the distal causes of sensory stimuli, not simply by raw signals. If a process put visual and haptic signals in common coordinates, the causal inference model could simply solve the correspondence problem in tool use. Rescaling of the sensory estimates from haptics in tool use enables the signals to be put in common object space.

Figure 3.11 illustrates the calculation based on this idea. Consider a tool that magnifies the size at the handle to the tool-tip. When pinching an object using such a tool, visual object and hand opening sizes are different. We presume an optimal tool user so that the haptic estimate is perfectly rescaled at the likelihood level, according to tool geometry, and its reliability (variance) is not changed by the rescaling. The model then operates in a similar manner as normal grasping. When tool-tips do not reach the visual object surface, the posterior distribution indicates that the possibility is much higher of two different objects and a decision would be made not to integrate (Figure 3.11 (a)). In contrast, when the tool-tips match the visual object size, the probability they refer to the same object would be higher than two different objects, and the decision would be made to integrate (Figure 3.11 (b)).

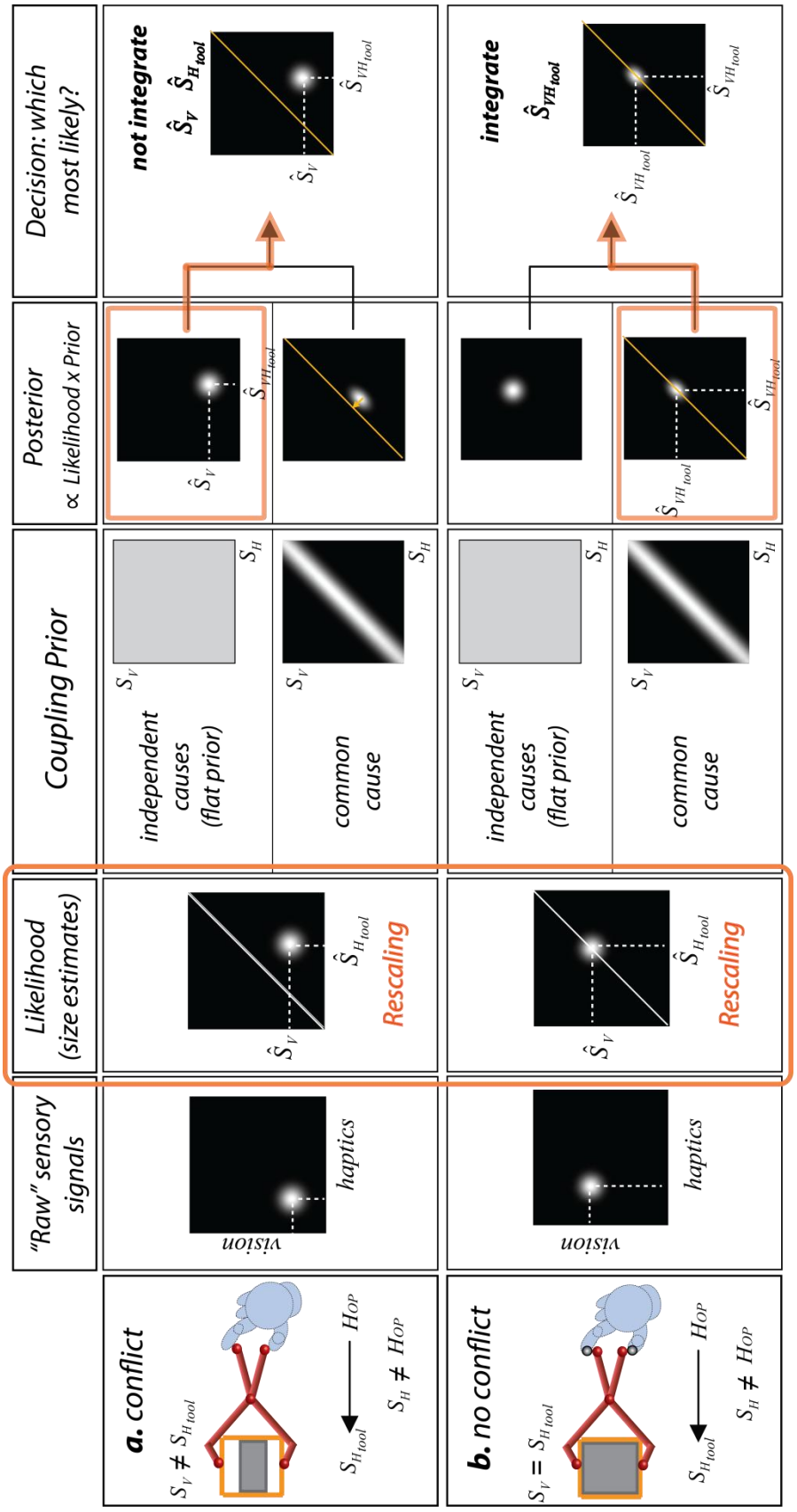


Figure 3.11. Bayesian model for visual-haptic correspondence problem in tool use. Coupling prior represents two possibilities: independent causes (flat prior) and common cause (correlated). Posterior distribution is product of the likelihood and prior. Decision would be made based on the posterior distribution (integrate or no-integrate). **a.** conflict object space. **b.** no conflict in object space.

We here explained about the magnitude case but the same analogy can be applied to explain the spatial offset case: Experiments 2-3. (Note: in this case, the tasks were size judgements, so not directly related to the spatial location; however, the location information was used for solving the correspondence problem, in the similar principle.) When judgement is made that signals come from two different objects, one size estimate would be chosen either chance level or biased vision or haptics, this would be close to single cue performance. However, if tool use increases uncertainty of the correspondence and the decision noise, the performance would be worse than single cue performance. This would provide an explanation of the obtained results in Experiment 4, where some observers showed larger JNDs than single cues in conflict conditions.

The coupling prior distribution would be affected by the degree of knowledge about a tool and also the correlation between two sensory signals introduced by the tool geometry. That is, tool use might change the scale or unit of the coordinates of the prior distribution or the slope of diagonal ridge and/or its degree of correlation. In general, sensory integration does not always share a common unit or involve sensory calibration (Smeets, van den Dobbelsten, de Grave, van Beers, & Brenner, 2006). However, even if there are different units or conflicting between vision and haptics in likelihood level, a certain statistical assumption can have an effect on the final estimation and the information can be robustly integrated with the prior constraints (Ernst, 2005; Ernst & Bülthoff, 2004; Knill, 2007a). Tool use could be the case. Thus, whether sensory information is processed with (i) to treat the common units, or scale, introduced by the tool geometry in the likelihood level, or (ii) to employ some perceptual bias (possible arbitrary unit) in the coupling prior level, would be a possible argument.

However, the rescaling in the prior level is highly unlikely because the wrong estimates are used to create the final percept (readout problem). If the rescaling occurs in the prior level, we could never estimate size in tool use without vision; i.e. if haptic estimates hold regardless of the tool use, size estimates from haptics-alone would be always wrong. Clearly this is not the case and the possibility of rescaling in the prior level can be ruled out. It is arguably straightforward if the rescaling occurred

in haptic sensory estimates. If remapping occurs in the likelihood level, haptic size estimates would be rescaled before the integration stage and the decision to the final size estimate could be inferred correctly. It also makes the strong prediction that haptic size estimates will be altered by changing tool gain, even if the hand opening is held constant. We examine this rescaling of haptic estimates in the next chapter.

Chapter Four

Perceived object size during tool use

4.1. Introduction: Do haptic size estimates change when using a tool?

We have previously shown that during tool use the brain solves the correspondence problem correctly, independent of (i) spatial offset in visual and haptic signals (Experiment 2 and 3) and (ii) conflict between visual size and hand opening (Experiment 4). This could be achieved by considering that if haptic estimates were rescaled from the hand coordinates by the tool, visual-haptic integration would occur based on the similarity between visual and haptic size in the common object coordinates. However, we did not determine whether haptic size estimates were actually affected by the tool geometry in Experiment 4, because we only measured discrimination threshold, and not perceived size.

Here we examine whether haptic size estimates are rescaled, taking account of tool geometry. To solve the correspondence problem, the estimates of size from haptics need to be appropriately rescaled in accordance with the tool dynamics. We measure haptic size estimates and examine whether the size estimates were correctly read out at the object coordinates from the posterior distribution and how they were affected by the hand coordinates.

To consider measuring haptic estimates, here we restate the weighted linear summation to examine the data. Assuming that the noise in visual and haptic estimates are independent, and Gaussian distributed, and that estimates are on average unbiased, the combined estimate $\hat{S}_{VH}$ can be computed:

$$\hat{S}_{VH} = w_V \hat{S}_V + w_H \hat{S}_H, \qquad w_V + w_H = 1 \tag{4.1}$$

where individual cue weights (w_V, w_H) are assigned based on their reliabilities:

$$w_V = \frac{1/\sigma_V^2}{1/\sigma_V^2 + 1/\sigma_H^2}, \qquad w_H = \frac{1/\sigma_H^2}{1/\sigma_V^2 + 1/\sigma_H^2} \tag{4.2}$$

We manipulated physical properties: object size (S_V) and hand opening (H_{OP}) independently, and generated conflicting conditions $(\hat{S}_V \neq \hat{S}_H)$, then we measured combined estimates $(\hat{S}_{VH})$. Using the above equations, we could calculate their weights from the combined estimates in normal grasping, assuming vision-alone and haptics-alone estimates were unbiased: $(\hat{S}_V \sim S_V, \hat{S}_H \sim H_{OP})$. When using the 1:1 gain tool, the opening size at the tool-tip and handle are the same with the certain spatial separation of the tool length; therefore, we could expect that the perceived object size would not be significantly different from the normal grasping. In contrast, if the haptic estimates are perfectly rescaled from the hand opening to the object size introduced by the different gain tools, we could expect that the combined estimates would be the object size at the tool-tip, not at the handle.

In general, measuring PSEs using a psychophysical task enables us to examine the haptic estimate change precisely. However, when designing a 2-IFC task, in which comparing object size with normal grasping and tool use, we found, through several pilot experiments, that the task became very difficult in the size judgements. Because the tool disappeared in one interval, two intervals were quite similar and just increased uncertainty. Therefore, here we decided to measure perceived object size from haptics while using three different gain tools, applying a method of adjustment. Observers explored the virtual object size using the tool in vision-plus-haptics and haptics-alone conditions. We varied haptic and visual sizes independently to measure the individual cue's contribution to the final percept. Observers then matched a visual and haptic stimulus to the previously explored object size. The observed data were compared with the linear summation model. We demonstrated that haptic estimates were rescaled from the hand coordinates to the object coordinates, according to the tool gain.

4.2. Experiment 5: Perceived object size changes during tool use

4.2.1. Experimental Design

We examined whether haptic size estimates were rescaled from the hand opening according to the tool geometry, by measuring perceived object size. We designed the experiment using conflict stimuli with a matching paradigm. We directly measured perceived object size while using three different tools that (i) preserved the normal 1:1 mapping between visual size and hand opening, (ii) reduced the hand opening 0.7:1, and (iii) magnified it 1.4:1. Each value represents the magnitude between the tool-tip and the handle, which is normalised by hand opening size.

A no-tool condition served as a baseline measurement for the 1:1 gain tool in the vision-plus-haptics condition. That is, we could check whether performance would change between normal grasping and the 1:1 gain tool under the condition ($S_H = H_{OP}$). Moreover, in the haptics-alone tool-use condition we made direct measurements of perceived object size from haptics. We compared the perceived performance with different tool gain between vision-plus-haptics and haptics-alone conditions. We varied three different sizes of vision and haptics independently. However, we used an "incomplete" design, aiming to avoid large conflict between vision and haptic sizes because observers might easily notice such large conflict stimuli. We examined whether rescaling only occurred when combining visual and haptic information or occurred in both conditions. If perceived size changed according to the tool gain under the equal hand opening in the haptics- alone condition, this result could be considered as powerful evidence for the idea that rescaling occurs at the likelihood level.

4.2.2. Methods

Observers

Twelve right-handed observers (aged 19-45 years, 3 males) took part in this experiment. Among them, only one observer (YY) participated in the previous

Experiment 4. All observers had normal or corrected to normal vision with normal stereo acuity, and no known motor deficits. All were naïve to the purpose of the experiment.

Apparatus & Stimuli

The apparatus used was the same as in Experiments 1-4. Figure 4.1 shows the set of visual and haptic stimuli. The visual and haptic object sizes were chosen independently (from 42, 60, 84 mm). The visual objects were the random-dot stereogram stimuli and their parameters, such as dot density and the ground plane size, were set as in Experiment 4. Haptic stimuli were also presented at the fingers in a similar manner as before.

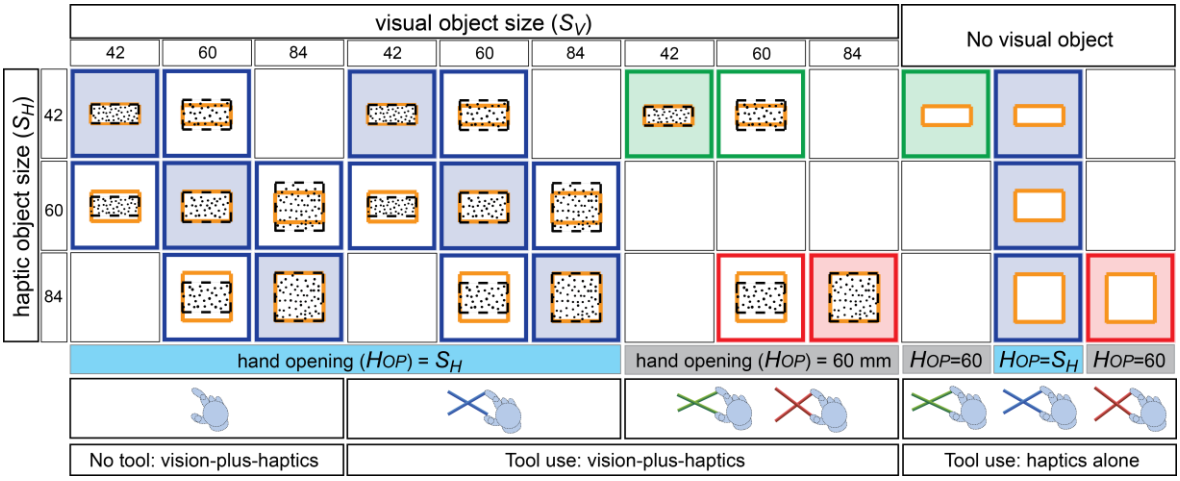


Figure 4.1. A set of visual and haptic stimuli. Visual object size S_V and haptic object size S_H were combination of 42, 60, and 84 mm. Squares coloured black with dots represent visual sizes and squares coloured orange represent haptic sizes. Stimuli were randomly assigned from the above combination. Observers explored object size under three conditions: (i) no-tool (vision-plus-haptics), (ii) tool-use (vision-plus-haptics), and (iii) tool-use (haptics-alone). The no-tool condition was equivalent to the 1:1 tool use in vision-plus-haptics condition, namely $H_{OP} = S_H$. Three types of tools were also randomly assigned on trial-by-trial basis. In the schematics, tool symbols represent three different gains (0.7:1, 1:1, 1.4:1) using different colours (green, blue, red) respectively, but simplified (the actual shape can be seen in Figure 4.2.)

Three different gain tools were used (Figure 4.2). Tools were visually defined pliers similar to Experiment 4. By moving the pivot position, we changed the size magnification between the tool-tip and at the handle. These tools were coloured differently: the blue tool had a 1:1 relationship between the tool-tip and the hand opening, the green tool reduced the hand opening with a 0.7:1 gain, and the red tool magnified it with a 1.4:1 gain. The tool length was constant (120mm) for the three tools. (In Experiment 4, we set the offset between the tool-tip and the handle for the standard stimuli as constant, rather than the tool length. In the current experiment, the set tool length meant the hand position was slightly moved depending on the tool geometry. However, based on our findings in Experiment 2, effects of such spatial offset could be ignored for perceiving the object size.)

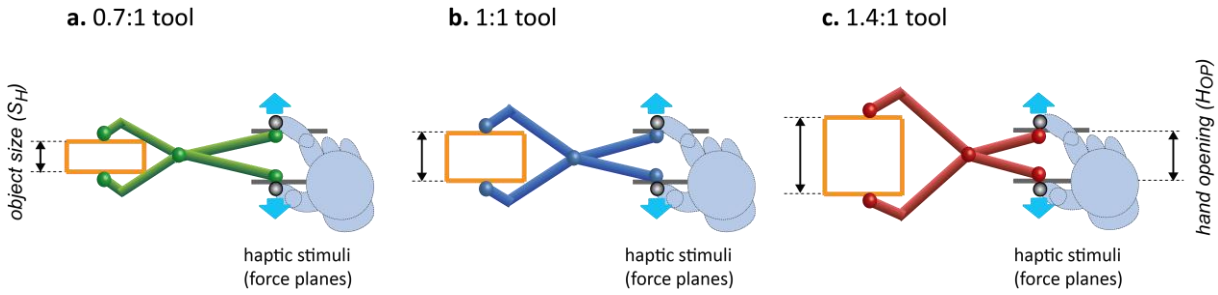


Figure 4.2. Three different gain tools. The relationship between the tool-tip and the handle was changed by moving the pivot position. **a.** 0.7:1 tool, **b.** 1:1 tool, and **c.** 1.4:1 tool. Haptic stimuli were generated at hand when the tool-tip touched the object. The light blue arrows represent the force. Haptic estimates of object size: S_H at object coordinate and H_{OP} at hand coordinate.

To avoid strong bias to a single modality (either visual or haptic dominance) in size estimation, we attempted to match visual and haptic reliabilities as much as possible. Ideally, visual and haptic reliabilities should have been matched for individual observers; however, considering the total experimental time, we decided to use the averaged data from the previous experiment. To do this, we calculated appropriate visual displacement noise for different object sizes from observed JNDs

as a function of visual added noise, which were averaged across 7 observers (See Figure 3.4). We added the averaged displacement on visual stimuli in the same way as Experiment 4. Under this manipulation, we can analyse the data, using the weighted linear summation model with unbiased estimators.

Procedure

Perceived sizes were measured under three different conditions: (i) no-tool (vision-plus-haptics), (ii) tool-use (vision-plus-haptics) and (iii) tool-use (haptics-alone). Each trial consisted of two periods. In the first period, observers perceived an object size, and then in the second period, they responded either visually or haptically to which size they perceived in the first period. Visual and haptic sizes (42, 60, 84mm) were varied independently and the combinations of these stimuli were randomly assigned from trial to trail.

As an initial setting of the workspace, we introduced a visual occluder, which was located at the left side of the workspace from the body midline. The size of the occluder was (width × height = 170mm × 200mm) and covered the whole tool when it reached the object. We used the occluder in order to minimise any effects of tool geometry on perceived object size. In addition, a white line was drawn along the screen edge and was the shape of a large rectangle. This was to facilitate fusion of left and right eye images into 3-D correctly, in the same manner as Experiment 4.

Firstly, observers perceived the virtual object size using visual and/or haptic stimuli. As in Experiment 4, visual and haptic stimuli were presented simultaneously in the vision-plus-haptics condition. In the no-tool condition (vision-plus-haptics), a starter cross indicated the location of the centre of stimuli. In contrast to the previous discrimination threshold experiments, no starter spheres were used in this experiment. The object was located behind an occluder (Figure 4.3). When the object was touched, the visual object popped up in front of the occluder for the duration of the grasp; so the object was only visible when the force planes were presented.

In the tool-use conditions, the three tools were randomly interleaved on a trial-by-trial basis. Observers pinched the object using the tool behind the occluder;

so even though observers knew which tool they were using, the tool was invisible when exploring the object size. In vision-plus-haptics with tool use, when the object was touched by the invisible tool, the visual object appeared in the same manner as no-tool condition; whereas, in the haptics-alone condition, the visual object did not appear. When the tool-tip touched the virtual object, force was generated at the hand in the same manner as the previous experiments.

We also introduced an occluder to avoid that visual tool geometry would directly affect the perceived size (Figure 4.3). Even though observers could not see the tool when perceiving the object size, observers should know which tool they were using. We coloured three gain tools differently and clearly informed the observers before the (visual) tool disappeared behind the occluder. Observers moved their hand toward the object position but they could not exactly know the tool-tip position behind the occluder. Therefore, we also generated a force plane at the moment one plane which vertically faced to the hand position was touched.

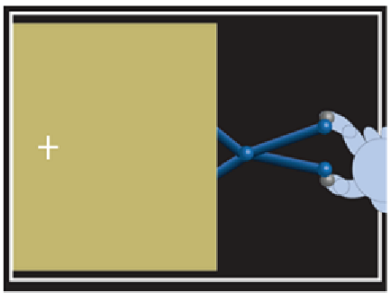


Figure 4.3. Occluder. Half of the screen was covered by a large rectangle. The object was located below the occluder and a white cross indicated the location of the object centre. The occluder was enough to cover the whole tool when the tool reached to the object. When the object was touched, the visual object was presented on the occluder.

In this first period, in contrast to the discrimination experiments, observers were allowed to explore the object size several times to perceive the object size within a certain period. Because, the discrimination experiments required observers to simply judge two object sizes and respond to which interval contained the bigger size. However, in this experiment, observers were asked to memorise one perceived object size and then to choose one matching size among the number of comparison stimuli.

In addition, we introduced a background force plane, to restrict finger movements on the plane. More precisely, the force-feedback devices limited the finger movements on the fronto-parallel plane to match the tool and the object orientations in the similar way to the Control Experiment 1. The additional force plane made it easier to set their finger orientations to match the tool orientation and this did not affect the performance itself at all.

Secondly, on each trial observers indicated the object size they perceived, either visually or haptically. There were two different response conditions: visual response and haptic response (See Figure 4.4). The sizes of the stimuli were varied from 34 to 98 mm with 4mm intervals, totalling 17 different sizes.

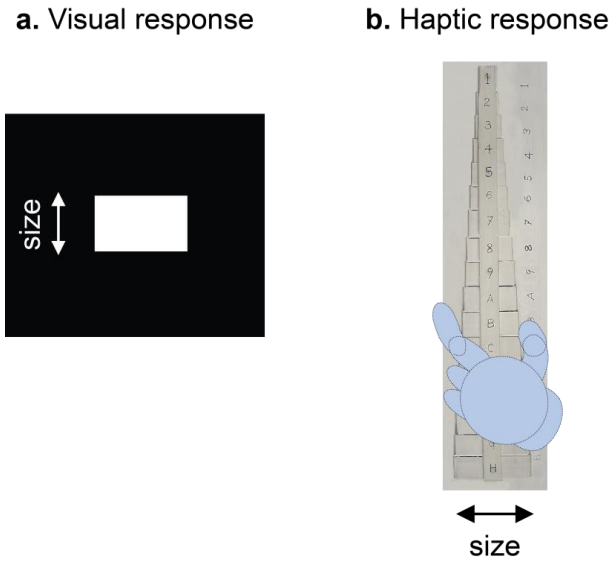


Figure 4.4. Visual and haptic responses. Observers were asked to respond to what the size of the object was. **a.** Visual response stimuli. Observers adjusted the size with key-presses. **b.** A set of haptic response stimuli. Observers chose the size from a set of wooden blocks felt by the left hand. Both stimuli were varied from 34 mm to 98 mm with 4 mm interval (total 17 sizes).

For the visual response, a grey coloured solid cuboid was presented on the screen (its width and depth were fixed as 100mm and 20mm respectively). The first size was randomly assigned from the set of sizes and then the size sequentially displayed by pressing a keyboard "up" and "down". Observers adjusted the visual stimulus size by making a button press on a keyboard and selected the appropriate size from the complete set of sizes. For the haptic response, a set of wooden blocks was provided (width and depth were fixed at approximately 30mm respectively). Observers grasped the set of wooden blocks with their left hand, and chose one from the set. The blocks were located behind a cover, so observers could see neither the blocks nor their hand. The experimenter recorded the observers' selection; there were no time limitations for the response.

The three conditions (no tool, tool use vision-plus-haptics, tool use haptics-alone) were blocked. A practical session in which observers learned the experimental procedure was followed by the main experiment. Half of the observers took part in the visual response first and the other half took part in the haptic response first. Stimuli and tools were randomly assigned on a trial-by-trial basis (See Figure 4.1).

4.2.3. Results & Discussion

Figure 4.5 shows perceived object size, averaged across twelve observers, in the no-tool condition. Half the observers made a visual response and then a haptic response; the other half conducted a haptic response first, and a visual response second. We did not find any response order effects on their performance, so all the data were analysed together.

We assumed that ideally perceived sizes in single modalities were equal to physical properties, or unbiased ($\hat{S}_V \sim S_V, \hat{S}_H \sim H_{OP}$). So, if combined perceived size ($\hat{S}_{VH}\ at\ S_V = H_{OP} = S_H$) was shifted from the physical size, we could consider the shift as response bias. Thus, we normalised the data by the individual performance at $S_V = S_H = H_{OP} = 60mm$. The diagonal dotted lines show that perceived sizes were perfectly matched to the visual object sizes regardless of haptic sizes. For no-conflict

stimuli ($S_V = S_H$), the perceived sizes were approximately on the diagonal dotted line in both visual and haptic responses.

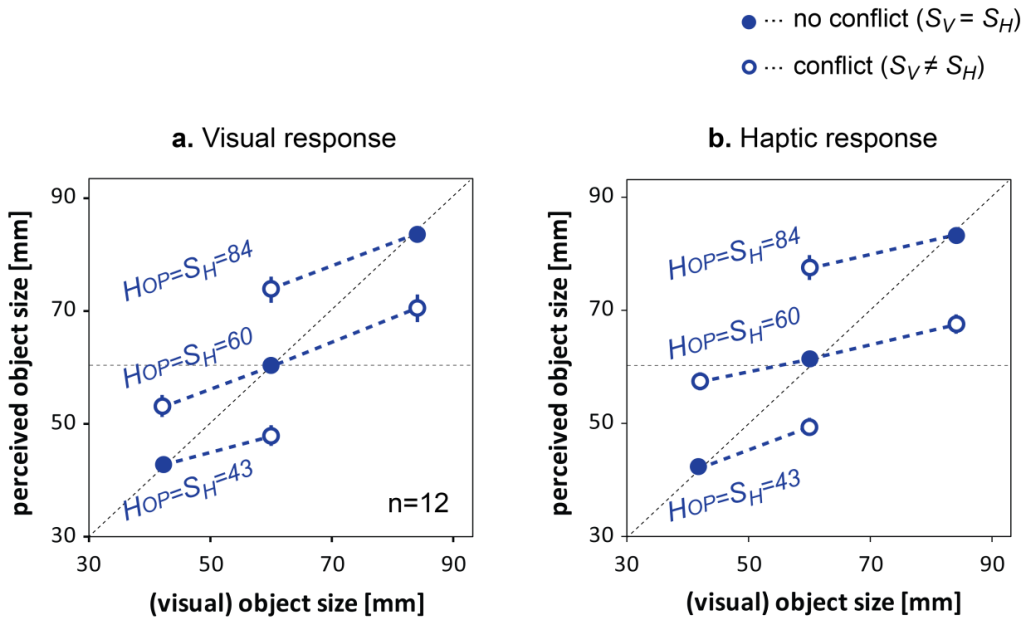


Figure 4.5. The perceived object sizes in normal grasping (no-tool condition, $H_{OP} = S_H$). Perceived object size with vision-plus-haptics, averaged across twelve observers. The dashed lines link the same size of hand opening. The filled circles represent no conflict conditions ($S_V = S_H$) and open circles represent conflict condition ($S_V \neq S_H$). **a.** visual responses **b.** haptic responses. Error bars denote ±1 standard error.

We set both cue contributions to the final estimate as similar as possible by adding the averaged visual displacement noise; so if both cue contributions were equal to the final estimates, the graph slope would be expected to be around 0.5. However, even though visual and haptic cue reliabilities were roughly matched by the manipulation, the graph slope was slightly deviated from 0.5.

The three dashed lines linked the data at the same haptic object sizes ($H_{OP} = S_H = 42, 60, 84\ mm$). The slopes of these dashed lines indicate single cue's

contributions $(\hat{S}_V, \hat{S}_H)$ to the combined perceived size $(\hat{S}_{VH})$, because the perceived size changed with variations in visual size although haptic size was constant. Therefore, the slopes clearly show that observers used both modalities (vision and haptics) to estimate the size (See Figure 4.5). Individual cue weights (w_V, w_H) were calculated from the individual graph plotting the relationship between object sizes and perceived object sizes using the Equation (4.1). Figure 4.6 showed visual and haptic weights, averaged across twelve observers. The observed performance was weighted more to haptics for visual responses $(\mathrm{w_V/w_H} = 0.37/0.63)$ and for haptic responses $(\mathrm{w_V/w_H} = 0.29/0.71)$.

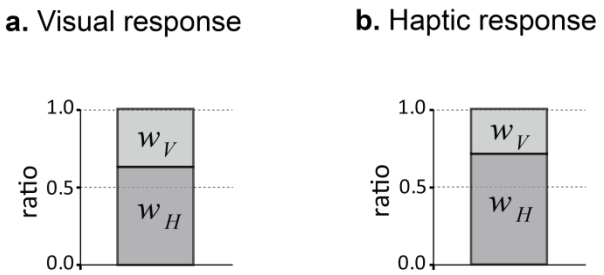


Figure 4.6. Visual and haptic weights, averaged across 12 observers. Weights were calculated from the slope in Figure 4.5 using Equation (4.1). **a.** visual responses **b.** haptic responses.

Figure 4.7 shows perceived object sizes when using the three different gain tools. The data were normalized by the individual performance in the no-tool condition at $S_V = S_H = H_{OP} = 60mm$. When using the 1:1 gain tool (blue), the hand opening and the tool-tip were the same; so, this was identical to the no-tool condition (excluding the visual-haptic spatial locations). The light blue lines linked the data when perceiving no-conflict stimuli $(S_V = S_H = H_{OP})$ using the 1:1 tool. Although the slopes were slightly tilted from the diagonal dotted line, the perceived sizes were

close to the diagonal in both visual and haptic responses. This means that the perceived size was matched to the actual object size.

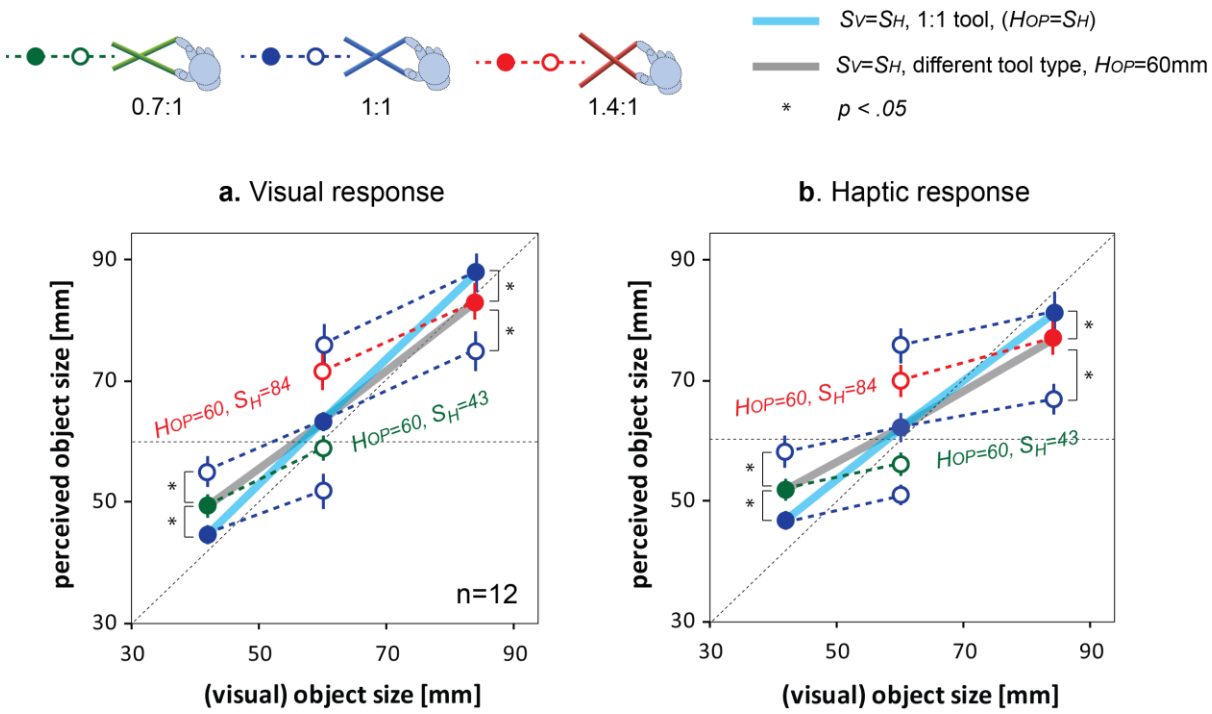


Figure 4.7. Perceived object sizes in vision-plus-haptics with three different tools. **a.** visual response **b.** haptic response. Perceived sizes were averaged across twelve observers. Here seven points are data when using the 1:1 tool, similar to the no-tool condition, with four additional points: (S_V, H_{OP}) = (43, 60), (60, 60) for the 0.7:1 tool and (S_V, H_{OP}) = (60, 60), (84, 60) for the 1.4:1 tool. The dashed lines link the same hand opening. Filled circles are no-conflict (S_V=S_H) stimuli and open circles are conflict (S_V≠S_H). Asterisks (*) indicate statistically significance differences (p < .05). Error bars denote ±1 standard error.

The three blue dashed lines linked the data at the same haptic object sizes ($H_{OP} = S_H = 42, 60, 84\ mm$) when using the 1:1 tool in a similar manner to the no-tool condition. Compared the performance between no-tool and 1:1 tool use, repeated-measures ANOVA analysis across seven size-combinations and twelve observers indicated significant differences on the main effect of sizes: $F(6,66)$ =

74.50, $p < .01$ for visual responses and $F(6,66) = 121.87$, $p < .01$for haptic responses. Importantly, there were no statistically significant differences on the main effect of the conditions (no-tool vs. 1:1tool): $F(1,11) = 2.48$, $p = .14$ for visual responses, and $F(1,11) = 0.06$, $p = .81$ for haptic responses. The slopes also clearly indicate that both visual and haptic estimates contributed to the combined perceived size in both visual and haptic responses.

As illustrated in the Figure 4.7, at visual size = 43mm, repeated-measures t-tests show statistically significant differences between 1:1 tool (no-conflict) and 0.7:1 tool, $t(11) = 2.53$ $p < .05$ for visual response and $t(11) = 3.69$, $p < .01$ for haptic response, and also between 1:1 tool (conflict) and 0.7:1 tool, $t(11) = 3.47$, $p < .01$ for visual response and $t(11) = 5.01$, $p < .01$ for haptic response. At visual size = 84mm, repeated-measures t-tests show statistically significant differences between 1:1 tool (no-conflict) and 1.4:1 tool, $t(11) = 2.32$, $p = .04 < .05$ for visual response and $t(11) = 2.30$, $p = .042 < .05$ for haptic response, and also between 1:1 tool (conflict) and 1.4:1 tool, $t(11) = 3.21$, $p = .008 < .01$ for visual response and $t(11) = 7.87$, $p < .01$ for haptic response. The Figure and these analyses indicate incomplete scaling; when using different gain tools (0.7:1 tool and 1.4:1 tool $S_V = S_H \neq H_{OP} = 60mm$), the perceived size were significantly different from the perfect rescaling (1:1 tool, $S_V = S_H = H_{OP}$) and conflicting (1:1 tool, $S_V \neq S_H = H_{OP} = 60mm$).

Importantly, even though hand opening was constant ($H_{OP} = 60\,mm$), observers rescaled haptic size estimates according to the tool gain. In the Figure 4.7, the grey line linked the "no-conflict" data when $S_V = S_H, H_{OP} = 60mm$ using different gain tools (0.7:1, 1:1, 1.4:1), and this line was close to the diagonal. Comparing with the 1:1 gain tool with no-conflict (the blue line) and three different gain tools with "no-conflict" (the gray line), repeated-measures ANOVA analysis across two sizes (43mm, 84mm) and two conditions (1:1 tool, different gain tool) indicated that they were statistically significant differences on the main effect of sizes: $F(1,11) = 275.47$ $p < .01$ for visual response and $F(1,11) = 204.02$, $p < .01$ for haptic responses; importantly not statistically differences on the main effect of conditions: $F(1,11) = 0.003$, $p = .96$ for visually matching, and $F(1,11) = 0.49$, $p = .50$ for haptically matching. The interaction (sizes × conditions) showed that

$F(1,11) = 6.64$, $p = .026 < .06$ for visual response and $F(1,11) = 10.50$, $p = .008 < .05$ for haptic response.

Further analysis was conducted for visual size (43, 84mm) vs. conditions (no-tool with conflict, tool use with no-conflict), under $H_{OP} = 60mm$. Repeated-measures ANOVA analysis indicates that there were statistically significant differences on main effect on sizes: $F(1,11) = 52.85$, $p < .01$ for visual response and $F(1,11) = 79.36$, $p < .01$ for haptic response, importantly there were not significant differences on main effect of the conditions: $F(1,11) = 3.385$, $p = .09$ for visual response and $F(1,11) = 1.153$, $p = .306$ for haptic response, and significant differences on the interactions (sizes x conditions): $F(1,11) = 18.32$, $p < .01$for visual response and $F(1,11) = 15.15$, $p < .01$ for haptic responses. These results show that perceived size estimates were changed, interacting with visual size and tool gain even when the hand opening was invariant. Therefore, using different gain tools seemed to cause haptic signals to be rescaled.

Individual cue weights (w_V, w_H) were also calculated from the individual performance in the similar manner with normal grasping. Figure 4.8 shows that visual and haptic weights, averaged across twelve observers, were roughly equal for visual responses ($\mathrm{w_V/w_H} = 0.46/0.54$) but haptic weighted more for haptic responses ($\mathrm{w_V/w_H} = 0.23/0.77$). Although the values were slightly changed, these trends (haptic weighted more for haptic response) were similar to the no-tool performance as shown in Figure 4.5. In addition, although it is hard to analyse these data statistically due to noisy data and large individual differences, these results were supported by another study such as Helbig and Ernst (2007). Because the same signals were used for both visual and haptic responses, these weight differences seemed to be caused by response biases and attention or other mechanisms might be involved in such response biases.

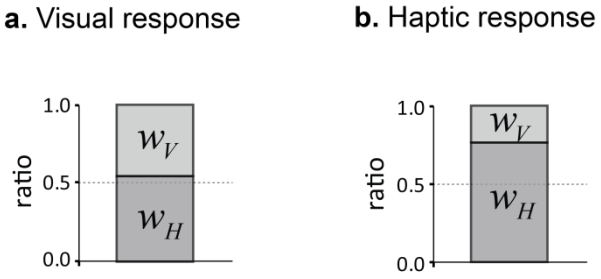


Figure 4.8. Visual and haptic weights in vision-plus-haptics with tool use, averaged across 12 observers. Weights were calculated from the slope in Figure 4.7 using Equation (4.1). **a.** visual responses **b.** haptic responses.

Figure 4.9 shows perceived object size with tool use in the haptics-alone condition. The data were also normalized by the individual performance in the no-tool condition at $S_V = S_H = H_{OP} = 60mm$. In a similar manner to the tool use in vision-plus-haptics condition, the blue lines link the data when using the 1:1 tool ($S_H = H_{OP}$). The perceived sizes were also close to the diagonal for visual responses, but the slope was slightly tilted from the diagonal dotted line for haptic responses. In comparison between vision-plus-haptics and haptics-alone with the 1:1 tool, repeated-measures ANOVA analysis on the main effect of object sizes (43, 60, 84 mm) indicates significant differences: $F(2, 22) = 497.28$, $p < .01$ for visual response and $F(2, 22) = 308.38$, $p < .01$ for haptic response; ANOVA analysis indicates no significant difference on the main effect of conditions (vision-plus-haptics and haptics-alone) for visual responses: $F(1,11) = 4.15$, $p > .05$, but a significant differences for haptic responses: $F(1,11) = 17.10$, $p < .05$; ANOVA analysis indicates no significant interaction (sizes x condition): $F(2,22) = .788$, $p = .467$ for visual response and $F(2,22) = .034$, $p = .966$. This means that with haptics-alone, with the 1:1 tool, the perceived size was well matched to the actual object size for visual responses, but not for haptic responses.

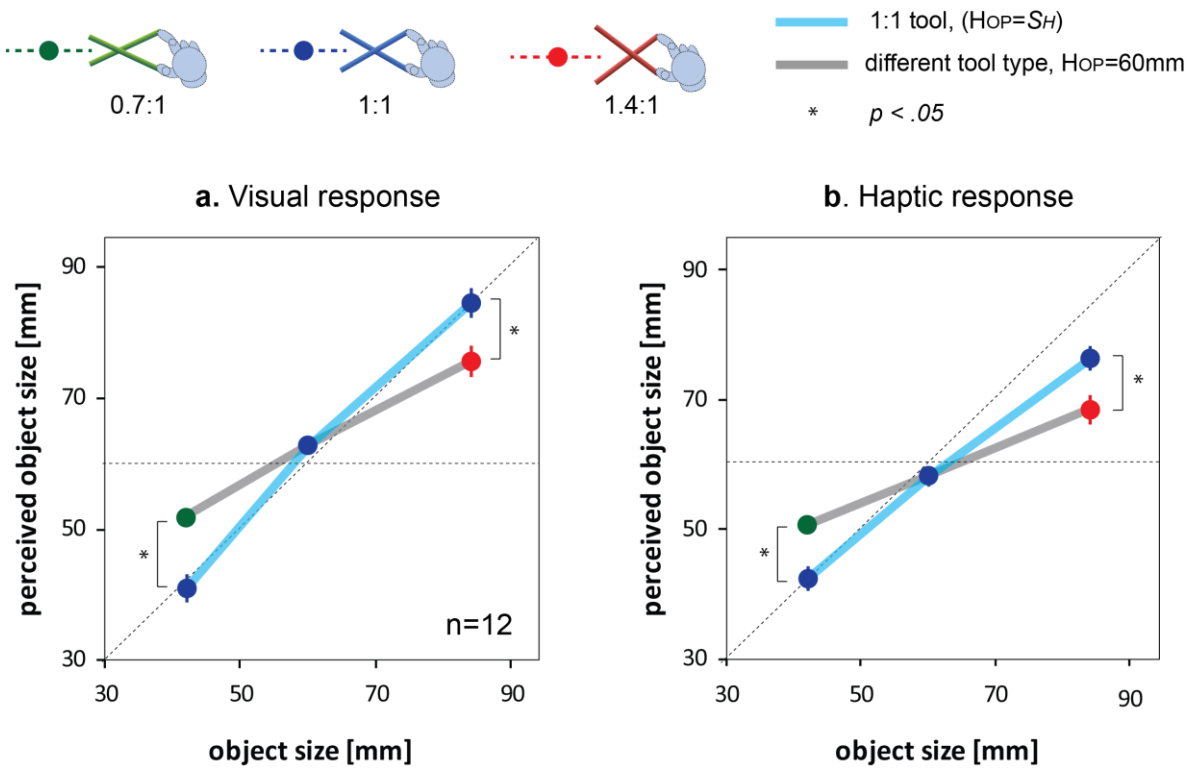


Figure 4.9. Perceived object size in tool-use, haptics-alone condition. The data were averaged across 12 observers. In the graph, 5 points: ($H_{OP} = S_H$) = 43, 60, 84 for 1:1 gain tool and (H_{OP}, S_H) = (60,43) for the 0.7:1 tool and (H_{OP}, S_H) = (60,84) for the 1.4:1 tool. The blue line links 1:1 tool use ($H_{OP} = S_H$) and the gray line links (H_{OP} = 60 mm) with different tool use. **a.** visual response and **b.** haptic response. Asterisks indicate statistically significance differences (p < .05). Error bars denote ±1 standard error.

The gray lines also link the data when using the different gain tools (0.7:1, 1:1, 1.4:1), where hand opening was constant $H_{OP} = 60mm$. This clearly shows that although the hand opening was constant, perceived size dynamically changed depending on the tool gain. The haptic estimates therefore were rescaled from the hand opening (hand coordinates) to the tool-tip (object coordinates) by tool geometry, even though no visual information was available. This rescaling was incomplete, however, resulting in biases in estimated size from haptics when using a tool. We found such rescaling for both visual and haptic responses.

Compared with 1:1 tool and different gain tools, repeated-measures t-test shows statistically significant differences at object size = 43mm, $t(11) = 5.25$, $p < .01$

for vision response, $t(11) = 3.92$, $p = .002 < .01$ for haptic response, and at objet size = 84mm, $t(11) = 3.023$, $p = .012 < .05$ for vision response and $t(11) = 7.21$, $p < .001$. It looks like rescaling was significantly reduced relative to the true tool geometry, which would be an interaction between object sizes and tool vs. no-tool conditions. Moreover, to examine which factor (object sizes and tool gain) strongly affected perceived size, we tested the observed data using repeated-measures ANOVA analysis. Compared with gray line and blue line, the analysis indicates a significant effect of object sizes (43mm, 84mm): $F(1,11) = 387.75$, $p < .001$ for visual response and $F(1,11) = 283.67$, $p < .001$ for haptic response, but no significant effect of tool gain (1:1 gain, different gains), $F(1,11) = 0.49$, $p = .50$ for visual response and $F(1,11) = 0.04$, $p > .8$ for haptic response; and the interactions (object sizes × tool gain) were significant $F(1,11) = 20.33$, $p < .01$ for visual response and $F(1,11) = 34.17$, $p < .01$ for haptic response.

We tested the observed data in tool-use cases (0.7:1 and 1.4:1 tools) between the vision-plus-haptics condition (Figure 4.7) and the haptics-alone condition (Figure 4.9) when hand opening was constant ($H_{OP} = 60mm$). Repeated-measures ANOVA tests showed statistically significant on the main effect of object sizes (43mm, 84mm), $F(1,11) = 100.43$, $p < .001$ for visual response and $F(1,11) = 107.97$, $p < .001$ for haptic response but no significant main effect of condition (visual tool, no-visual tool), $F(1,11) = 3.37$, $p = .094 > .05$ for visual response but significant for haptic response $F(1,11) = 12.38$, $p = .005 < .05$; and a significant interaction (size × condition), $F(1,11) = 8.63$, $p = .014 < .05$ for visual response and $F(1,11) = 12.22$, $p = .005 < .05$ for haptics. These results and statistical analyses support the idea that perceived size from haptics was rescaled in accordance with the type of tool, rather than the opening of the hand.

Moreover, we tested the observed data between no-tool (vision-plus-haptics) (Figure 4.4) and tool-use (haptics-alone) (Figure 4.8). Repeated-measures ANOVA tests showed statistically significant main effects of object sizes (43mm, 84mm), $F(1,11) = 84.31$, $p < .001$ for visual response and $F(1,11) = 102.61$, $p < .001$ for haptic response but interestingly no significant main effect of condition (no-tool, tool-use), $F(1,11) = .74$, $p = .41 > .05$ for visual response and $F(1,11) = 3.60$, $p = .084$

> .05 for haptic response; and also no significant interaction (size × condition), $F(1,11) = 1.39$, p = .26 for visual response and $F(1,11) = 2.48$, $p = .14 > .05$ for haptics. These statistical analyses indicate that even when hand opening (60mm) was only available to perceive the object size, the perceived sizes were deviated from 60mm. The perceived sizes in the haptics-alone with tool-use condition (Figure 4.9) were similar to the observed data when visual sizes were available in the no-tool condition (Figure 4.5).

Here, we examine whether the incomplete rescaling of the haptic estimates, which were found in the haptics-alone conditions, similarly contributed to the perceived size in vision-plus-haptics. Even though some response bias existed, overall we set visual and haptic cue contributions roughly equal to the final estimate in the vision-plus-haptics condition. Therefore, using the weighted linear summation model (See Equation (4.1)), we predict perceived size ($\hat{S}_{VH}$) with three different tools in vision-plus-haptics and compare with the observed data (linked with the gray line in Figure 4.7). We assume that (i) visual and haptic weights (w_V, w_H) might not change across different tools; so we use observed cue weights (in Figure 4.8), (ii) vision estimates ($\hat{S}_V$) would be unbiased and close to the real object size; so we use observed data with the 1:1 tool (linked with the blue line Figure 4.7), and (iii) haptic estimates ($\hat{S}_H$) would be rescaled by the tool; so we use observed data from haptics-alone condition (linked with the gray line in Figure 4.9).

Figure 4.10 shows the prediction obtained through the above calculation, compared with the observed data (filled circles) from Figure 4.7. The observed data were close to the prediction, within the 95% confidence intervals, in both visual and haptic responses. The haptic bias seemed to be the same, independent of which cues were available during tool use. That is, the visual estimate ($\hat{S}_V$) and the incomplete rescaling of the haptic estimate ($\hat{S}_H$) were combined to create a unified percept of object size when both cues were available in tool use.

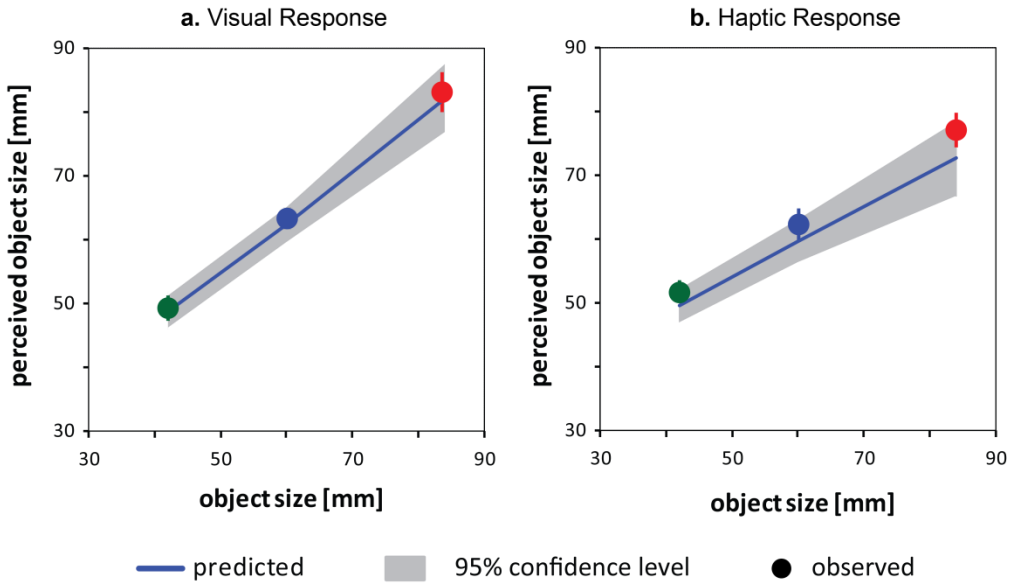


Figure 4.10. Prediction and experimental data. The prediction is plotted with 95 % confidence level, compared with the observed data at Figure 5.4. The prediction is calculated from observed data (cue weights in vision-plus-haptics condition, visual bias from 1:1 tool in vision-plus-haptics condition, and haptic bias in haptics-alone condition), using the weighted linear summation model.

4.3. Discussion

4.3.1. Chapter Summary

We showed that perceived size from haptics was rescaled with the type of tool, independent of the opening of the hand. This rescaling was incomplete, however, resulting in biases towards size estimates from hand opening. Similar rescaling of haptic estimates was found in both vision-plus-haptics and haptics-alone conditions, as well as across response modalities (visual and haptic), suggesting that the rescaling is the same whether or not haptic signals must be combined with vision. These results support an idea that haptic size estimates were dynamically rescaled, taking account of the geometry of tools.

4.3.2. Rescaling in tool use

Compared with psychophysical experiments to measure discrimination thresholds such as JNDs and PSEs, observing perceived object size using a matching paradigm was not hard evidence of rescaling because a possibility of intentional response cannot be eliminated completely. However, combining the findings of Experiment 4, our results could be considered as evidence that haptic size estimates were remapped in order to solve the visual-haptic correspondence problem during tool use.

We examined whether rescaling occurred at haptic size estimates and whether haptic size estimates could be remapped from hand coordinates to object coordinates, based on the tool geometry. This rescaling would allow the visual-haptic correspondence problem to be solved in common object coordinates. Even though rescaling was incomplete, our results showed that perceived size was changed according to the tool geometry not only in vision-plus-haptics but also in haptics-alone conditions. Because tool geometry affected haptic size estimates without visual information about the object size, the results suggest that the rescaling likely occurred in haptic estimates. In other words, haptic size estimates were dynamically remapped from the hand coordinates onto the common object coordinates, where the tool-tip is actually acting to pinch an object, regardless of visual information. This rescaling process might be obligatory to establish the relationship between hand opening and haptic size estimates at the likelihood level, bearing upon the tool geometry. The brain therefore appears to combine visual and haptic size information to be appropriately estimated in the common object coordinates.

4.3.3. Is rescaling a conscious or automatic process in tool use?

In addition, in our experiment, even though observers could not see the tool and the geometry during size judgement (Experiment 2-4), observers knew which tool they were using. We changed the tool type randomly on a trial-by-trial basis, aiming to avoid visuo-motor adaptation. This could also work for avoiding conscious judgements when measuring JNDs. However, unlike the discrimination experiment, it

is hard to eliminate the possibility of intentional modifications to size estimates when measuring perceived size using adjustment methods (Experiment 5). Observers might consciously bias their haptic perception using the tool geometry when they were responding. For example, it is possible that observers intentionally responded a larger size when using the 1.4:1 tool or a smaller size when using the 0.7:1 tool than their actual percept.

Viewing tools conveys the information of the tool, not only the geometry which directly relates to the spatial mapping but also its contexts such as colour (red, blue, and green). Observers might categorise tools according to the context, not geometry; for instance, the "red" tool as "big" and the "green" tool as "small", and then might make sizes adjustments according to such contextual information. Such indirect information for spatial mapping might affect the perceived size. This idea is quite problematic because it is hard to determine whether the tool remapping occurred by either actual spatial remapping or involving some different high level of conscious process. This issue will be discussed more in the General Discussion. The possibility of conscious response is hardly eliminated when using a matching paradigm to measure a perceived size; therefore, this can be considered as an experimental limitation of Experiment 5. We will therefore examine whether haptic size sensitivity to the same object size is affected according to tool geometry, by measuring JNDs in the next chapter.

Chapter Five

Cue weights in tool use

5.1. Introduction

We previously showed visual-haptic integration in a statistically optimal fashion during tool use, independent of spatial conflicts of raw sensory signals in their locations (Experiments 2 and 3) and magnitudes (Experiment 4). Moreover, we also showed dynamic rescaling of haptic estimates (Experiment 5). These results support the notion that tool use can induce dynamic change of spatial mapping. The results are also consistent with Bayesian accounts of cue integration (e.g. Ernst, 2007; Körding et al., 2007a).

As described in previous chapters, in normal grasping, vision and haptic weights are determined by the reliability of each cue. As a result of changes in the tool gain, the change in the hand opening caused by a given object size might affect the haptic reliabilities. For example, feeling a 50mm object with a tool that causes a hand opening of 80mm could change the precision with which the size of the same object can be estimated in normal grasping. This may occur because the "base" hand opening and the tool gain changed the mapping between hand opening and changes in object size. If so, it is interesting to ask whether cue weights are set correctly, taking into account these changes. We therefore examine cue weight changes during tool use in this Chapter. This examination is also an important control for our earlier experiments. In Chapter 3, we assumed that haptic JNDs at the hand are linearly rescaled from "hand" to "object" space according to tool geometry. Measuring haptic sensitivity (as a function of size when using different tools) allows us to determine whether the previous assumption was correct.

5.1.1. Predictions of haptic sensitivity during tool use

As illustrated above, tool gains might alter haptic reliability even when grasping the same object size, and this change might induce alteration of cue weights. Here two hypotheses can be made by considering the spatial mapping process introduced by tool geometry.

Hypothesis 1: Haptic sensitivity to object size is determined by the object size *per se*, regardless of hand opening in tool use. If the limiting factor in haptic sensitivity is determined by object size (for example by higher-level neural structures representing size), there would be no differences between normal grasping and tool use. This means, even though hand size is changed according to the tool gain, the sensitivity to object size would not change.

Hypothesis 2: Haptic sensitivity to object size is determined by hand opening and the sensitivity at the hand is linearly rescaled by tool gain, which represents the relationship between object size and hand opening. In this case, the process is not limited by the high level size representation to determine sensitivity to size with a tool, rather low level sensitivity (hand opening) is interacting with the tool geometry. The tool gain would dynamically affect the haptic sensitivity to object size.

Figure 5.1 depicts size-discrimination performance predicted by the two hypotheses. In this calculation, we used a polynomial equation to illustrate the JND changes as a function of size with three different gain tools, as an example[3]. If Hypothesis 1 is correct, we predict no difference in performance in object space between tools with varying gain in object space (5.1 (a)), although we would predict differences in hand space (5.1 (b)). In contrast, if Hypothesis 2 is correct, we predict different performance in object space according to the tool gain in object space (5.1 (c)) and no difference in hand space (5.1 (d)). If the JND changes are expressed by a simple linear equation, they would fall on to the same line regardless of the tool gain in Hypothesis 1 (5.1 (a)), but they would be separated onto three different lines according to the tool gain in Hypothesis 2 (5.1 (b)).

[3] This was based on completed data, reported in Experiment 4.

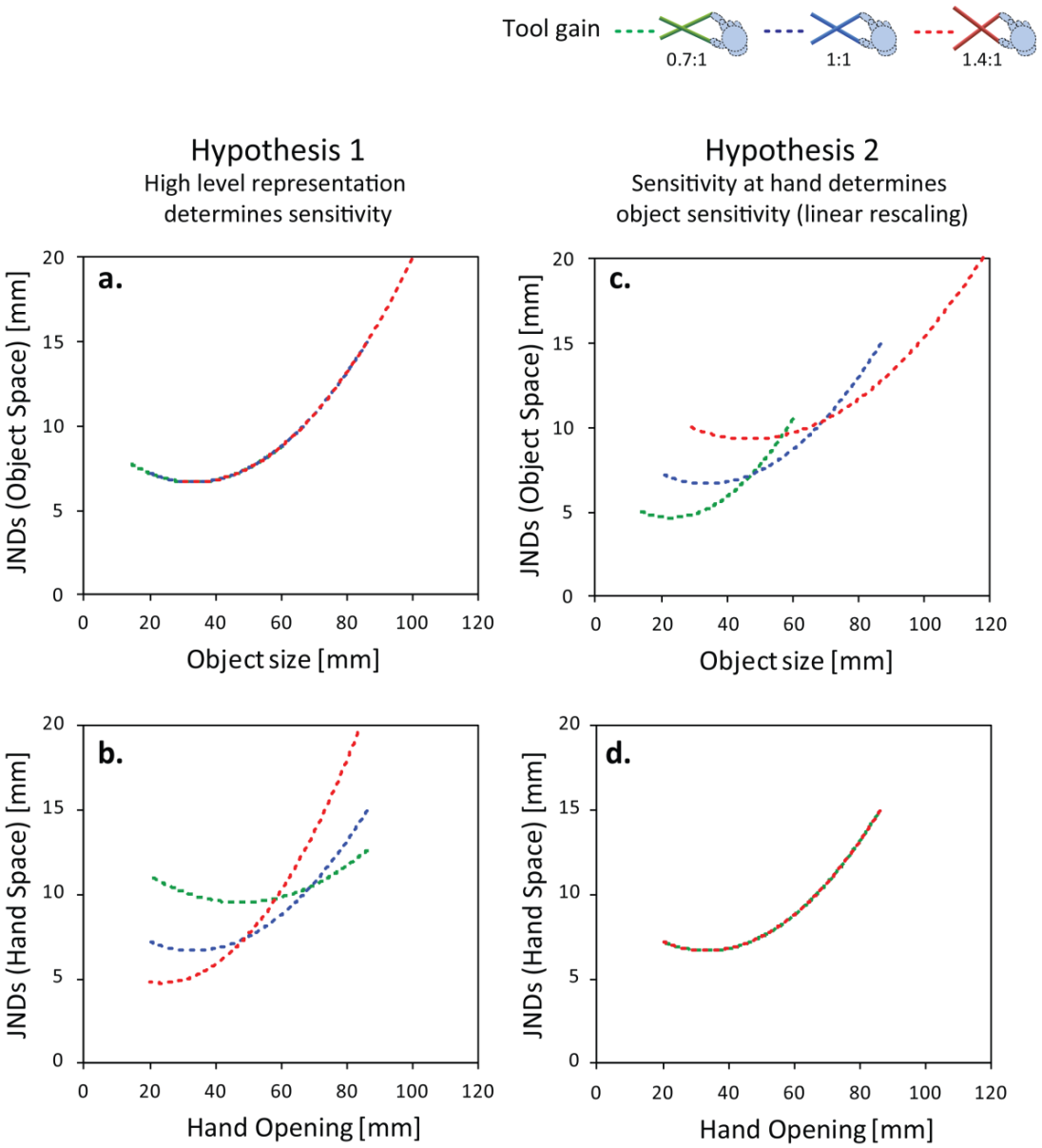


Figure 5.1. Size discrimination performance predicted by Hypothesis 1 and 2. JNDs are plotted as a function of object size (a, c) and as a function of hand opening (b, d). Figures illustrate three different gain tools, 0.7:1, 1:1, and 1.4:1, coloured green, blue, and red respectively. For the calculation, hand opening was varied from 20 to 86 mm, and we used a second degree polynomial curve.

We tested these hypotheses by measuring JNDs in both normal grasping and tool use and comparing performance to determine whether tool geometry affects haptic size discrimination in Experiment 6. Note that in specifying the predictions for

Experiment 4, we assumed that Hypothesis 2 was correct. This experiment, therefore, also serves as a control to test this assumption.

5.1.2. Prediction that cue weights change during tool use

If Hypothesis 2 is correct, haptic sensitivity would vary even for the same object size when grasped with a different tool. Therefore, it may be the case that haptic reliability, and therefore cue weights should dynamically change according to tool gain (if integration is to be optimal). Here, we examine how two possibilities of haptic sensitivity to object size – ruled by either Weber's law or a different relationship – would affect optimal cue weights.

Does haptic sensitivity follow Weber's law?

In general, perceptual sensitivities (ΔI), or just noticeable differences, are proportional to the magnitude of the reference stimulus (I) with a constant ratio, expressing the simple equation: $\Delta I / I = \text{Constant}$, known as Weber's law. The constant values are determined by physical stimuli. Although Weber's law is not always valid, the relationships are relatively preserved across various properties and different modalities. Therefore this law has been considered as a general principle, or rule-of-thumb, through a long history of psychophysical research in human perception (Stevens, 1957).

Several studies have supported that Weber's law holds for haptic sensitivity to object size. Historically, haptic sensitivity to object length was measured by finger span and adjustment methods were widely used (van Doren, 1995, Stevens & Stone, 1959). Recently, Pettypiece, Goodale, and Culham (2010) considered kinematic analysis of grasping an object into measuring the sensitivity to object size. Their participants grasped an object and then indicated size of the object with index finger and thumb (finger span, or manual estimation). Although kinematics properties of grasping such as grip aperture did not follow Weber's law, they demonstrated that manual estimation of object size was consistent with Weber's law. This consistency

was similar in manner to binocular sensitivity so that the sensitivity could be scaled with object size. However, several studies have reported the violation of Weber's law for haptic size sensitivities. For example, Durlach, Delhorne, Wong, Ko, Rabinowitz, and Holerbach (1989) reported that Weber's law was violated in haptic discrimination of object length. The discrimination performance in haptics was deviated from Weber's law when sizes were relatively small. The authors also suggested that the haptic sensitivity was likely to be affected by the finger lengths and palm size.

According to the above analysis, Weber's law is often thought as a rule, but actually often violated, especially where the hand is concerned. Haptic sensitivities may be relatively task-dependent and possibly limited by individual hand size differences, it is therefore important to know whether or not Weber's law in haptic sensitivities is preserved across sizes in our experimental environment.

Predicted cue weights for Weber and non-Weber cases

We first calculated how haptic sensitivity and cue weight in object space are affected by the above two possibilities. For haptic sensitivity, we used a simple linear equation for Weber's case and a second degree polynomial equation as an example for non-Weber's case. As Weber's law generally holds for the relationship between visual sensitivity and object size, we used a simple linear equation for vision.

Figure 5.2 shows sensitivity and optimal weight changes with three different gain tools. The patterns of haptic sensitivities in hand space (either Weber's law or a polynomial relationship) are clearly different in the object space (a and b). If Weber's law is preserved and Hypothesis 2 is correct (hand opening multiplied by tool gain determines object size sensitivity), the sensitivity in object space would fall on the same line and the maximum sensitivity would be at the smallest size in a given object size range. However, if the relationship between size and sensitivity follows a polynomial function, haptic sensitivity would be strongly affected by the tool gain.

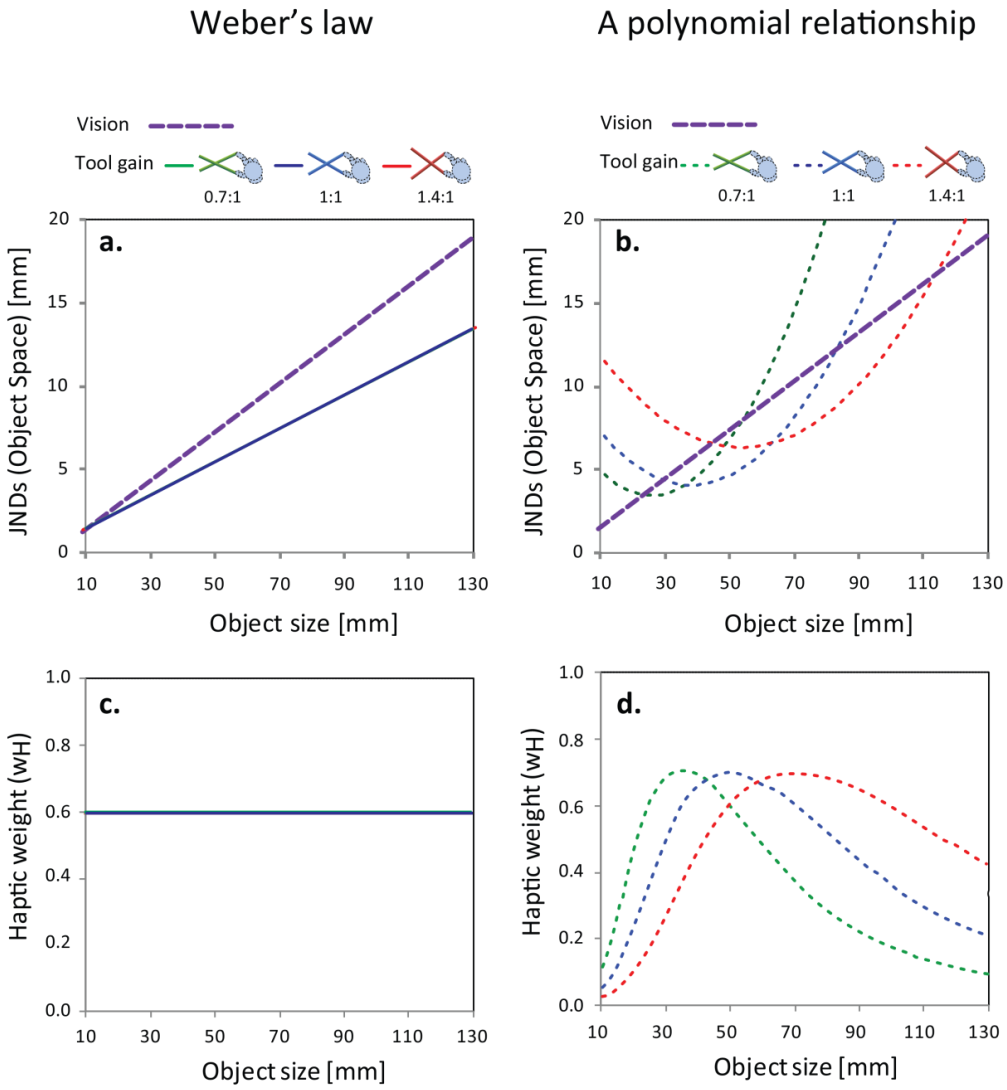


Figure 5.2. JND and weight changes according to two haptic sensitivity patterns. Figures illustrate performance differences between the two possibilities on haptic sensitivity. The left panel shows when haptic sensitivity follows a linear relationship for Weber's law, and the right panel shows a second degree polynomial relationship, which is based on the empirical data (Describe in Experiment 6.) The figures calculated are based on Hypothesis 2. JNDs are plotted as a function of object size (a, b). Weight changes are calculated by combining visual and haptic reliabilities (c, d). Figures illustrate three different gain tools, 0.7:1, 1:1, and 1.4:1, coloured green, blue, and red respectively.

To calculate predicted cue weights, we assumed visual sensitivities followed Weber's law and we used the Equations (4.1) and (4.2). To use these formulas, we also assumed that visual and haptic cues are unbiased. If haptics sensitivity is ruled by the Weber's case, the cue weight would not change, because both visual and haptic sensitivities follow linear equations; the ratio between them is unchanged across object size (Figure 5.2 (c)). In contrast, the cue weight would dynamically change when haptic sensitivity follows a polynomial relationship (Figure 5.2 (d)).

That is, regardless of whether haptic sensitivity is a function of object size or hand opening, if Weber's law holds, cue weights would not be expected to change between different tools because both cues are ruled by linear equations. While in the non-Weber's case for haptic sensitivity, cue weights would change with the different tool gains. Thus, understanding whether haptic sensitivity follows Weber's law is important for predicting the consequences for sensory integration of changing the tool gain. Here we explore this relationship, and examine whether the brain adjust cue weights appropriately with different tool gain.

In Experiment 6, we measured haptic discrimination thresholds in normal grasping and using three different gain pliers, across varied object sizes, in a similar way to Experiment 4. In Experiment 7, we measured the weights given to vision and haptics with each tool, using a cue-conflict paradigm, in a similar way to Experiment 5. We demonstrated that haptic weight was appropriately changed based on the tool geometry.

5.2. Experiment 6: Haptic sensitivity in size-discrimination performance

To investigate whether sensitivity follows Weber's law, we measured haptic size-discrimination thresholds across various object sizes (from 30mm to 80mm) with either normal grasping or tool use. We also examined how such haptic sensitivity varies across object size when using three different gain tools. We conducted both these experiments in a haptics-alone condition.

5.2.1. Methods

Observers

Six right-handed observers (aged 19-36, 3 males: AC, AW, WW, and 3 females: CW, LC, LH) took part in this experiment for both no-tool and tool conditions. All observers had normal or corrected to normal vision with normal stereo acuity, and no known motor deficits. All observers were naïve to the purpose of the experiment.

Apparatus and stimuli

The haptic stimuli represented an object (a rectangular cuboid) which was made by two parallel rectangular planes, similar to those used in Experiments 1 - 5. As before, the stimuli were rendered by two PHANToM force-feedback devices, one each for the index finger and thumb of the observer's right hand. No visual information relating to the object size was provided in any condition.

Haptic object size was varied according to a method of constant stimuli: 6 standard sizes 30, 40, 50, 60, 70, 80 mm and 8 comparison stimuli ± 1, 3, 6, 9 mm for each standard size. In this experiment, we set the stiffness slightly more rigid 1.05 [N/mm] than previous experiments 0.8 [N/mm], to prevent squashing the minimum size (21mm) of spongy-like virtual object for all the observers.

The three different types of tools were visually defined "pliers", similar to Experiment 4 and 5. By moving the pivot position in the same manner as before, we changed the size magnification relationship between the tool-tip and the handle which either: (i) preserved the normal 1:1, (ii) reduced with a 0.7:1 gain, or (iii) magnified with a 1.4:1 gain. The tool length was 180mm for all three tools. These tools were coloured differently, (i) blue, (ii) green, and (iii) red. Haptic stimuli were presented at the hand based on the tool geometry. As in previous experiments, we added the fronto-parallel force plane, to limit the movements on the plane.

Procedure

The procedure was similar to the haptics-alone conditions in Experiments 1 and 5. We measured size-discrimination thresholds (JND) in haptics-alone, with normal grasping (no-tool condition) and with three different tools (tool condition). Observers were required to judge the object size in a two-interval forced-choice (2-IFC) task. Haptic object size was varied according to a method of constant stimuli, and observers indicated which interval contained the larger size.

In each interval, two visual spheres appeared indicating the position of the upcoming stimulus (but no size information). When observers inserted their finger and thumb in the no-tool condition into the 'start zones', the start zone colour changed from yellow to green indicating they were ready to grasp the stimulus. All visual information disappeared immediately when the observers moved their digits inward from the start zone position. Observers were trained to grasp the stimulus for approximately 1 sec in each interval and then released it. The presentation timing was restricted to between 900 msec and 1200 msec, as in the previous experiments. The object position varied randomly from trial to trial to prevent the position of one plane being used to complete the task.

In the tool use condition, three different types of tools were randomly interleaved. Observers inserted the tool-tips into the start zones. All visual information including the tool disappeared when tool-tip moved inward from the green coloured start zone.

Both no-tool and tool-use blocks were completed separately. Each block contained six different standard stimuli and their eight comparisons, chosen at random. In each condition, observers completed 30 repetitions of each standard stimulus level. No feedback was given at any stage during the experiment. We measured psychometric data in size discrimination for each observer and the data analyses were conducted in a similar manner as the previous experiments.

5.2.2. Results & Discussion

Figure 5.3 shows the results of size-discrimination performance (JNDs) in normal grasping, averaged across all six observers. A regression analysis showed that the data were fitted effectively by a second degree polynomial curve, rather than a linear line. The minimum point of JNDs was 43.8mm, not the minimum size of the set of stimuli. Haptic sensitivity therefore does not simply follow Weber's law, in particular when deviations from the law are apparent at relatively small object sizes.

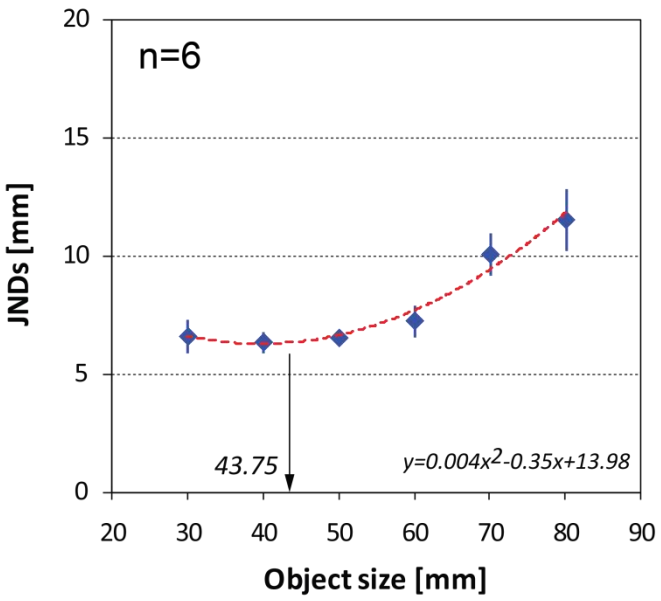


Figure 5.3. Haptic sensitivity in the no-tool condition. JNDs as a function of object size, averaged across six observers. Note here all data are Object size = Hand opening. Error bars denotes ±1 standard error.

Two typical observers' performances can be seen in Figure 5.4 and all individual data in Appendix 6. Like AC, some observers clearly showed a maximum sensitivity at the middle of the range of object sizes. In contrast, like LC, other observers' maximum sensitivity was at the smallest object size. Despite these differences, no one showed a linear trend for their sensitivity. Instead, a second degree polynomial curve better represented all observers' data. It is possible that both types of performance are similar, however. The maximum sensitivity for the latter observers might be located below the smallest object size in the current set of stimuli. That is, the varied object stimuli might not be enough to measure their maximum sensitivity. However, we could not measure smaller sizes, because some observers still could squeeze the small size stimuli until their finger and thumb came in contact, even though surface stiffness increased more than the previous experiments.

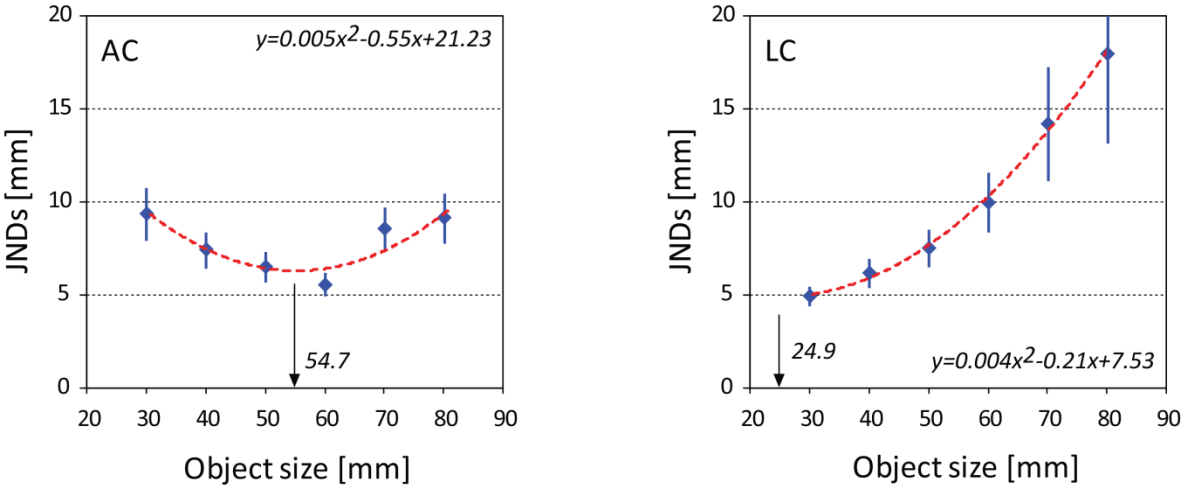


Figure 5.4. Haptic sensitivity across sizes. Typical observers: AC and LC. Both were fitted by a second degree polynomial curve. Note here all data are Object size = Hand opening. Error bars denotes ±1 standard error.

Note that in our setting, when two finger sheds made a contact for a while, the apparatus started to vibrate. For the observers' safety, we did not measure less than

20mm. This "squishiness" was the experimental limitation when using such PHANToM devices to measure haptic sensitivity. However, consider the real situation: we can still make size judgement of a sponge-like object, even when fingers come to a contact. When such sponge-like object becomes smaller and smaller or softer and softer, sensory noise on size judgement might increase more and more. Thus the Figure trend might relate to the object squishiness and would be found even in the real situation. Further examination would be needed to compare this trend with the real object. Moreover, this would become clear if we measure JNDs as a function of object stiffness.

Figure 5.5 shows haptic sensitivity when using three gain tools, averaged across six observers. Predictions for Hypothesis 1 and Hypothesis 2 are plotted by calculations using a no-tool condition, considering the three different gain tools. In addition, individual data can be seen in Appendix 7. The data were analysed at the object coordinates, where haptic size was magnified from hand opening according to the tool gain. If limits on haptic sensitivity are high level of process, representations could be in object coordinates; and if they are low level; representations may be in hand coordinates. We therefore considered two predictions: Hypothesis 1 predicts that haptic reliabilities are determined by object size and Hypothesis 2 predicts that haptic reliabilities are determined by hand opening. To test Hypothesis 1, haptic reliabilities were calculated with the tool at object coordinates, directly from the no-tool performance. To test Hypothesis 2, haptic reliabilities were linearly rescaled based on the ratio between object size and hand opening. These calculations are superimposed on the observed data in Figure 5.5.

Haptic sensitivities between different gain tools, analysed at object coordinates, were significantly different: $F(2,70) = 66.67$, $p < .01$. This suggests that hand opening significantly affected the performance. Moreover, a statistical analysis was conducted to examine which hypothesis is supported when using a tool. Compared with two hypotheses, repeated-measures t-test show that there were statistically significant differences between two hypothesis: $t(107) = 3.52$, $p < .01$. Compared with JND measurements at object and Hypothesis 1, a repeated-measures t-test shows significant differences: $t(107) = 3.15$, $p < .01$; however, compared with

JND measurements at object and Hypothesis 2, a repeated-measures t-test shows no significant differences: $t(107) = .50$, $p = .62$. These analyses indicate that when using tools, observed haptic sensitivity in object coordinates were better fitted to Hypothesis 2 than Hypothesis 1.

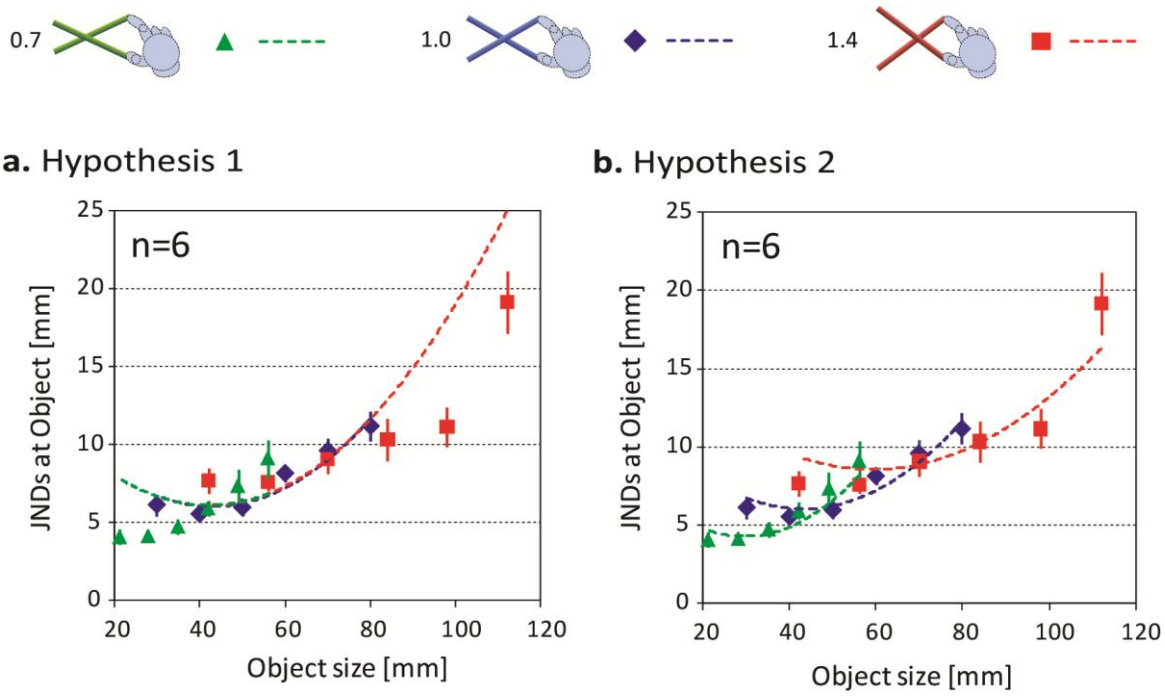


Figure 5.5. Haptic sensitivity at object space in the tool-use condition. JNDs as a function of object size, averaged across six observers. The filled markers represent observed data and the dotted lines represent prediction using observed data in normal grasping (Figure 5.3). **a.** Hypothesis 1. **b.** Hypothesis 2. Note: here all data and predictions are of object sizes, not hand opening. Error bars denotes ±1 standard error.

Figure 5.6 shows scatterplots of haptic reliabilities in the object coordinates between observed data and predictions to examine their correlations by a linear regression analysis. The line of best fit slope of Hypothesis 2 is 0.85 and the coefficient determination (R^2) is approximately 0.6, contrasting 0.50 and 0.48 respectively for Hypothesis 1. Both values indicate that Hypothesis 2 better fits the data than Hypothesis 1. Observed data (Figure 5.5) and statistical analyses (Figure

5.6) clearly show that when using different gain tools, we can conclude that haptic sensitivity to object size was affected by hand opening.

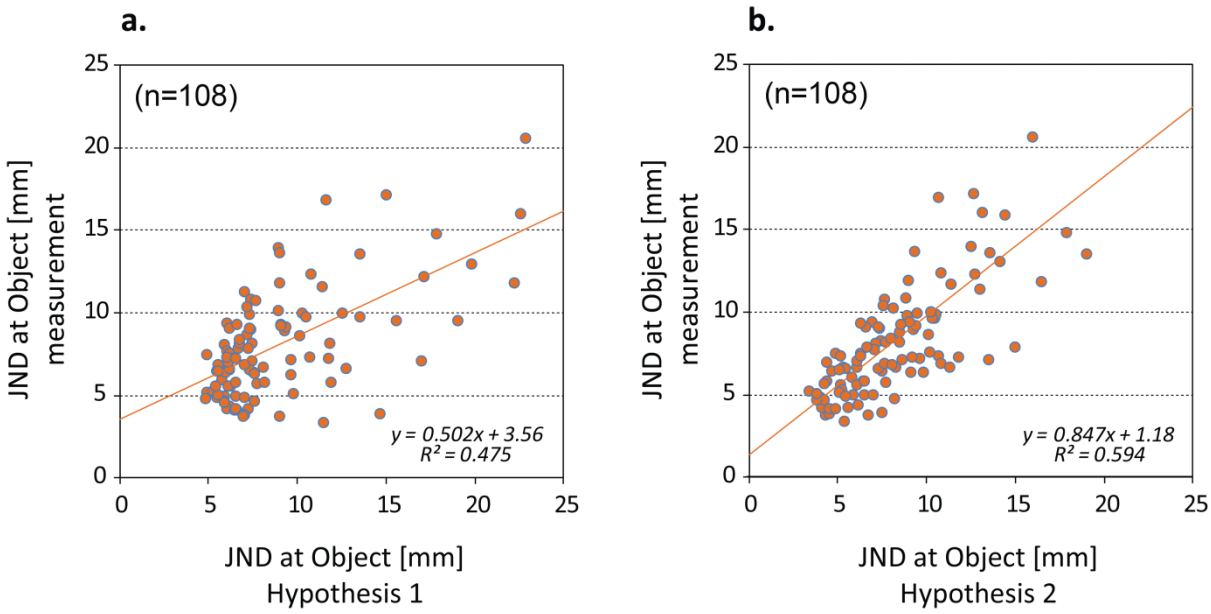


Figure 5.6. Scatter plots of haptic reliabilities compared with two predictions. The number of data points was 108 (6 different standard sizes x 3 different tools x 6 observers). Data were calculated at the object coordinates. **a.** Hypothesis 1 and **b.** Hypothesis 2. A linear regression analysis shows a best fit equation and R-square value and the trend-line is added to each graph.

In addition, we analysed data at the hand coordinates and the tool data were superimposed onto the no-tool data in Figure 5.7. The Figure shows that tool use performance was close to the normal grasping performance, regardless of different gain tools. Figure 5.7 shows a similar pattern to Hypothesis 2 (Figure 5.1 (d)), not Hypothesis 1 (Figure 5.1 (b)). This could be considered as strong evidence to support Hypothesis 2: linear scaling from haptic sensitivity at hand to object space.

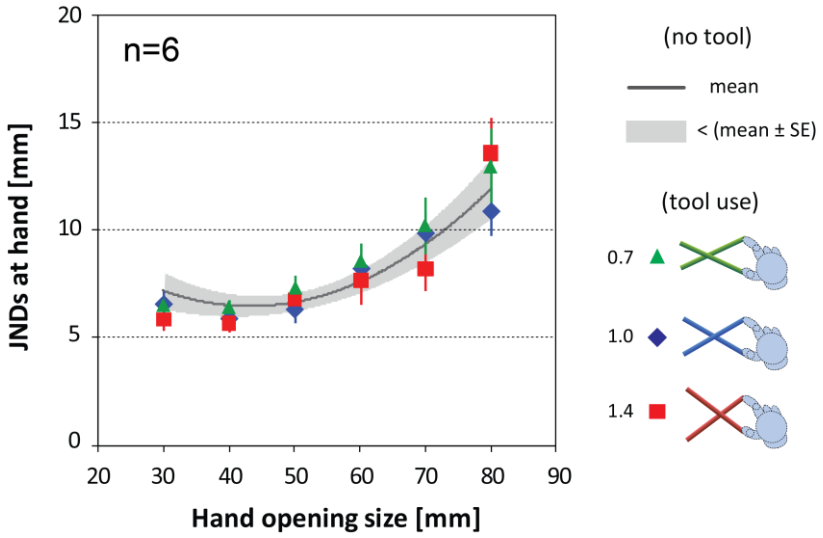


Figure 5.7. Haptic sensitivity at hand space. JNDs as a function of hand opening size, averaged across six observers. Data are compared with the no-tool and tool use conditions. The Shaded area represents within 1 standard error of the mean value. Error bars denotes ±1 standard error.

To merge the above analyses, we could conclude that haptic sensitivity at the hand strongly affected object size discrimination during tool use. In other words, changes in haptic object sensitivity rely on sensitivity at the hand, not object size. Thus it can be considered that the haptic sensitivity determined by hand opening was linearly rescaled onto haptic sensitivity of the object size, determined by the tool gain. This confirms our assumption used in Experiment 4, that is consistent with Hypothesis 2.

5.3. Experiment 7: Do cue weights change with different gain tools?

As shown in Experiment 6, haptic sensitivity to object size in size-discrimination performance did not follow Weber's law in the no-tool use condition. Previously we set two predictions: Hypothesis 1 assumes that haptic sensitivities are determined by object size, regardless of hand opening. Hypothesis 2 assumes that haptic sensitivities are determined by hand opening and they are linearly rescaled by a tool gain. Through the examination, Hypothesis 2 showed better fitting to the observed haptic sensitivities in the object coordinates. That is, haptic sensitivity to object size was strongly affected by hand opening given by the tool gain. Therefore when combining vision and haptic cues to estimate object size, an optimal system should change cue weights with tool gain changes. Here, Experiment 7 examined whether cue weights changed depending on different tool gains.

Figure 5.8 shows the point predictions for three different object sizes with three different gain tools, using the calculation in Figure 5.2 (d), where haptic sensitivity follows a polynomial curve and visual sensitivity follows Weber's law. We made point predictions using observed data in Experiment 6. The Figures demonstrated how haptic cue weight changes depending on the tool gain, across object sizes. The 0.7:1 tool results in haptic dominance when object size is small; in contrast, the 1.4:1 tool results in relative visual dominance when object size is small. Importantly, haptic cue weights using 0.7:1 tool and 1.4:1 tool are reversed between small and large sizes. The haptic weight when using 0.7:1 tool decreases with an increase of object size; whereas, using 1.4:1 tool increases haptic weight with an increase of object size. Although these are very rough predictions based on noisy observed data, the Figure importantly indicates that haptic weights dynamically change across object sizes with different gain tools. If we could measure this trend, it would be strong evidence for the brain optimally changing cue weights according to changes in tool gain.

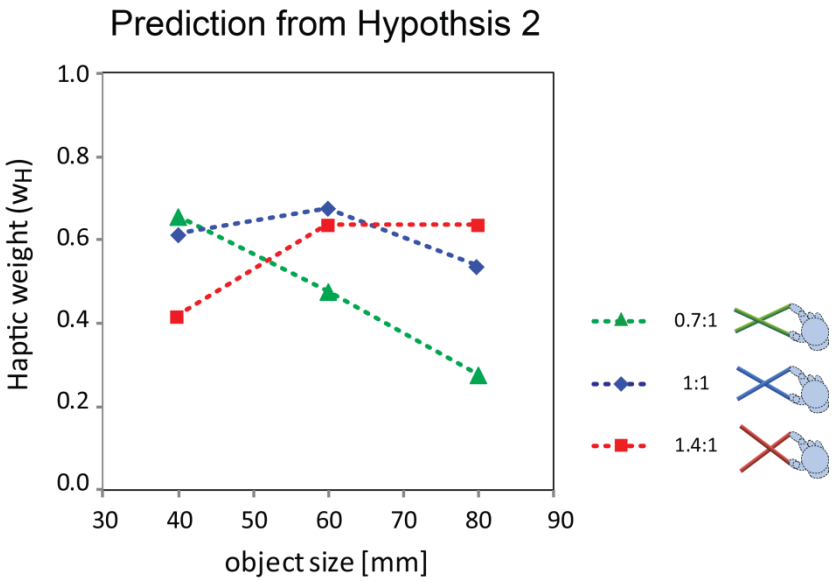


Figure 5.8. Prediction of cue weight change with three different gain tools across object sizes. Point predictions from Hypothesis 2: haptic sensitivity is linearly rescaled to object coordinate depending on the tool gain. Each cue weight is calculated by the linear summation model based on visual and haptic reliabilities (Figure 5.2 (d)). The prediction points are for three object sizes (40, 60, 80mm) and three gain tools, 0.7:1, 1:1, and 1.4:1.

We first estimated each cue's reliability by measuring size-discrimination thresholds in vision-alone and haptics-alone conditions. This enabled us to make predictions for cue weights in vision-plus-haptics conditions when different tools were used. Visual and haptics reliabilities were roughly matched at the middle size of a set of stimuli. We measured perceived size in no-conflict and conflict conditions between visual and haptic object sizes, with the three different gain tools. We then examined whether tool geometry changed the cue weight and discuss this, by comparing experimental data and predictions.

5.3.1. Methods

Observers

Six right-handed observers (aged 19-36 years, 2 males: AC, WW, 4 females: JG, LC, LH, LS) took part in this experiment. Among them, 4 observers (AC, WW, LC, LH) participated in both Experiment 6 and 7. All observers had normal or corrected to normal vision with normal stereo acuity, and no known motor deficits. All observers were naïve to the purpose of the experiment.

Apparatus and stimuli

The apparatus was the same as that used in Experiments 1-6. As in Experiment 4, the random-dot stereograms were used as a visual object and force planes were rendered at the index finger and thumb. We varied haptic reliability using three different gain tools, the same as Experiment 6.

Procedure

Vision-alone and haptics-alone conditions

We first measured visual and haptic size-discrimination thresholds (JNDs) independently, in order to predict visual and haptic cue weights in the vision-plus-haptics condition. Object size was varied according to a method of constant stimuli. In the visual-plus-haptic condition, three standard sizes 40, 60, 80 mm were used; however we measured JNDs at four standard sizes 30, 50, 70, 90 mm with eight comparison sizes: ± 1, 3, 6, 9 mm for each standard size for two reasons: (1) we assumed that visual sensitivity would held to Weber's law, so we could calculate JNDs at 40, 60, 80 mm from the data at 30, 50, 70, 90 mm, and (2) our finding showed the relationship between noise level and JNDs relatively followed a second degree polynomial curve (See Experiment 4), so for curve fitting, JNDs at 4 points were needed. We added two different visual noises: either low (50~100%) or high (150~200%) to the object depth step.

In the haptics-alone condition, we used three gain tools (1:1, 0.7:1, 1.4:1). These tools were visually rendered in the same manner as Experiment 6. Haptic object size was varied using three standard sizes (40, 60, 80 mm) and eight comparison sizes (± 1, 3, 6, 9 mm) for each standard size. Consequently, hand opening size was varied depending on the tool gain. When using the 1:1 tool, standard and comparison sizes at the hand were the same as the haptic object size. When using the 0.7:1 tool, standard sizes at the hand were 57.1, 85.7, 114.3 mm, and comparison sizes at the hand were ± 1.4, 4.3, 8.6, 12.9 mm. Similarly, when using the 1.4:1 tool, standard sizes at the hand were 28.5, 42.9, 57.1mm and comparison sizes at the hand were ± 0.7, 2.1, 4.3, 6.4 mm.

In both vision- and haptics- alone conditions, the object position was randomly set from trial to trial to prevent the position of one plane being used to complete the task.

Vision-plus-haptics condition

In the vision-plus-haptics condition, we measured perceived sizes under the different object size combinations, using a cue-conflict method. Each trial consisted of two periods, the same as Experiment 5. In the first period, visual and haptic stimuli were presented simultaneously when touching an object with a tool. In the no conflict condition, visual object sizes were equal to the haptic object sizes (40, 60, 80 mm). In the conflict condition, visual object sizes were varied ± 10 mm from the haptic object sizes. To use unbiased estimators, we needed to control visual and haptic reliabilities to be equal, using the similar methods in the previous experiments. Based on the individual's single cue performance, appropriate visual displacement noise level was chosen for each observer so that visual reliability matched haptic reliability. We adjusted the noise level as $w_V = w_H \sim 0.5$ at 60mm, using obtained data in the vision-alone and haptic-alone conditions.

In the second period, a light grey coloured solid cuboid was presented on the screen. Its width and depth were fixed at 100mm and 20mm respectively, and its height varied with 5mm intervals from 20mm to 120mm with a total of 21 different

sizes. Observers adjusted the stimuli size until it matched the previously perceived object size by pressing a keyboard with their left hand. In this second part, we measured only visual responses because perceived object sizes showed similar trends between visual and haptic responses in Experiment 6 (Figure 4.5 - Figure 4.10). Even though it is hard to state the similarity between visual and haptic responses using statistical analysis because the measurements were noisy and individual differences were relatively large, the similarity was also supported by Helbig and Ernst's study (2007b).

This condition consisted of eight blocks: four blocks (labelled as S1, E1, S2, and E2) using no-conflict stimuli, and four main blocks (labelled as M11, M12, M21, and M22) using randomised no-conflict and conflict stimuli. Here S, E, M represent "start", "end", and "main" respectively. The no-conflict blocks (Ss and Es) had only nine combinations: three object sizes ($S_V = S_H$) and three gain tools. The main blocks (Ms) had 27 (no-conflict and conflict) combinations: three haptic object sizes (S_H), three visual object sizes (S_V) and three gain tools. Observers completed the first session (S1-M11-M12-E1), and then the second session (S2-M21-M22-E2). We measured 20 repetitions of per each combination and these were randomised in each block. The no-conflict blocks (Ss and Es) were set to examine learning effect of repetitive tool use. The obtained data from these blocks will be analysed in the General Discussion.

Note that for examining cue contributions to the unified percept, measuring PSEs using the cue conflict method with a 2-IFC task is preferable to the method of adjustment, as described in the previous chapter. However, if a 2-IFC task is applied, tool-use or no-tool is assigned each interval and the target object size is changed between two intervals; therefore, the task difficulty could increase. Because the tool disappears during size judgement, observers may confuse whether they are touching an object using a tool or by hand, when perceiving the object size in the 2-IFC task. Also the required perturbations in stimulus level increase the number of combinations needed between visual and haptic sizes. For these reasons, we employed the adjustment method, and measured perceived object size while using three different gain tools.

5.3.2. Results & Discussion

Figure 5.9 shows JNDs for vision-alone (a) and haptics-alone (b and c) conditions, averaged across six observers. The data of vision-alone performance were fitted by a linear regression that assumes Weber's law held across the range of object sizes. Haptics-alone performance is plotted in both hand coordinates (Figure 5.9 (b)) and object coordinates (Figure 5.9 (c)). As shown in Experiment 6, JNDs at hand coordinates showed a second degree quadratic curve. JNDs at object coordinates were consistent with the previous findings in Experiment 6.

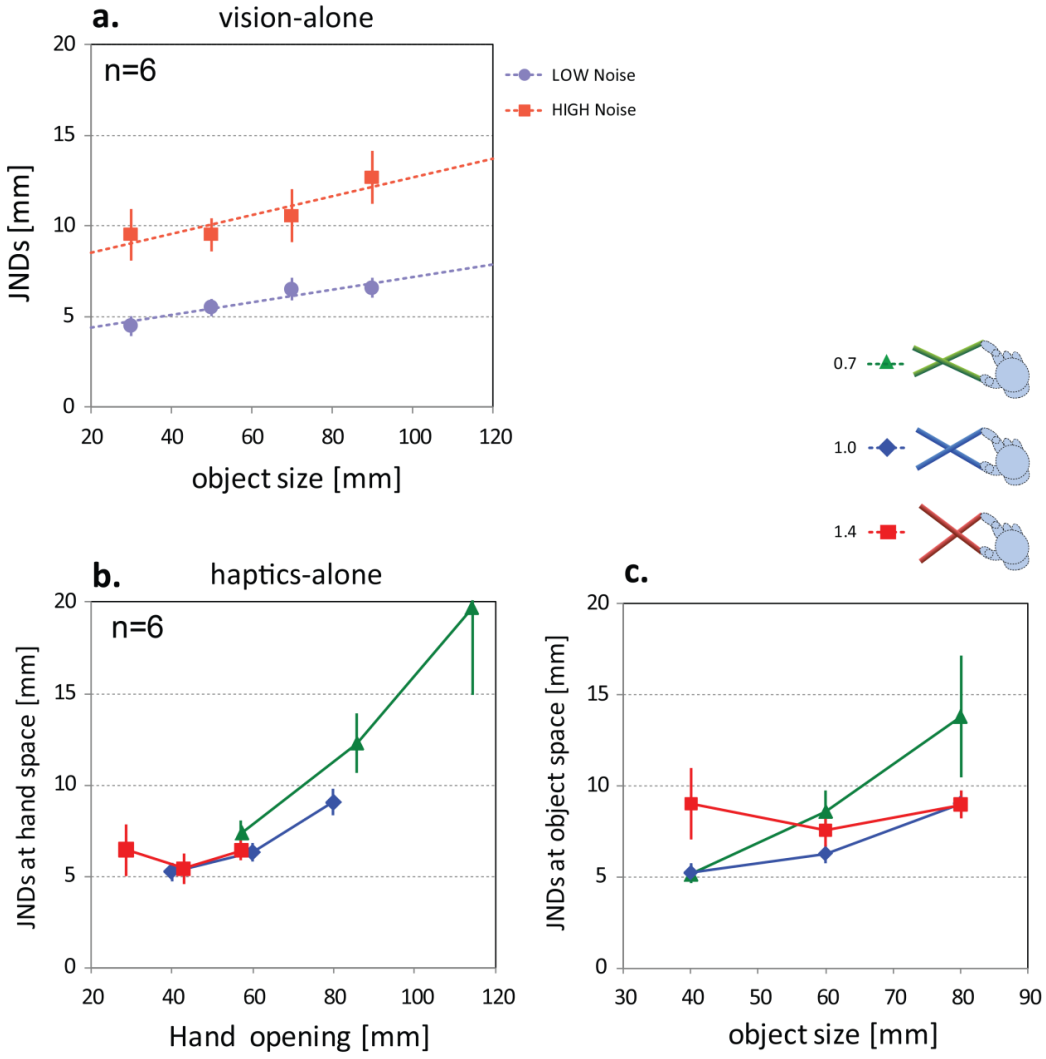


Figure 5.9. Experimental data averaged across six observers. **a.** size discrimination performance in the vision-alone condition across four different object sizes with two different displacement noises. **b.** size discrimination performance in the haptics-alone condition with three different gain tools. This graph represents at the hand coordinates (c.f. Figure 5.1 (d)). **c.** This graph re-plots the haptics-alone data but using object coordinates, rather than hand opening (c.f. Figure 5.1 (c)). Error bars denote ±1 standard error.

Figure 5.10 shows perceived size across object sizes with three different gain tools, averaged across four main blocks and six observers. We assume that a perceived size $\hat{S}_{VH}$ is estimated by combining object size estimates $\hat{S}_V$ from vision and $\hat{S}_H$ from haptics following the weighted linear summation model as described in Chapter1: $\hat{S}_{VH} = w_V\hat{S}_V + w_H\hat{S}_H,\ w_V + w_H = 1$.

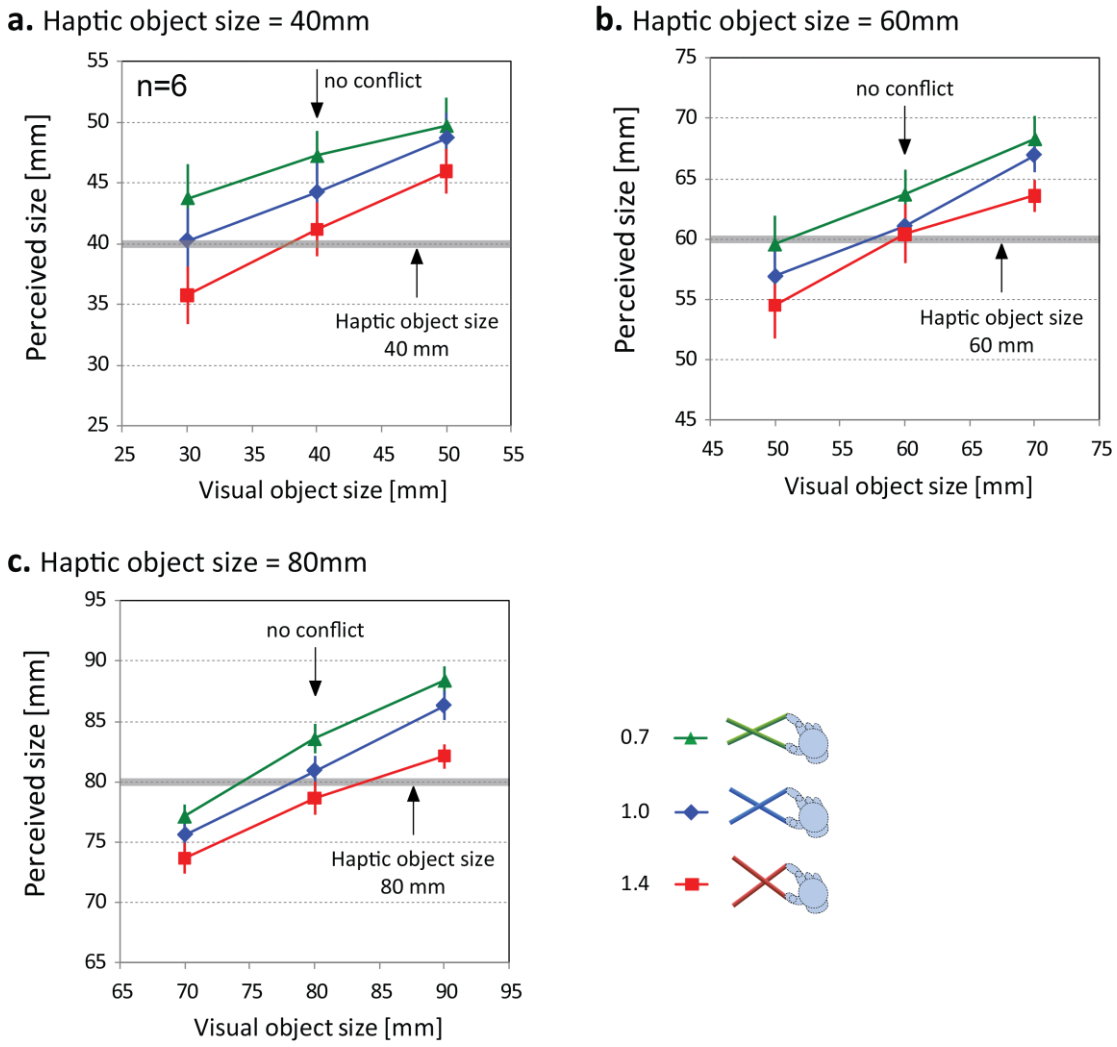


Figure 5.10. Experimental data of perceived sizes in vision-plus-haptics condition using three different gain tools, averaged across six observers. **a.** haptic object size = 40mm. **b.** haptic object size = 60mm. **c.** haptic object size = 80mm. Visual object size is varied ± 10mm (conflict) and 0mm (no-conflict). Error bars denote ±1 standard error.

Each graph shows perceived size when haptic object size was constant ($S_H = 40$, 60, 80 mm respectively). We assumed that visual size estimates were not biased: $\hat{S}_V \sim S_V$ and haptic size estimates were affected by hand opening sizes as we showed "incomplete rescaling" in Chapter 4. However, even if haptic estimate $\hat{S}_H$ was changed according to different gain tools, it was constant for three data points: $S_V = S_H \pm (0, 10$ mm) for the individual tool gain; therefore, slopes of three lines directly represent cue weights for different gain tools.

As such, individual cue weight changes were derived from individual perceived data. Figure 5.11 shows haptic cue weight change across three object sizes (40, 60, 80 mm) with three different gain tools, averaged across six observers. Individual data can be seen in Appendix 8. We predicted cue weights from visual and haptic single cue reliabilities at object coordinates based on Hypothesis 2, where haptic reliabilities are determined by hand opening and they are linearly rescaled by a tool gain. The predictions are plotted on the observed data.

As described above, Hypothesis 2 assumed that haptic reliabilities with tool use at the object are determined by haptic reliabilities at the hand and they are linearly rescaled by a tool gain. The prediction was calculated from visual reliabilities (σ_V) in the vision-alone condition, and haptic reliabilities (σ_H) in the haptics-alone condition, using Equation (4.2). The Figure shows that haptic weights changed according to the tool gain, and the trends of 0.7:1 tool and 1.4:1 tool were clearly reversed in the object size range.

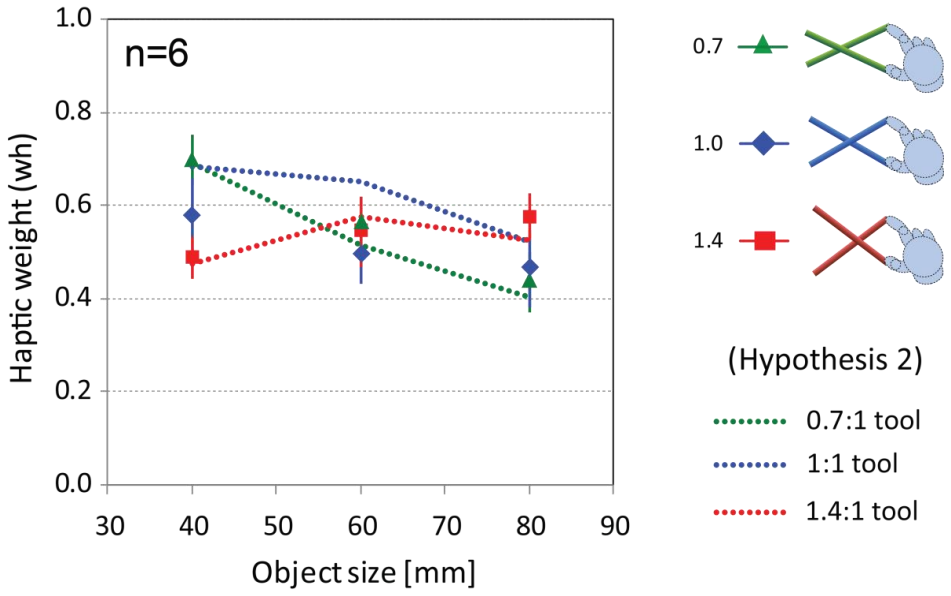


Figure 5.11. Experimental data with Hypothesis 2 in vision-plus-haptics condition. Haptic cue weight changes across object sizes using three different gain tools, averaged across six observers. Tools are 0.7:1 (reduced, green), 1:1 (normal, blue), and 1.4;1 (magnified, red) from hand to object coordinate, across object sizes with three different gain tools. Three lines, connecting three points calculated from Hypothesis 2, are superimposed on experimental data. Error bars denote ±1 standard error.

A repeated-measures ANOVA was used to test for haptic cue weight changes among three object sizes and three different gain tools. Weight changes differed significantly across the three object sizes, $F(2, 10) = 6.80$, $p = .014 < .05$, but not significant between the three tools, $F(2, 10) = 0.68$, $p = .53 > .05$, however the interaction between object size and tool was significant, $F(4, 20) = 10.25$, $p = .0001 < .05$. Post-hoc comparisons were conducted for all conditions. Tukey HSD test indicates that when object size = 40mm, 1.4:1 tool use ($M = 0.44$, $SD = 0.11$) showed significantly lower haptic weights than 0.7:1 tool use ($M = 0.70$, $SD = 0.13$), $p = .0009 < .05$. Conversely, when object size = 80mm, comparisons showed 1.4:1 tool use ($M = 0.58$, $SD = 0.12$) had a significantly higher cue weight than 0.7:1 tool use

(M = 0.49, SD = 0.16), p = .040 < .05. Based on these statistical analyses, cue weights changed significantly with tool gains. Figure 5.12 shows a linear regression analysis (a) between measurement and Hypotheses 1, in which haptic sensitivity is determined by object size, and (b) measurement and Hypothesis 2, in which haptic sensitivity is linearly rescaled from hand coordinates to object coordinates.

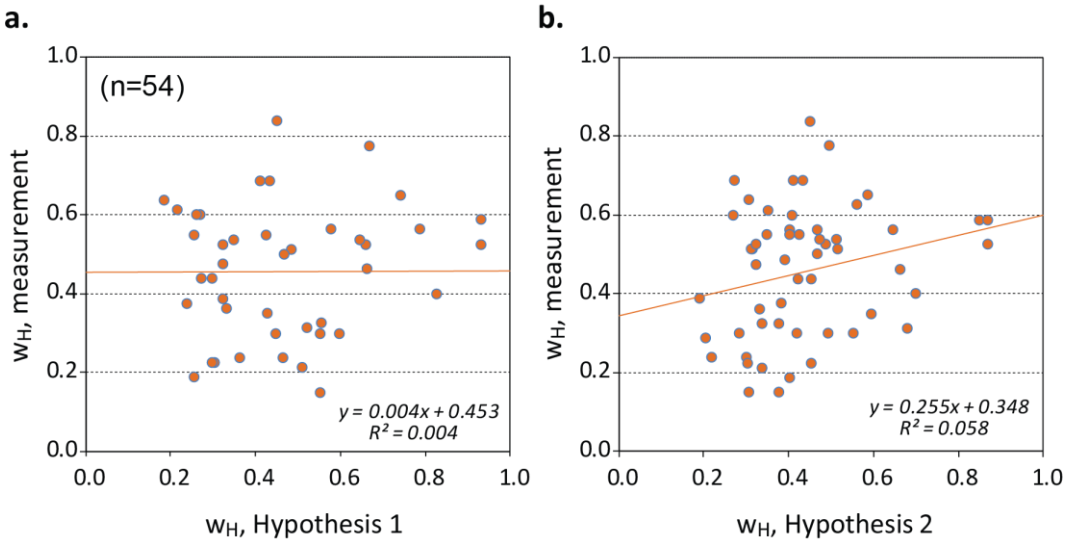


Figure 5.12. Correlation between prediction and measurement. **a.** Hypothesis 1 and **b.** Hypothesis 2. A linear regression analysis shows a best fit equation and R-square value and the trend-line is added to the graph.

Compared with Hypothesis 1 and 2, a repeated-measures t-test shows statistically significant differences: $t(53) = 3.58$, $p < .01$ (two-tailed). Compared with measurement and Hypothesis 1, a repeated-measures t-test shows significant differences: $t(53) = 3.48$, $p < .01$. In contrast, compared with measurement and Hypothesis 2, a repeated-measures t-test shows no significant differences: $t(53) = .82$, $p = .42$. It could be considered as that Hypothesis 2 would represent the experimental data better than Hypothesis 1.

Further statistical analysis was made. Pearson Correlations for Hypothesis 2 is $r(52) = .240$, $p = .04 < .05$ (*1 tailed*). Although R-square values of linear regression are quite small indicating weak fitting ($r^2 = 0.058$), estimated correlation coefficients β_1 (constant) = 0.348 and $\beta_2 = 0.255$. When setting null Hypothesis (*H0*) $\beta_1 = 0$ against Hypothesis 2 (*H2*): $\beta_1 \neq 0$ and *H0*: $\beta_2 = 0$ against *H2*: $\beta_2 \neq 0$, tests of significance of coefficient indicate: t-value for $\beta_1 = 5.277$, $p = .000$ and the t-value for $\beta_2 = 1.781$, $p = .081$. Although the correlation analysis is still too statistically weak to reject the null hypothesis (*H0*), combining with the results of Experiment 6 and also the above comparison between Hypothesis 1 and 2, Hypothesis 2 could be considered as a useful predictor of the observed data.

Overall, the obtained results showed that when using a tool, cue weights were appropriately adjusted to account for changes in haptic reliability introduced by the tool geometry (gain).

5.4. Summary & Discussion

5.4.1. Chapter Summary

Firstly, we showed that haptic sensitivity did not follow Weber's law in no-tool use, where the hand and object coordinates were matched. We also demonstrated that haptic sensitivity to object size was linearly rescaled from hand to object coordinates, according to tool gain. Importantly, different gain tools induced different relationships between visual and haptic size information. Therefore, haptic sensitivity to object size was strongly affected by hand opening given by the different gain tools; so cue weights were dynamically changed depending on the tool. These results support the idea that the brain appropriately adjusts cue weights when tool gain changes.

5.4.2. Haptic sensitivity limited by hand sizes

Firstly, Weber's law did not hold for haptic sensitivity across object sizes in our discrimination experiments. The result was consistent with several studies that haptic size sensitivity was likely affected by hand size, and that Weber's law was violated when object size was relatively small compared with finger and palm sizes. In our virtual environment, we could generate a force resisting against hand movement, but we could not introduce a rigid surface like real wood or a metal object. Therefore, some observers squeezed the small size with too much force and could not discriminate the size because their finger and thumb were nearly touching. Although there were such experimental limitations and individual differences such as hand size and exerted force, overall the data were effectively fitted by a second degree polynomial curve.

Although further investigation would be required to see whether this finding could be generalized to different situations, there is an important implication. Because hand size is limited, an optimal sensitive size likely exists in haptics. Such size may somehow relate to statistically natural hand postures. From our experimental data, the polynomial equation indicates that the best haptic sensitivity, averaged across six observes, was at a hand opening of 43.8mm (See Figure 5.3). This observed size is well matched to the size between index finger and thumb when taking a resting posture in everyday situations, reported by Ingram, Körding, Howard, and Wolpert (2008). According to their study, the statistical natural posture was roughly 40mm. We speculate one possibility that such resting posture is more natural and produces the most sensitive performance in size judgment. In our results, there were no statistically significant relationship between individual hand size components and their best sensitive sizes. However, this might be caused by the small number of observers used. Further experiments and meta-analysis would be needed using a large population to examine resting hand posture and the haptic sensitivity.

Secondly, we examined two possibilities about the haptic sensitivity during tool use. We found that the haptic sensitivity was not directly derived from the object size itself; rather it is relatively affected by hand opening. The sensitivity was linearly rescaled from the size sensitivity at the hand opening according to the ratio between

tool-tip and handle opening. If haptic sensitivity follows Weber's law (linear equation, not polynomial) with hand opening, optimal precision in object size would not happen because highly reliable performance would always be achieved at the smallest object size when choosing any gain tools. However, because haptic sensitivity is limited by hand size, when the polynomial sensitivity at the hand would be linearly rescaled from hand coordinates to object coordinates the optimally sensitive size could change depending on the tool gain. Therefore, the most sensitive size would be determined by the object size interacting with tool gain. This effect would be reduced when vision and haptics were combined.

5.4.3. Dynamic cue weight changes

Taken together, these findings demonstrate that the weight given to haptics varies with tool type in a manner which was predicted by the single-cue reliabilities, assuming the linear weighted cue combination rule (also, the MLE model; c.f. Ernst & Banks, 2002). Variations in haptic reliability introduced by different tool geometry therefore provide an account for the process of visual-haptic integration. The cue weights of the same object size were dynamically altered by changing the tool geometries on a trial-by-trial basis. The dynamic cue weight change was a good demonstration that haptic sensitivity to object size was affected by the sensitivity at the hand.

As the haptic sensitivity is limited at the hand opening, the sensitivity is not directly derived from the object coordinates. However, this does not mean that the correspondence problem was solved between raw sensory coordinates, or that haptic size estimate and its sensitivity did not share the same coordinates with vision. Instead, the cue integration occurred in the object coordinates where the haptic sensitivity was linearly rescaled from the hand coordinates in accordance with the tool geometry (in a similar manner as we measured haptic size estimates in Experiment 5). This implies the important notion that, although the correspondence problem during tool use should be solved in the object coordinates, the haptic estimate and its sensitivity are derived from the sensory information in the hand coordinates. The degree of visual-haptic integration would be determined by both

"low" level noisy sensory inputs and "high" level (object) representation noise and uncertainty, interacting with tool use and the rescaling.

In addition, these observed data would be useful to examine how size sensitivity is optimised in visual-haptic interfaces. In General Discussion, we will explore further how tool geometry does influence haptic sensitivity and the visual-haptic integration, from several different aspects.

Chapter Six

General Discussion

6.1.Summary of results

Tools, such as pliers, systematically perturb the normal spatial mapping between vision and haptics in terms of offset and gain. This thesis has explored how the brain solves the correspondence problem during tool use, in order to decide whether or not to integrate the sensory information across modalities despite these spatial changes. We have reported the results of several experiments, in Chapter 2 through Chapter 5, focusing on different aspects of the correspondence problem and cue integration, while examining the underlying mechanisms in tool use.

First, we showed that visual and haptic size information was integrated near-optimally, but only when appropriate to do so, independent of the spatial offset between object and hand positions (Experiment 2) and conflicting size between object and hand opening (Experiment 4). To explain the size-discrimination performance, we explored a computational model of the correspondence problem in Chapter 3. This analysis suggested that rescaling of haptic estimates, to account for tool geometry, could allow the correspondence problem to be solved using "object" coordinates. We then showed that haptic size estimates were rescaled when using tools with different gains (Experiment 5). The rescaling was not perfectly matched to object size, however. Further experiments showed that haptic sensitivity to object size when using a tool was not determined by object size *per se*, but was determined by sensitivity to different hand openings. That is, when using a tool, JNDs in object space were simply equivalent to JNDs in hand space, multiplied by the tool gain (Experiment 6). We then showed that, because haptic size sensitivity does not follow Weber's law, haptic reliability is altered by tool gain. We then demonstrated that cue

weights were dynamically altered when the tool gain changed (Experiment 7), in a manner consistent with these changes in reliability.

Overall, the results presented in this thesis provide important information for understanding the mechanisms that support tool use. In this chapter, we will restate the several principal hypotheses we have tested, with regard to key concepts such as Bayesian inference and visuo-motor processing. Moreover, we will suggest future research directions and implications for designing haptic interfaces.

We have been aware of several experimental limitations through the series of experiments. We ran all experiments using PHANToM devices and manipulated visual and haptic stimuli independently. The devices introduced unnatural spatial mapping and observers were required to judge object sizes in the unnatural situations, regardless of visually defined no-tool or tool representations. From this point of view, all settings were considered as using a "type of tool". We generated tools visually defined, which smoothly followed hand movements but did not have weights in an unnatural manner. In addition, various cues were unnaturally deprived in the virtual environment and available stimuli were quite limited. Although we manipulated spatial parameters, visual and haptic stimuli were presented simultaneously. This might facilitate observers to solve the correspondence problem from temporal aspects, not from spatial aspects; so we could not eliminate the possibility that observed performance might be influenced by such temporal correspondence. Such experimental settings were also used in other studies (e.g. Ernst & Banks, 2002; Gepshtein et al., 2005). As far as we know, visual-haptic integration has not been demonstrated in a natural situation yet. Even though all were in the unnatural situation, compared with tool and no-tool conditions, we clearly showed different performances in the visual-haptic integration.

6.2. Solving the correspondence problem during tool use

6.2.1. How to integrate visual and haptic size estimates when using a tool?

Experiments 2 through 4 showed that visual and haptic size information were integrated near-optimally, if a tool was appropriately grasping the object. Thus, in tool use, the correspondence problem was "correctly" solved, based on statistical similarities of sensory information in object space. The degree of integration reduced with increasing spatial offset between object and tool-tip, in a similar manner to increasing spatial offset between object and hand. The brain accounts for the statistical relationships between visual and haptic size, taking into account the geometry of the tool. In this way, both during tool use and normal grasping, the similarities could be treated in common object coordinates. If so, the correspondence problem in tool use therefore reflects general principles in multisensory integration.

With respect to understanding visual-haptic integration during tool use, we used the cue-integration framework, applying Bayesian approaches, and ideas from visuo-motor control. The Bayesian perspective provides a probabilistic framework to solve the correspondence problem in visual-haptic integration. We used two key concepts: a coupling prior (e.g. Ernst, 2005, 2007) and causal inference (e.g. Körding et al., 2007a). Recent ideas from the visuo-motor control literature are consistent with the idea that haptic estimates are rescaled from hand coordinates to object coordinates, according to the tool geometry. In Chapter 3, we demonstrated that combining these ideas is applicable to explain our tool-use results.

Figure 6.1 summarises our model and the possible process in tool use. The diagram shows how raw haptic signals could be remapped onto object coordinates and how the correspondence problem could be solved to generate a final percept. In this framework, the spatial remapping appears to occur at the sensory estimation, as suggested by the results of Experiments 5-7. The remapping affects haptic estimates and sensitivities in relation to the tool dynamics and geometries. The rescaled properties can then be combined with a coupling prior in object coordinates to solve the correspondence problem, and then generate the final percept (depending on whether the decision is that the signals have a common cause or two independent

causes). Thus, the decision would be made probabilistically from chance level to absolute certainty for choosing common cause or two independent causes. The visual-haptic integration would be gradually changed from complete integration to no integration depending on the degree of the correspondence.

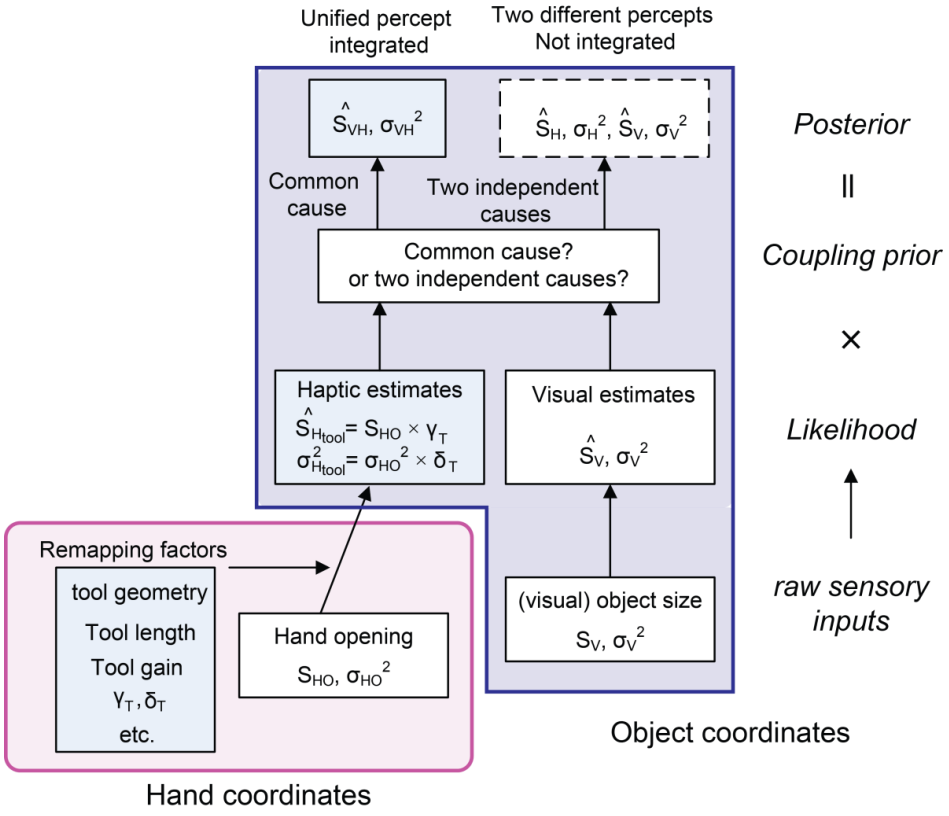


Figure 6.1. Schematics of the possible cue-integration mechanism when using a tool. As an example, the geometry of a tool affects haptic estimates in terms of both bias and variance. The brain would decide whether visual and haptic signals had a common cause or two independent causes at the posterior level. If common cause is assumed, the two sensory signals are combined to produce a unified percept. If independent causes are assumed, they are not combined, producing different percepts from vision and haptics.

The model predicts a degree of cue integration depending on the extent of the correspondence between signals. For example, as the conflict in magnitude between signals increases, the probability that the signals are from a common cause

decreases, therefore the signals become less integrated. We showed such a progressive reduction in visual-haptic integration with an increase of conflict, where size-discrimination performance was approaching to or even worse than single-modal performance when the conflict was large (spatial offset in Experiment 1-3, magnitude in Experiment 4 and Control Experiment 2). This progressive reduction was consistent with other empirical studies; for example, the integration in location judgements reduced with increased spatial separation between vision and audition (Battaglia, Jacobs, & Aslin, 2003; Beierholm et al., 2008).

The visual-haptic correspondence problem appears not to be solved on the spatial mapping of the raw sensory stimuli. Instead, it could be solved based on the spatial remapping to the inferred source of those stimuli. This remapping could occur at the level of the sensory estimate from haptics. The results of Experiment 5 showed that haptics-only size estimates did indeed change when tool gain changed. This would allow visual and haptic signals to be represented in a common space, despite very different sensory inputs. Other studies have shown similarity in perceptual space between vision and haptics when exploring 3-D objects (Cooke, Kannengiessen, Wallraven, & Bülthoff, 2006; Cooke, Jäkel, Wallraven, & Bülthoff, 2007; Gaiβert, Wallraven, & Bülthoff, 2010). For example, although sensitivity to object orientation is significantly different visually and haptically (Lawson, 2009), observers showed similar performance in both visual and haptic recognition of mini-scaled familiar objects, suggesting that size changes were generalised across modalities (Craddock & Lawson, 2009a, 2009b). Spatial remapping at the sensory-estimates level might make this possible if spatial representations in vision and haptics are treated in common coordinates.

Related research in computational models of tools

As far as we know, only one study has proposed a computational model of tool use. Extending from their causal inference model (Körding & Tenenbaum, 2007a), Körding and Tenenbaum (2007b) considered the causal structure of sensory experience with different coordinate systems between vision and hand movements.

They showed that visual-proprioceptive integration could also be formulated by the causal inference model. In order to deal with the relationship between tool-tip and hand, they introduced ‘movement vectors’ and ‘configuration of joints’ and treated the combination in the same coordinates. They demonstrated how to estimate cue reliability changes from individual coordinates, and they also predicted cue integration during tool use, by assuming a weighted linear model. Their basic idea about how the coordinate system changes between tool-tip and hand seems to be similar to our studies; however, their study treated a cursor as a tool. Thus, further examination would be required to determine, for example, whether the model works for more realistic tools.

6.3. Plausible underlying mechanisms in tool use

6.3.1. Visuo-motor control with a tool and mental representation of tool use

Bayesian inference ideas provide a good account of how the brain perceives the most probable states in the world, interacting with motor control and sensory feedback (Berniker & Körding, 2011; Körding & Wolpert, 2006; Wolpert & Ghahramani, 2000). Several studies have attempted to understand human goal-directed behaviour by combining Bayesian decision theory with ideas from motor control (e.g. Trommersäuser et al., 2003, 2008). Action and space are strongly interconnected; visuo-motor remapping, or coordinate transformation, occurs when moving the arm with visual feedback (Ghahramani, Wolpert, & Jordan, 1996). The forward and internal models could naturally extend to include tool dynamics. These models could produce the required neural representations to allow the system to predict the future state of the arm in object space interacting with a tool (Imamizu, Miyauchi, Tamada, Sasaki, Takino, Rülz, Yoshida, & Kawato, 2000; Kluzik, Diedrichsen, Shadmehr, & Bastian, 2008). It has been suggested that the visuo-motor remapping process can be explained by such models (e.g. Mehta & Schaal, 2002;

Wolpert et al., 1995). Here, we discuss tool use mechanisms from this visuo-motor perspective.

When controlling a tool, the brain needs to understand how muscle commands relate to the movement of the tools, in order to compute (predict) where the tool-tips are going to interact with objects in visual space. The process possibly modifies and also recalibrates any relationships in relation to tool use, for example between force and motion, interacting with space; then these may produce a new mapping between tool and arm (Cardinali, Frassinetti, Brozzoli, Urquizar, Roy, & Farnè, 2009; Debats, van de Langenberg, Kingma, Smeets, & Beck, 2010; Kluzik et al., 2008). As such, the brain could transform haptic estimates from hand coordinates to object coordinates according to tool geometry to share properties in the common coordinates.

There have been a few systematic analyses of the types of human tools and their categorical transformations. Beisert et al. (2010) investigated human tool use behaviour under different transformation rules, which are defined as the relation between operating hand movement (e.g. open and close fingers) and its functional movement (e.g. release and squeeze an object), by applying a tool-switching paradigm. They measured reaction time and error rates under different conditions by manipulating tools' functional structure and the transformation rule: compatible or incompatible and by providing pictures or words stimuli. The results implied that a tool structure provided by pictures might induce both implicit facilitation functionally (low-level processing) and explicit retrieval from restored transformation rules (high-level processing). In our experiment, we used a simple linear transformation rule from tool-handle to tool-tip. Further exploration therefore would be needed to investigate how different transformation rules affect visual-haptic integration.

Several studies have proposed that the brain produces internal models of tools, which could be differentiated from the body representation (Imamizu et al., 2000; Imamizu, Kuroda, Miyauchi, Yoshioka, & Kawato, 2003). There is evidence that the nervous system understands the dynamics of tools even when they are novel (Imamizu et al., 2000). In addition, Beisert et al. (2010) showed that even when tool structure and the transformation rule were incompatible, the rule can be embodied,

using a tool-switching and stimuli-response paradigm. They suggested that a specific transformation rule could be acquired through repetitive tool usage. Previous research has shown that a new relationship between arbitrary signals can be learned and in terms of the model, this new relationship can be expressed by the coupling prior (Ernst, 2007). With respect to our tool study, through repetitive tool usage, the variance of the prior distribution might be getting narrower and the correspondence problem could be solved more effectively through this remapping processes.

Neurophysiological and behavioural findings also imply that the mental representation of a tool could be consolidated through observing tools and through active tool use (e.g. Bach, Knoblich, Gunter, Friederici, & Prinz, 2005; Botvinick, Buxbaum, Bylsma, & Jax, 2009; Massen & Prinz, 2007). The process requires knowledge of the tool geometry and the dynamics of the manipulated object and tools. Mediating motor activities with a specific tool possibly functions to transfer the tool knowledge to a general representation (Schwartz & Holton, 2000) or multiple representations of tools specialised by contexts or functions (Ingram et al., 2010). The cerebellar activity measured by fMRI indicates that an internal model of a new tool is obtained when controlling an external object interacting with forward and/or inverse kinematics (Imamizu et al., 2000). Even when imagining using a common tool, the cerebellum is activated (Higuchi, Imamizu, & Kawato, 2007). Thus, such mental representations, or even imagery of tools and tool use, may facilitate prediction of environmental consequences of motor actions with tool use.

6.3.2. Switching vs. quick adaptation

One important question is what sensory cues are used to tell the nervous system which tool it is using at a given time. Here, it should be highlighted that in our experiments, tools were randomly assigned on a trial-by-trial basis, in order to prevent visuo-motor adaptation. As shown in the Control Experiment 1 in Chapter 2, normal visuo-motor adaptation in the felt position of the fingers did not occur during tool use. That is, observers knew where their hand was located when using a tool. So the spatial remapping of haptic signals could not be explained by normal visuo-motor

adaptation. Rather the brain appears to dynamically and flexibly switch between different remapping "states". These results could also not be explained by a "body schema" account, that the body representation itself is changed by tool use. Instead, the brain seems to differentiate between tool and hand (Povinelli, Reaux, & Frey 2010) using a switching strategy and the sense of body-ownership does not change during tool use (De Preester & Tsakiris, 2009). Our results are therefore more consistent with previous research such as multiple representations of tool dynamics (Ingram et al., 2010) and transformation rules in tool use (Beisert et al., 2010; Massen & Prinz, 2007). We suggest that the solution to the visual-haptic correspondence problem during tool use therefore could be explained by "switching" or more dynamic changes in mapping processes. In our experiment, we manipulated the tool gain on the trial-by-trial basis but the transformation rule from hand coordinates to object coordinates were quite simple. The pliers' type of tool operated with simple linear scaling; that is, when the tool handle was widening, the tool-tip was also widening with different gain (minifying or magnifying). Therefore, we need to consider further different transformation rule: for example, when the tool handle is widening, the tool-tip is getting narrower (e.g. pincer and clothespin types) and it would be interesting to examine whether visual-haptic integration still occurs even when the transformation rules are dynamically switching.

Different internal models therefore might be established depending on individual tools, and these could be switched from one to another and/or to normal grasping (the arm/hand model), via interacting with contextual information and sensory feedback. The correct internal model of a tool could then be retrieved by comparing spatial information about muscle commands and tool movements in space with stored models, and the predictions for what would happen. Alternatively the correct internal model of a tool could be selected on each trial by using 'high level' contextual information such as shape and colour. Several studies reported that cerebella activities reflected an internal model of tool in relation to motor control (Imamizu, et al., 2000, 2003; Kluzik, et al., 2008). If the internal model relates to information processing in tool use, brain activities would be found at different brain regions: e.g. parietal lobes (processing sensory information), temporal lobes (object

representation on categorical or memory basis) or at frontal lobes (high-level processing and control), rather not only the cerebella activities (e.g. Lewis, 2006).

A final possibility is that the brain does not model the tool, but instead there is a rapid adaptation, taking place within a single trial. Several studies have reported that visuo-motor adaptation involves a rapid process, which can occur trial-by-trial, and termed as "trial-by-trial adaptation" (Diedrichsen, Hashambhoy, Rane, & Shadmehr, 2005; Fine & Thoroughman, 2007; Thoroughman, Fine, & Taylor, 2007). A previous research showed rapid visuo-motor calibration (Martin, Keating, Goodkin, Bastian, & Thach, 1996); when wearing prism glasses, observers adjusted their throwing direction by rapid adaptation of the optical distortions. Then, after the prism was removed, visuomotor calibration rapidly occurred. In addition, several studies using an illusion paradigm (visual size information) reported that body representations could be modified more rapidly than normal adaptation, and such modified body representations could influence size estimates (Bruno & Bertamini, 2010; Haggard & Jundi, 2009). Bruno and Bertamini (2010) demonstrated that haptic perception was rapidly recalibrated, when a fake hand size was changed. If merely viewing a fake hand affects object size judgements, our visually defined gain tools might somehow affect the size estimates and the sensitivities.

Thus, there are at least two possible mechanisms for remapping in tool use: (i) the brain actually recalibrates hand space and object space on-the-fly, according to knowledge of the tool geometry, or (ii) the brain retrieves an appropriate model from stored multiple internal models, based on contextual information or knowledge of the tool geometry. Here, we suggest further experiments to explore these two possibilities: first, "compute on-the-fly", or second, "memory-based categorisation", or may be both. Each experiment will measure size-discrimination thresholds and rescaling of haptic size estimates, as in our earlier studies, and compare the observed data between conditions.

One condition would require computation of tool gain on a trial-by-trial basis. Different tools would be generated with continuous variations in gain (for example, 0.5:1 to 1.5:1), by manipulating the pivot position to provide more than seven, different-gain tools (exceeding normal human working memory capacity).

Here it should be difficult to learn the tool gains and store them. A second condition would allow categorisation based on tool geometry. A small number of tools would be used, for example, three (0.75:1, 1.0:1 and 1.25:1). In both conditions, tools would not convey any other contextual information in association with gain properties. That is, different gain tools will share properties such as colour and length. If the brain dynamically computes spatial mappings based on the tool geometry, we predict similar performance in cue integration and appropriate rescaling in both conditions because each new trial will be treated the same. If, however, the brain uses different strategies such as memory-based categorisation (for example, 0.75:1 tool is "minified" tool, and so on), we would predict that cue integration and appropriate rescaling will only occur in the second condition, where individual tools can be reliably identified.

If the results support a memory-based hypothesis, we would examine whether contextual cues in relation to a tool type, alone, are sufficient to recover the appropriate internal model. We will not display any tool geometry, but will instead just present different physical properties such as colour (for example, red screen is 0.75:1, blue, 1:1, and green, 1.25:1). In this case, an observer will need to learn about the association between tool gain and colour. If the brain learns the association and can use the information to complete the task, we predict that cue integration and appropriate rescaling will occur. This would show that information about tool geometry is not required on every trial to retrieve the correct model.

These future experiments would evaluate which of the two possibilities is more likely involved in tool use: dynamic computation of spatial mapping or categorisation with knowledge/memory base. In both possibilities, the learning process may be involved, even if calculated on the fly, to reduce uncertainty of the spatial mapping interacting with the tool geometry. Furthermore, the results would provide useful information to answer several related issues such as whether the rescaling process occurs in either early (low) or late (high) levels of processing, which relates to switching or visuo-motor adaptation, and uses different mental processing such as memory and categorisation. In addition, if we use more complex tools, such as opposite movements between hand and tool-tip activities, or not linear remapping from hand to object coordinates, it would be interesting to explore whether

the mapping still occurs. Such a complex tool might show how important predicting tool-tip position is.

6.3.3. Visuo-motor learning process

If tool remapping is not computed on-the-fly, tool geometry must presumably be learned over a period of time. Previous studies provide evidence for this. For example, people show more precise control of a tool, and make better size judgements, through repeated tool use (e.g. Bril, Rein, Nonaka, Weban-Smith, & Dietrich, 2010; Imamizu et al., 2003; Osiurak, Jarry, & Le Gall, 2010). Here we examine whether there is evidence for such learning in our situation.

Does rescaling change through repeated tool usage?

In experiment 5, we showed that different tool gains affected haptic size estimates, but that the rescaling was incomplete. One possibility is that estimates of size with different gain tools are always biased. Alternatively, it could be that our observers did not practice sufficiently with the tools to completely learn the appropriate mappings. To examine this, we used the data in Experiment 7 to see whether perceived size changed throughout the experiment. (Experimental details can be seen in Chapter 5.) We measured perceived size with no conflict, before and after the main block, so we had a measure of the current level of tool rescaling at that point. Three different gain tools were used; so hand opening size was varied according with the tool geometry.

Figure 6.2 shows the perceived-size change across the time course of the experiment. The data are plotted in terms of a "rescaling ratio" (observed perceived sizes divided by the actual physical sizes). Thus, if the ratio is 1.0, it can be considered as perfect rescaling. The rescaling ratio changed through the trials from the beginning (S1) to the end (E2) of the time course. The black arrows in Figure 6.2 show the difference of the ratio between 0.7:1 tool and 1.4:1 tool, and the difference clearly became smaller from S1 to E2. Although the rescaling was still incomplete at

E2, the degree of incomplete scaling gradually approached 1.0 (perfect rescaling) through repeated tool use (although statistical analysis did not show any significant differences through the time courses). Repeated-measures ANOVA analysis between 4 phases (S1, E1, S2, E2) showed significant improvement from S1 to E2, when using 0.7:1 tool, and when using 1.4:1 tool, $F(3, 51) = 3.80$, $p = .016$ (< .05), and also when using 1.4:1 tool, $F(3, 51) = 4.00$, $p = .012$ (< .05), but not for 1:1 tool, $F(3, 51) = .038$, $p = .990$ (NS).

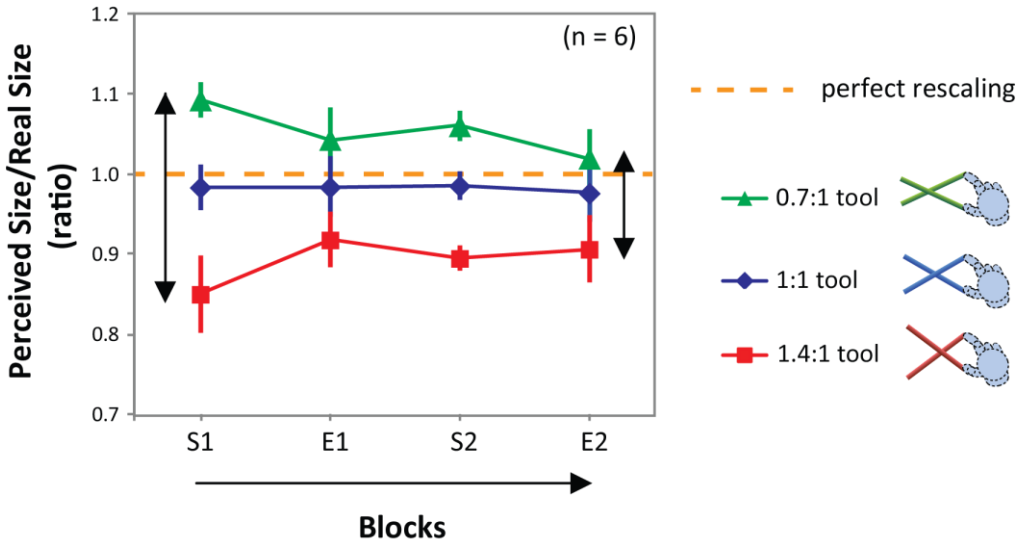


Figure 6.2. Perceived size change through time courses. The data were averaged across six observers. Object sizes were varied: 42, 60, 84 mm at the object coordinates. Three different gain tools were used: 0.7:1, 1:1, and 1.4:1. Four blocks were named as (S1-E1) and (S2-E2), representing Start and End for the first and the second sessions. The vertical axis shows the rescaling ratio, calculating so that perceived sizes were divided by the actual physical sizes. If perceived size was equal to the object size, rescaling ratio is 1.0 (perfect rescaling). The black vertical arrows represent the difference between 0.7:1 tool and 1.4:1 tool. In each block observers completed 20 repetitions of each stimulus level (180 trials per block.) Error bars denote ±1 standard error.

This improvement in rescaling through repeated tool use suggests a learning process. If observers could have explored the object size much longer, or if the tool type did not switch on a trial-by-trial basis, we might expect that the haptic rescaling could become more accurate and approach to the perfect rescaling. One also needs to consider this in relation to the question of whether the rescaling occurred in the sensory estimates or in the inference levels. The learning process might improve the degree of rescaling itself and/or strengthen knowledge of the degree of correlation between vision and haptics. The former could change the bias (and/or variance) of the likelihood and the latter could change the variance of the coupling prior; both would reduce the uncertainty of tool use. Through learning, sensory remapping would approach the perfect, and the correlation between vision and haptics would become strengthened.

In addition, in our experiment (Experiments 4-7), we used simple pliers-like tools. Such tools produce a linear transformation between size at the tool-tip and at the handle. Moreover, when closing the hand opening, the tool-tip size gets smaller, and when widening the hand opening the tool-tip size gets bigger. These movements are quite common in everyday tools such as scissors. That might explain why observers were able to understand how to use the tool after very few practices. Observers might already have a general internal model of such a tool, knowing how to predict the tool-tip position based on the motor command and how to use the tool geometry to solve the correspondence problem appropriately. Such learning effects on tool use can also be seen through child development (van Leeuwen, Smitsman, & van Leeuwen, 1994). It would be interesting, therefore, to explore how and indeed, if more complex tool relationships are so readily learned.

6.3.4. Attention in multisensory integration?

Several studies have claimed that spatial attention plays an important role in multisensory integration (Driver & Spence, 1998; Macaluso & Driver, 2001; Spence & McDonald, 2004). A series of studies have shown how body schema and peripersonal space are changed when using a tool, either shifting or enlarging the

space, mediated by spatial attention (e.g., Bonifazi et al., 2007; Maravita & Iriki, 2004; Holmes et al., 2004b). Although certain effects on the relationship between attention and space coding systems during tool use could be derived from such experiments, an attentional role in cross-modal integration has not yet been identified.

Helbig and Ernst (2008) showed that the visual-haptic cue weighting in size judgements is independent of modality-specific attention. They measured size-discrimination performance, manipulating attention with/without a distractor task for vision- or haptics- alone and vision-plus-haptics and found no effect of the distractor on cue weights. This led them to conclude that cross-modal integration likely occurs at pre-attentive levels. They proposed an early integration model in which cross-modal integration occurs at early (or pre-attentive) rather than late (or post-attentive) processing levels. However, they changed the attentional load rather than manipulating spatial attention. In our discrimination experiment, observers presumably set their attention to the space where the upcoming stimuli at the tool-tip would appear. If we clearly manipulated the spatial attention, for example sometimes at the tool-tip and sometimes at the handle, we may see effects on solving the correspondence problem.

6.3.5. Neural mechanisms in tool use

The observed data are consistent with the MLE cue integration model, suggesting that the final percept would likely result in a process of statistical estimation taking into account cue reliabilities and weights. How does the brain actually achieve this? In other words, how does the brain know about cue reliabilities and weights? It is unlikely that the brain could explicitly compute cue reliabilities and weights from the noisy environment to combine cues on a trial-by-trial basis. In contrast, a computational model provides an explanation of neural population activities effectively resulting in optimal integration of multisensory signals. Thus, reliabilities and weights might be implicitly represented by the neural population code. This probabilistic population coding (PPC) explanation would be biologically plausible and parsimonious than the idea of the explicit computation.

The previous research showed that accumulating noisy sensory neuron responses would form neural population codes. Sensory integration as behavioural outcomes could be explained by these population codes (Deneve, Latham, & Pouget, 2001). Moreover, linearly combining multiple neural population distributions would likely follow the MLE model (Deneve et al., 2001) and the Bayesian inference model (Deneve & Pouget, 2004). The likelihoods of sensory estimates could be computed based on the weighted sum of neural responses and the model can account for the behavioural outcomes (Fetsch, Pouget, DeAngelis, & Angelaki, 2012; Jazayeri & Movshou, 2006). As such, it would not be necessary that cue reliabilities and weights are explicitly encoded to produce the cue integration; rather, it could be directly ruled by neural activities responding to sensory inputs. However, because numerous brain regions are involved in the processing, the neural network of multisensory integration has been not yet fully revealed (Ghazanfar & Schroeder, 2006; Oruç et al., 2003; Welchman, Deubelius, Conrad, Bülthoff, & Kourtzi, 2005).

Neural activities in sensory integration

A number of neurophysiological and behavioural studies have identified several brain regions involved in multisensory integration (for reviews, see Calvert & Thesen, 2004; Stein & Stanford, 2008; see also Calvert, Brammer, & Iversen, 1998). A series of single neuron studies of cats showed that modality specific responses to visual, auditory, and somatosensory stimuli converge on the superior colliculus (SC) neurons in the mid-brain when stimuli are coincident in space and time (e.g., Meredith & Stein, 1983, 1986, 1996; Stein, Huneycutt, & Meredith, 1988; Stein, Meredith, Huneycutt, & McDade, 1989; Wallace, Meredith, & Stein, 1998). Accumulating human and animal research has shown that neural networks of sensory processing are widely projected to various brain regions (e.g. the intraparietal sulcus (IPS), the superior temporal sulcus (STS), and the ventrolateral prefrontal cortex (VLPFC). These would be linked between “early” uni-sensory areas and “late” stages of processing in both interactive and parallel ways (Calvert & Thesen, 2004; Stein & Stanford, 2008). Furthermore, neural population coding could also explain some

sensorimotor transformation in space by a probabilistic approach (Deneve et al., 2001).

Moreover, several studies have suggested that neural population codes could well represent Bayesian processing in the nervous system (Knill & Pouget, 2004; Ma, Beck, & Pouget, 2008a). Neurophysiological findings have shown the specific brain regions engaged in Bayesian-like multisensory processing in the macaque monkey (Gu, Angelaki, & DeAngelis, 2008) and cat (Rowland, Stanford, & Stein, 2007). Computational approaches have also shown that probabilistic neural responses would well reflect outcomes from the probability distributions of Bayesian inference (Ma, Beck, Latham, & Pouget, 2006; Ma et al., 2008a, Ma & Pouget, 2008b). For example, specific neurons responded to the degree of spatial coincidence (Deneve & Pouget, 2004) and spiking signals were observed in the specific time sequence (Deneve, 2008). The similar neural activities might be involved during tool use, if the brain applies this general probabilistic strategy to solve the correspondence problem.

Neural activities in relation to a tool

Previous research suggests that the mental processes of observing tools may involve two neural activities: retrieval and direct inference. This notion is supported by a number of behavioural and brain imaging studies for both clinical patients such as aparaxic (e.g. Goldenberg & Hagmann, 1998) and normal people (e.g. Creem-Regehr & Lee, 2005). A clinical research study investigating an aparaxia patient, who has a neural network disruption as a result of a stroke, showed dissociation between perceptual understanding of tool use and actual tool use when producing a gesture (Heath, Almeida, Roy, Black, & Westwood, 2003). Furthermore, even though observers did not employ actual actions (i.e. they only viewed and imagined tools), neural networks between the dorsal pathway (motor representation) and ventral pathway (object recognition) seemed to be selectively activated in high-level visual-processing systems (e.g., Creem-Regehr & Lee, 2005; Chao & Martine, 2000). These studies indicate that even when just viewing a tool, neural activities are facilitated.

Tool use involves a purposeful action and establishes a strong association in the mental representation between the tool and a target object. Previous research showed that even when simply observing tools and target objects, the objects' spatial features afford the appropriate tool selection, and also tool-use actions evoke the appropriate object selection (Bach et al., 2005; Järveläinen, Schürmann, & Hari, 2004). Several fMRI studies have reported that a region in the middle temporal gyrus selectively responds to tools (e.g. Beauchamp, Lee, Haxby, & Martin, 2003; Chao & Martin, 2000); however, some studies were more sceptical about such region specificity for tools (e.g. Downing, Chan, Peelen, Dodds, & Kanwisher, 2005). Even though specific regions have not yet been identified, a considerable amount of research has shown that in the context of observing either tools or tool use, in both humans and primates, neural networks are likely formed between the temporal, parietal, and frontal cortices, in particular the left cerebral hemisphere, (e.g. Creem-Regehr & Lee, 2005; Goldenberg & Hagmann, 1998; Johnson-Frey, 2004; Lewis, 2006; Peeters, Simone, Nelissen, Fabbri-Destro, Vanduffel, Rizzolatti, & Orban, 2009).

Mirror neurons, which fire when a particular object-directed action is carried out or observed, have been discovered in the ventral premotor cortex of monkeys (e.g. Ferrari, Rozzi, & Fogassi, 2005; Rizzolatti & Craighero, 2004). The mirror-neuron system is considered as an important component of human evolution and development, because it could support social interactions such as language acquisition, and observing and understanding others (Iriki, 2006). Several studies have suggested that because tool use closely relates to action representations, mirror neurons may convey important functions in tool use. For example, action affordance in relation to tool use may involve mirror neuron activities (Iriki, 2006). Ferrari et al. (2005) measured monkey's neural activities and introduced a new concept of "tool-responding mirror neurons", which are selectively responding when a monkey observes tool-use performance such as using a stick or a pair of pliers. They suggested that visual exposure to tool actions results in the formation of a visual association between hand and tool, and this association may assimilate a tool representation into an elongated representation of the hand, in a similar idea to the peripersonal space studies (Iriki et al., 1996). In contrast to these studies, our findings

from adaptation experiment (Control Experiment 1) imply that peripersonal space might not be changed even though the brain likely treats both visual and haptic information at the tool-tip. Further investigation would be required to explore the relationship between such body representation and dynamic tool use performance.

Brain imaging techniques would be useful to identify which regions are involved in spatial remapping or the rescaling process by tool geometry. In our experiment, we measured perceived size using a matching paradigm but could not completely eliminate observers' conscious bias. The imaging technique would be powerful to differentiate between conscious and unconscious level of processing. Such investigation would provide information about which brain regions are involved in solving the correspondence problem and cue integration during tool use, and whether they have a low-level or high level neural processing. Several studies reported that cerebella activities reflected an internal model of tool in relation to the motor control (Imamizu et al., 2000) and various regions were activated at parietal lobes in relation to somatosensory systems and also at temporal lobes and frontal lobes in relation to cognitive processing of tool use (Lewis, 2006). As previously stated whether the brain dynamically computes haptic rescaling according to the tool geometry or categorises spatial mappings using mental representations according to contextual cues, further research might reveal the tool use mechanisms by identifying the related brain regions and comparing with other imaging studies.

6.4. Implications for designing haptic-tool devices

Because humans have expanded their reachable space using various types of tools, investigations into how tool dynamics and geometries affect cue integration is an important applied topic in multisensory processing. Through a series of experiments, we demonstrated haptic sensitivity at the new mapping that could be dynamically changed by the spatial rescaling of hand movements. Therefore, optimal hand opening and the sensitivity is not always optimal at the introduced new coordinates. By examining changes in haptic sensitivity with hand size and tool type,

we could provide practical information for the design of interfaces that manipulate these variables.

Optimal tool design for best haptic sensitivity

There is a considerable amount of tool research to provide an optimal design of hand tools for many purposes. A tool consisting of a handle needs to regard human anthoropometric dimensions of palm and fingers, for example (Baber, 2003). Several factors, such as object stiffness, starting hand posture, and total movable distance to touch the object, could influence the haptic size sensitivity, but they are in close relation with each other. For example, a tool's heaviness and length should be configured, considering its comfortableness and size-weight balance when holding it (e.g. Lee, Kong, Lowe, & Song, 2009). Lee et al. (2009) showed the handle grip span required to optimise finger-specific force capability as a function of hand size. In addition, from motor control perspectives, Berrymann, Yau, & Hsiao (2006) showed that haptic posture (opening between digits) at initial contact with an object influenced haptic size-discrimination threshold. However, for the purpose of manipulating an object precisely, investigating haptic size sensitivity to the target object is important both for normal grasping and tool use.

Ecological approaches used to investigate human perception showed that tool structures of action-referential properties can facilitate and guide grasping action in so-called affordance (e.g. van Leeuwen, Smitsman, & van Leeuwen, 1994 and Osiurak et al., 2010 for a review). This action affordance affects response time and error rates (Phillips & Wward, 2002). Action fluency would influence emotion (Cannon, Hayes, & Tipper, 2010; Hayes, Paul, Beuger, & Tipper, 2008). Although the design of tools frequently considers ergonomics and the comfort of the operator, little consideration is made of favoured tools, as far as I have been aware of it, there is little systematic exploration from visual-haptic integration and correspondence points of view.

We have shown that raw haptic sensitivity does not follow Weber's law; instead, in our observed data the relationship between sensitivity and object size was

fitted by a second degree polynomial equation (Experiment 6). We also showed that the sensitivity in object space was linearly rescaled from sensitivity at the hand, according to the tool gain (Experiment 6). We then demonstrated that the haptic sensitivities, and the cue weights, were altered accordingly when changing the tool gain (Experiment 7). Therefore, these results suggest that an optimal tool gain may exist for particular object sizes.

In this section, we examine how haptic sensitivity might be optimised, using the averaged experimental data from Experiments 6 and 7. According to the results, when using different gain tools, haptic sensitivity to object size is relatively affected by hand opening rather than object size itself. When this sensitivity is linearly rescaled by a tool to the object coordinates, the size to which one is optimally sensitive would alter with the tool gain (See Figure 6.3). Here, we calculate haptic sensitivity with a limitation of hand opening from 20mm to 120mm and with a variation of tool gain from 0.5 to 2.0.

Figure 6.3 (a) plots JNDs as a function of object size and Figure 6.3 (b) plots JNDs as a function of tool gain. Figure 6.3 (c) illustrates how the most sensitive point would be determined by two parameters (object size and tool gain). When using a 1:1 tool, the maximum sensitivity occurred at 43.8mm object size (Experiment 6). In Figure 6.3, the diagonal dashed line represents the changes in the best haptic sensitivity as a function object size and tool gain and indicates that it would be altered depending on object sizes and tool gains. Thus, to produce the best precision, a combination of the parameters should be chosen optimally. Figure 6.3 (d) shows the change if the sensitivity follows Weber's law. In contrast to Figure 6.3 (c), the maximum sensitivity here would be always achieved at the smallest object size regardless of the tool gain (Figure 6.3 (d)). Therefore, in the Weber's case, the optimal combination would not exist.

These graphs indicate that an appropriate tool gain could be chosen for producing highly sensitive performance for the target object size (Figure 6.3 (c)). This suggests that if the range of object sizes is known a haptic interface could be designed, taking into account the sensitivity at the hand. Optimising haptic sensitivity would be beneficial for its precise usage when designing haptic interface geometry.

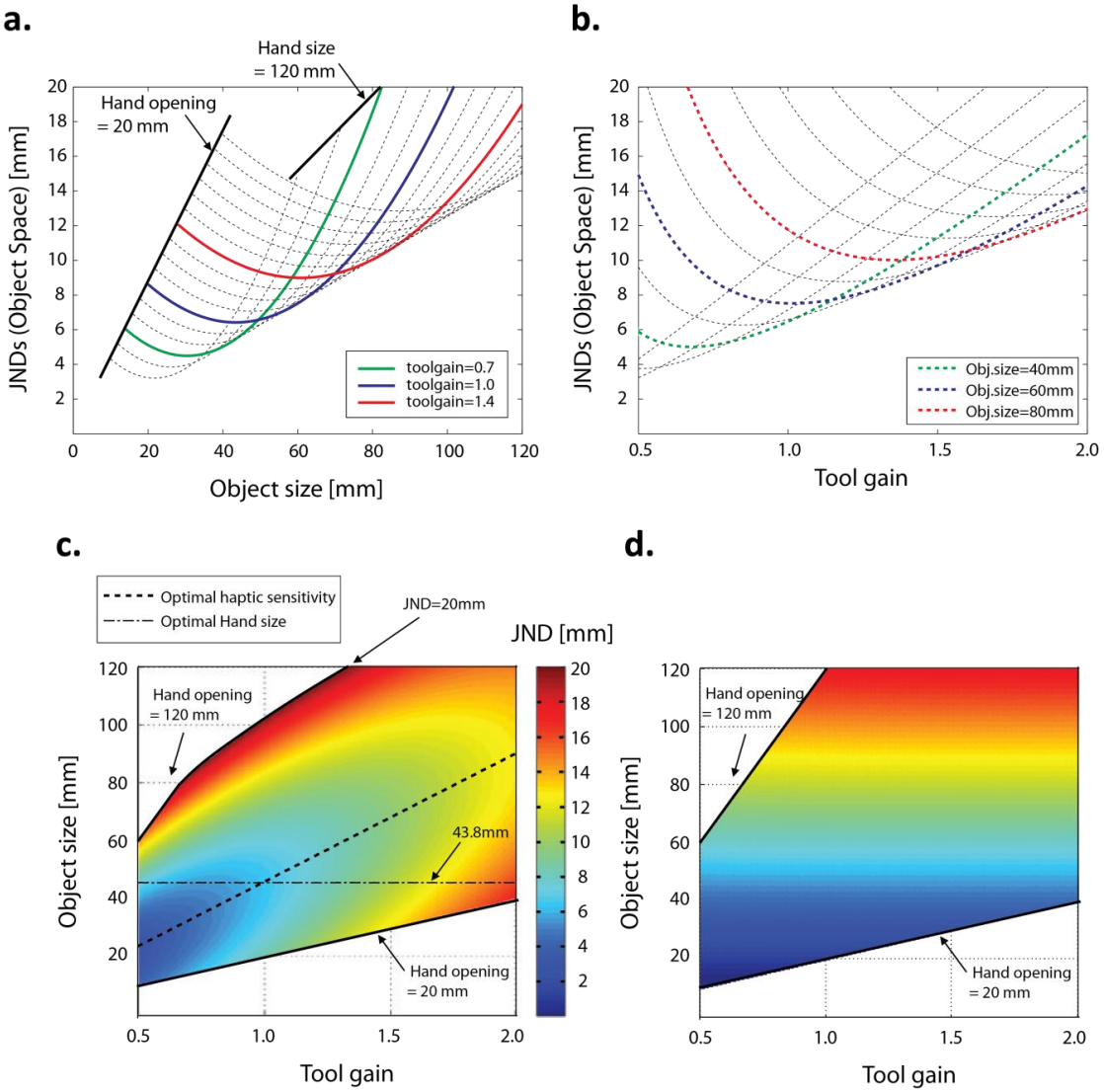


Figure 6.3. Haptic sensitivity changes across varied tool gains and object sizes. Calculation based on the experimental data averaged across six observers. **a.** JNDs in object coordinates as a function of object size with different gain tools. Three legends represent tool gain = 0.7:1 (green), 1:1 (blue), and 1.4:1 (red). **b.** JNDs in object coordinates as a function of tool gain with different object sizes. Three legends represent object size = 40mm (green), 60mm (blue), and 80mm (red). **c.** a 2-D surface plot of JNDs. **d.** a 2-D surface plot of JNDs when following Weber's law.

Design of visual-haptic interfaces

The above calculations are for only haptic sensitivity. Here we examine how the performance changes when combining visual and haptic sensitivities. We assumed that visual sensitivity follows Weber's law. Mathematically, combining linear and polynomial sensitivity curves, the outcome would be polynomial, although the power of the polynomial would reduce and approach to a more near-linear pattern. Therefore a similar pattern to Figure 6.3 (c) can be predicted in the vision-plus-haptics situation. Thus, an optimal value still exists, but with less variation overall. When vision is more reliable than haptics, this optimisation is less important. However, when visual information is less reliable than haptics (for example, in some surgical situations, there is a lack of normal visual cues such as depth and occlusions), maximising or optimising haptic sensitivity would be more important for the precise usage of the device. Therefore, these calculations still would be useful when considering optimising haptic precision in practical tool-use environments. Modern computer-simulated devices often introduce new spatial mappings; therefore, dynamic haptic sensitivity changes, depending on the scaling, should be integrated into the optimal design. Further investigation of human tool use could provide useful information for practical design of visual-haptic devices.

Here, we have explored only one simple case, which was the linear relationship between the tool-tip and the handle opening. However, by extending this, we can expect that the polynomial haptic sensitivity would be an important factor in design, even when using more complex tools. Further examination would be made for different type of tools with different transformation rules (between the tool-tip and the handle). Research about more complex tool geometry would be valuable to find a general rule in tool use. A greater understanding of the underlying mechanisms of solving the correspondence problem in tool use is potentially useful for improving the performance of haptic devices and creating more natural usage. If we provide a novel tool which facilitates perfect rescaling even at the initial use, it could reduce the training period and minimise any risks caused by less integration.

6.5. Summary of contributions

This thesis has explored the brain mechanisms in visual-haptic tool use and discusses these from a general probabilistic standpoint. Our findings support the idea that this general statistical strategy is used to solve the correspondence problem during tool use. We showed that the brain combines visual and haptic information in a statistically-optimal manner during tool use. The haptic size estimates appear to be remapped from hand coordinates to object coordinates. The haptic sensitivity to object size was linearly rescaled from sensitivity at the hand, according to the tool gain. Then we demonstrated that tool geometry affected haptic size estimates, and the weight given to haptics in the combined visual and haptic size estimates.

To explain our results, we examined possible computational models, which provide good accounts of sensory integration. However, whether this could be achieved by constructing a forward or internal model of tool use or involving a more straightforward strategy is still hard to specify. Moreover, the mechanisms would interact with various processes such as switching, memory, learning, and/or adaptation in tool use. The brain might dynamically compute an appropriate coordinate remapping or flexibly establish multiple coordinate systems according to different tool dynamics. Whatever the involved mechanisms are, the spatial remapping seems certainly to occur during tool use, and the remapping would be essential to solve the correspondence problem during tool use.

While the underlying mechanisms in tool use still remains to be determined in several aspects, our results provide instructional information for understanding multisensory integration during tool use and also suggesting an optimal design of visual-haptic interfaces.

6.6. Conclusions

To conclude, the brain integrates visual and haptic information appropriately, in a near-optimal fashion, during tool use, based on the similarity of size estimates in common object coordinates. This may be achieved by remapping haptic estimates on to the object coordinates according to the dynamics and geometry of the tool. To solve the visual-haptic correspondence problem in tool use, the brain is more likely to choose the most probable states in the world, taking into account the spatial mapping process imposed by the tool. As shown through this thesis, it is increasingly important to understand the underlying mechanisms of dynamic rescaling and solving the correspondence problem in the new spatial mapping in tool use.

Appendices

Appendix 1. Individual Performance (Experiment 1: Matching reliability)

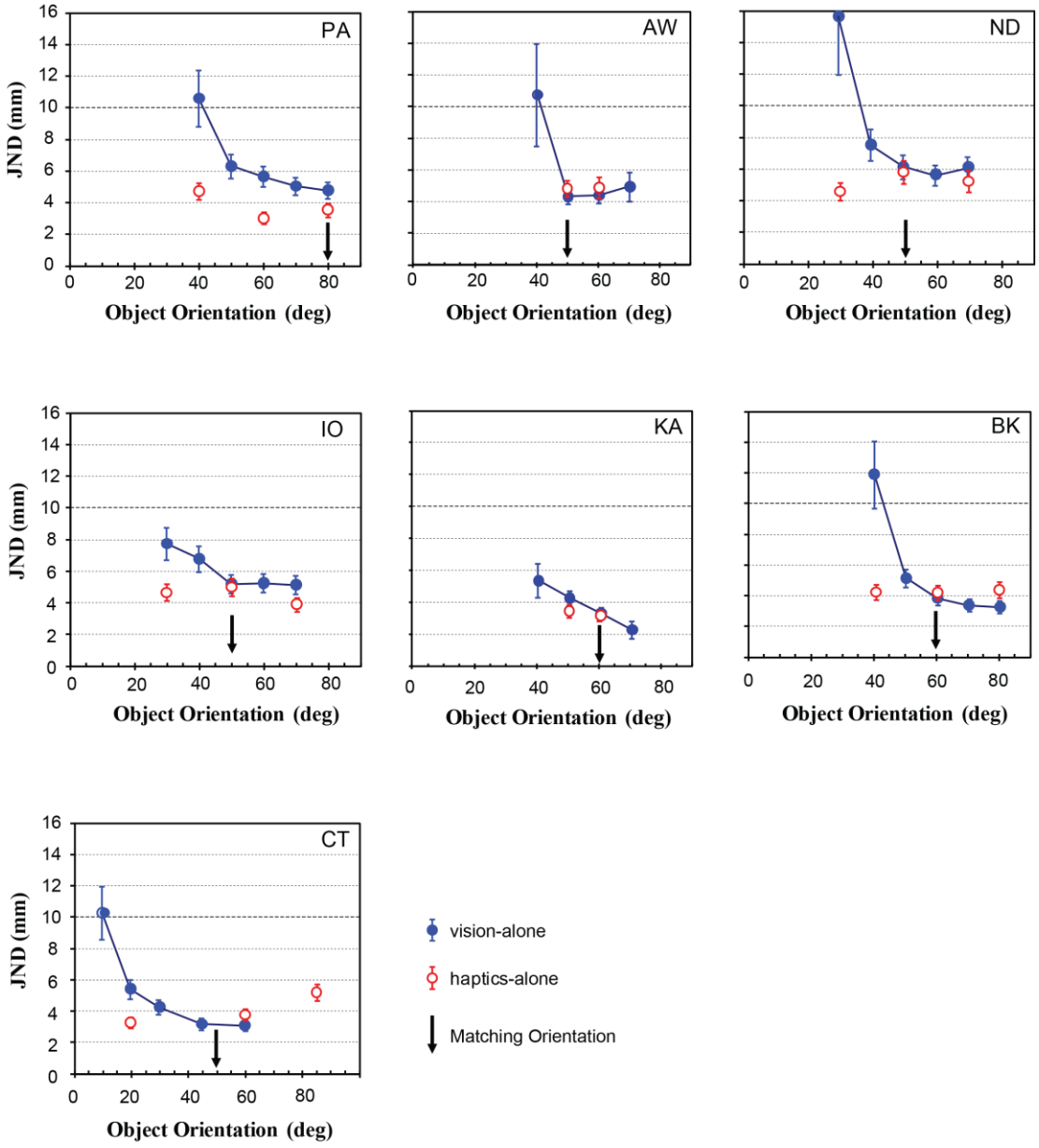
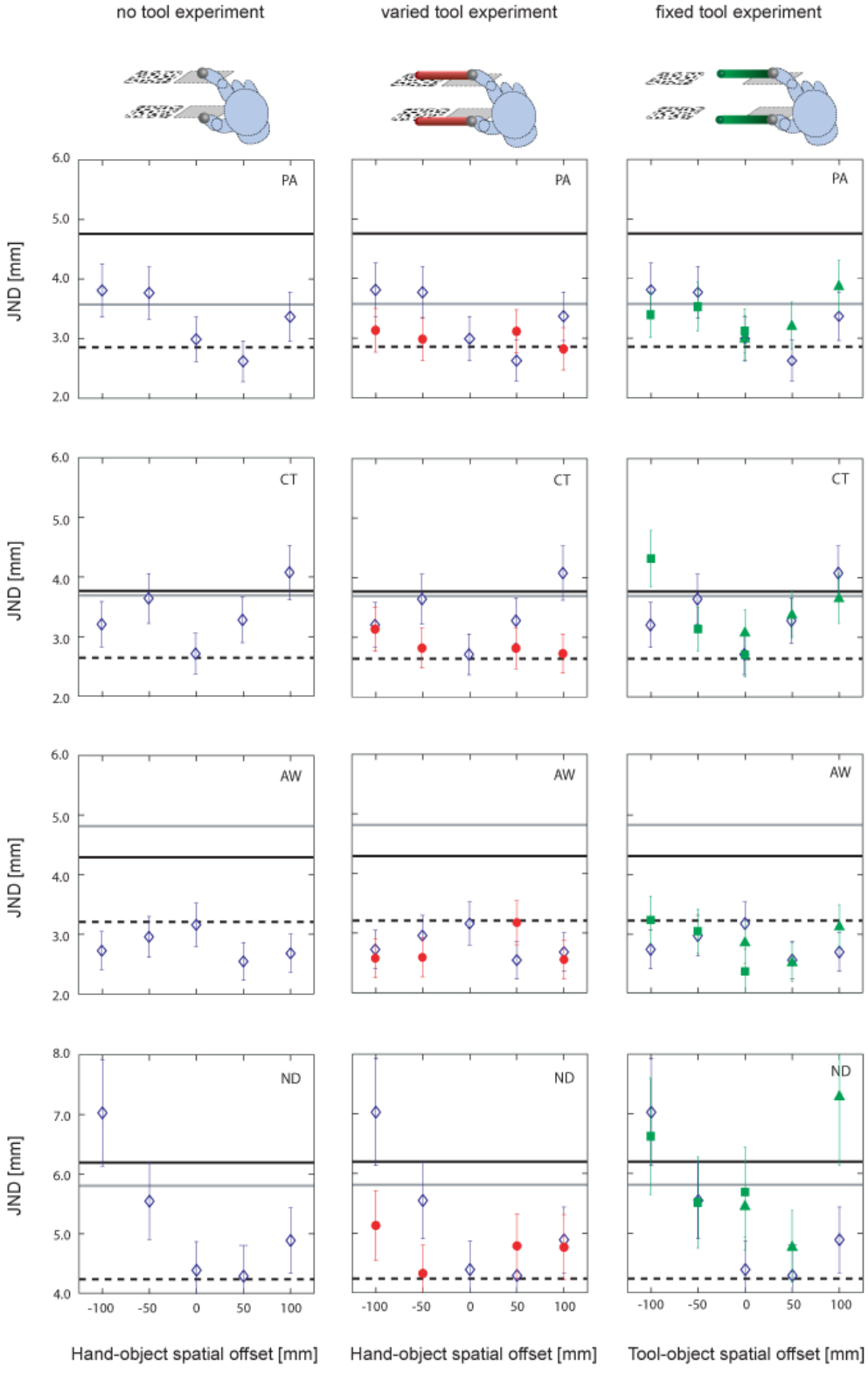


Figure A1. Distance discrimination JNDs in vision-alone and haptic-alone conditions as a function of surface orientation. Error bars denote +/- 1 SEM. The arrow represents the matching orientation for each observer.

Appendix 2. Individual Performance (Experiment 1, 2, 3)

Figure A2 plots individual observer's data from Experiment 1 (no-tool condition), Experiment 2 (zero tool-object offset condition), and Experiment 3 (variable tool-object offset condition). The data are plotted in the same format as Figure 2.8, 2.10, and 2.12, subsequently. (See caption for details).



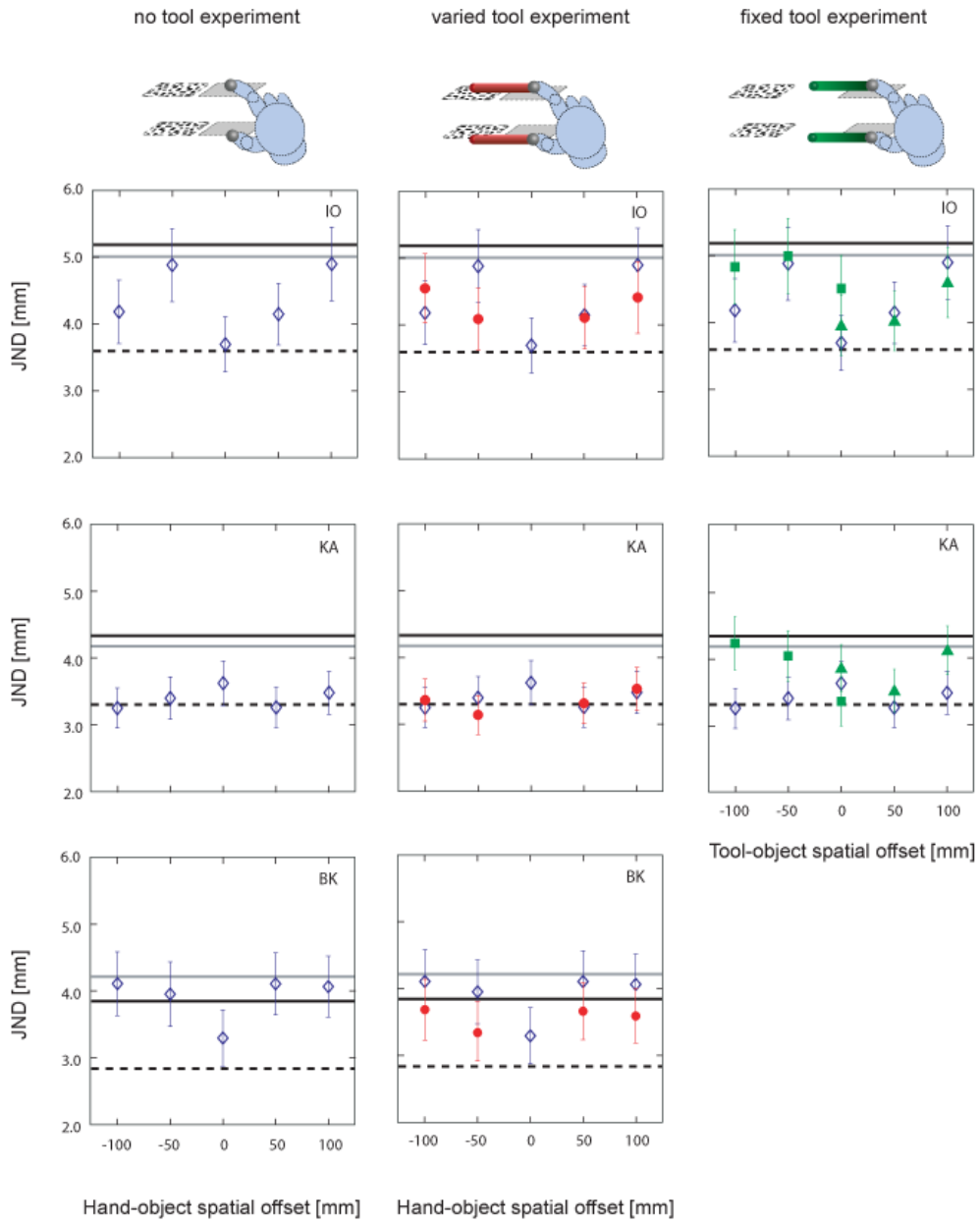
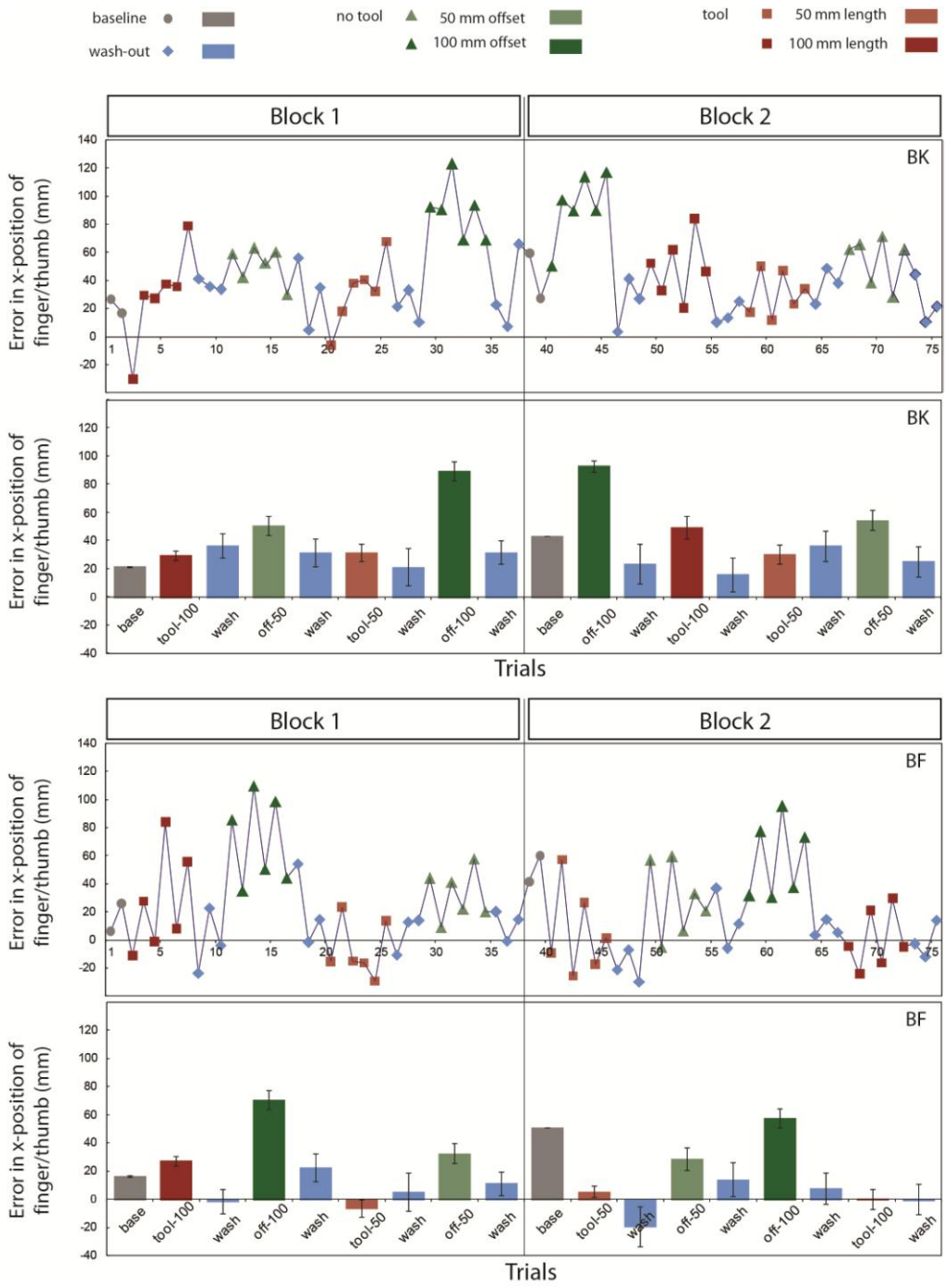


Figure A2. plots individual observer's data from Experiment 1 (no-tool condition), Experiment 2 (tool condition), and Experiment 3 (variable tool-object offset condition). The data are plotted in the same format as Figure 2.8. The effects of hand-object offset on discrimination performance in Experiment 1 (no tool), Experiment 2 (zero tool-object offset), and Experiment 3 (variable tip-offset), plotted separately for each observer. Mean size-discrimination performance (JND) is plotted as a function of spatial offset between the visual and haptic stimuli, (Experiment 1), and as a function of the tool-object offset (the offset between the tool-tip and the visual stimulus, Experiment 3). Negative offsets are denoted by green squares, and positive offsets by green triangles. The black and grey horizontal lines show mean discrimination performance from vision- and haptics-alone, respectively. The dashed line shows performance if the signals are integrated optimally, calculated from the single-modality JNDs using Equation 1. Error bars represent +/- 1 SEM.

Appendix 3. Individual Performance (Control Experiment 1: Does the felt position of the fingers adapt during tool use?)



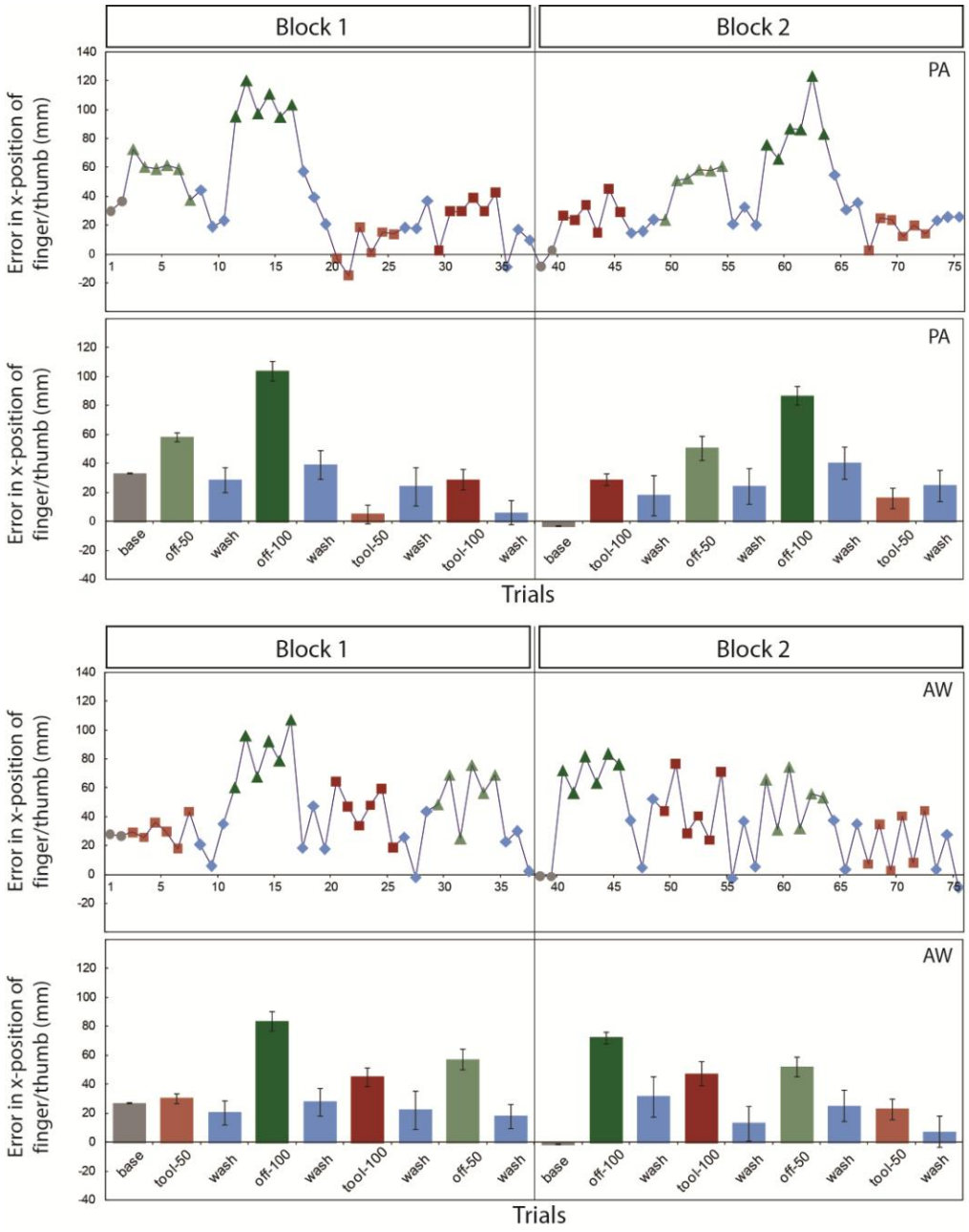


Figure A.3. Individual performances with sequences in Control Experiment 1. Four observers (BK, BF, PA, AW). Line charts show errors in the x-position of the finger and thumb on each test trials and bar charts show the average in each condition.

Appendix 4. Individual Performance (Experiment 4: Matching reliabilities increasing with visual displacement noise)

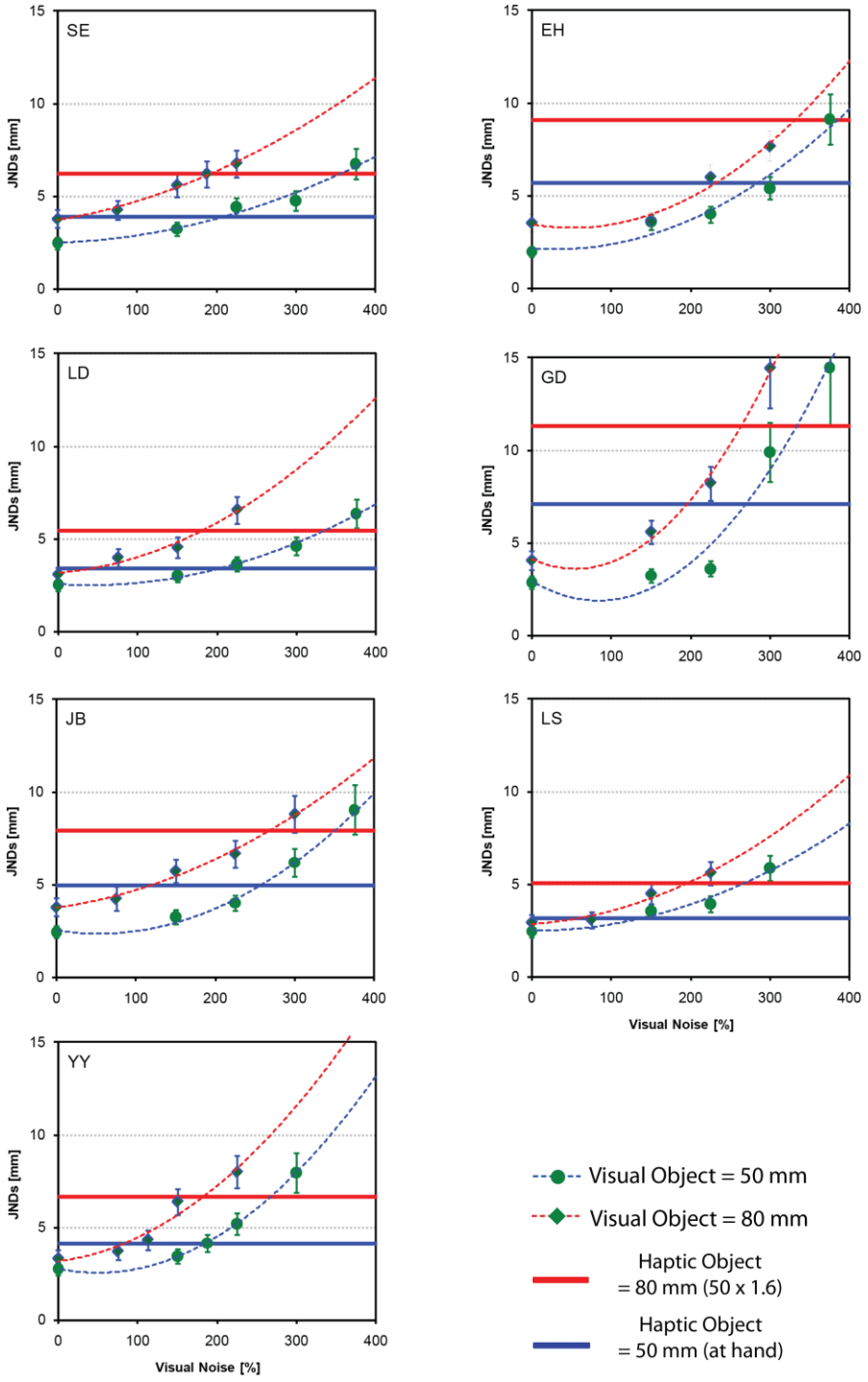


Figure A.4. Plots discrimination performance (JND) of 7 observers for matching visual and haptic reliabilities. Error bars denote ±1 standard error. The dashed fitting lines are second order polynomial curves. The percentages of visual noise were set for the four conditions (Figure 3.1) in the similar manner, as described in Figure 3.4.

Appendix 5. Individual Performance (Experiment 4: four different conditions)

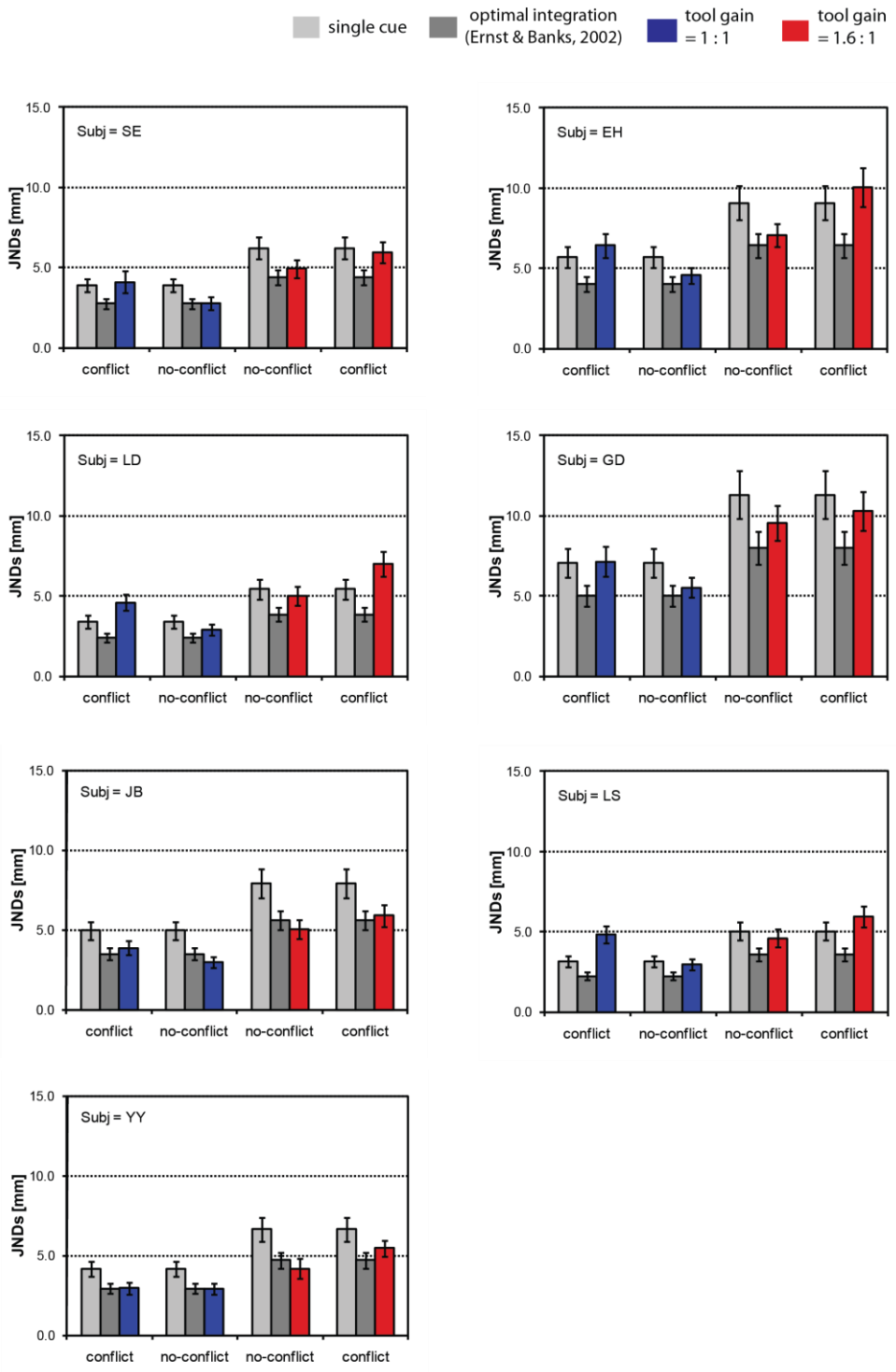


Figure A.5. Plots discrimination performance (JND) of 7 observers in 4 conditions in comparison with single-cue and predicted performance. Error bars denote ±1 standard error.

Appendix 6. Individual Performance (Experiment 6: no tool condition)

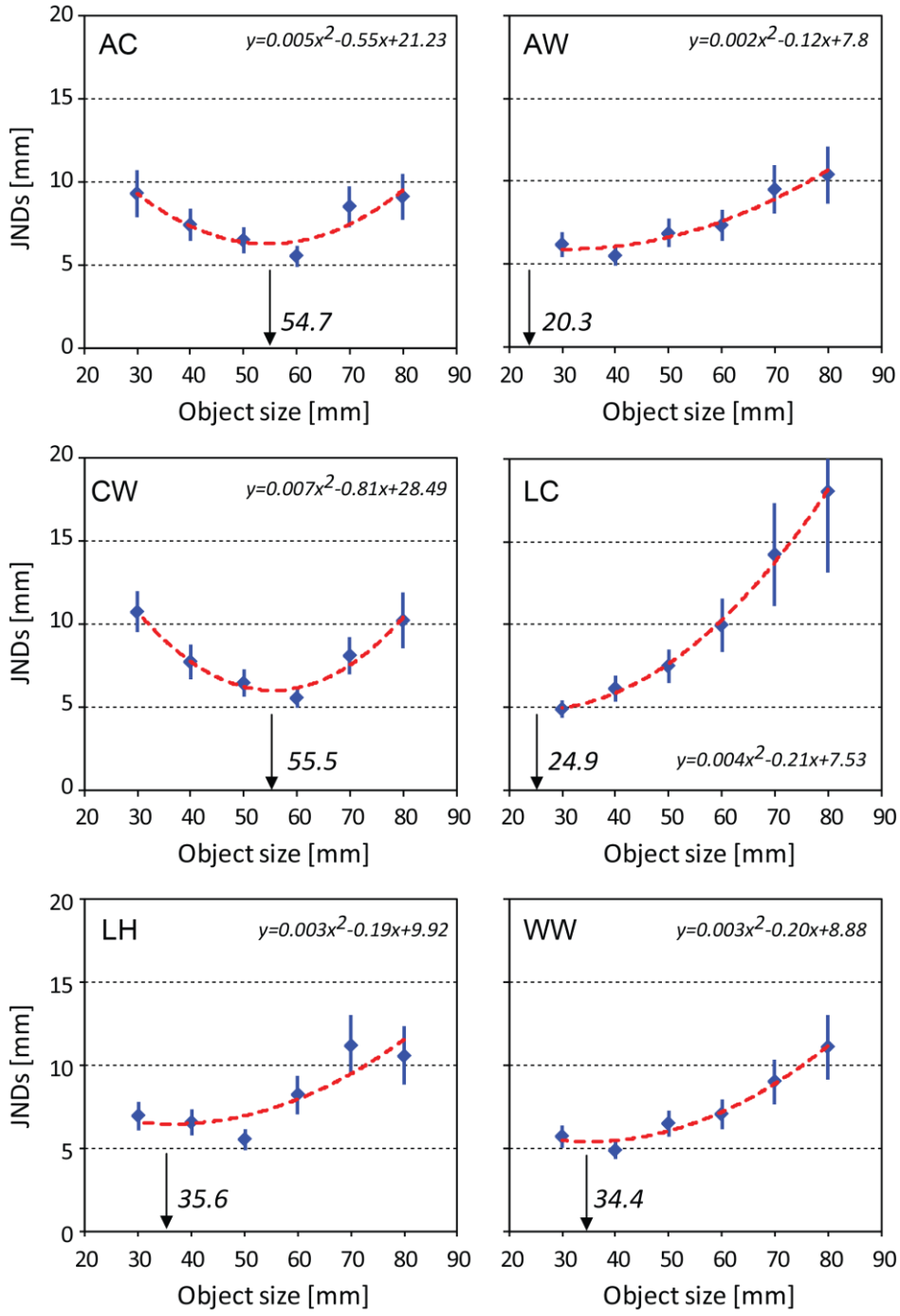


Figure A6. Individual experimental data for 6 observers. Each graph shows haptic sensitivities across object sizes. Object size = Hand opening. Error bars represent +/- 1 SEM.

Appendix 7. Individual Performance (Experiment 6: tool condition)

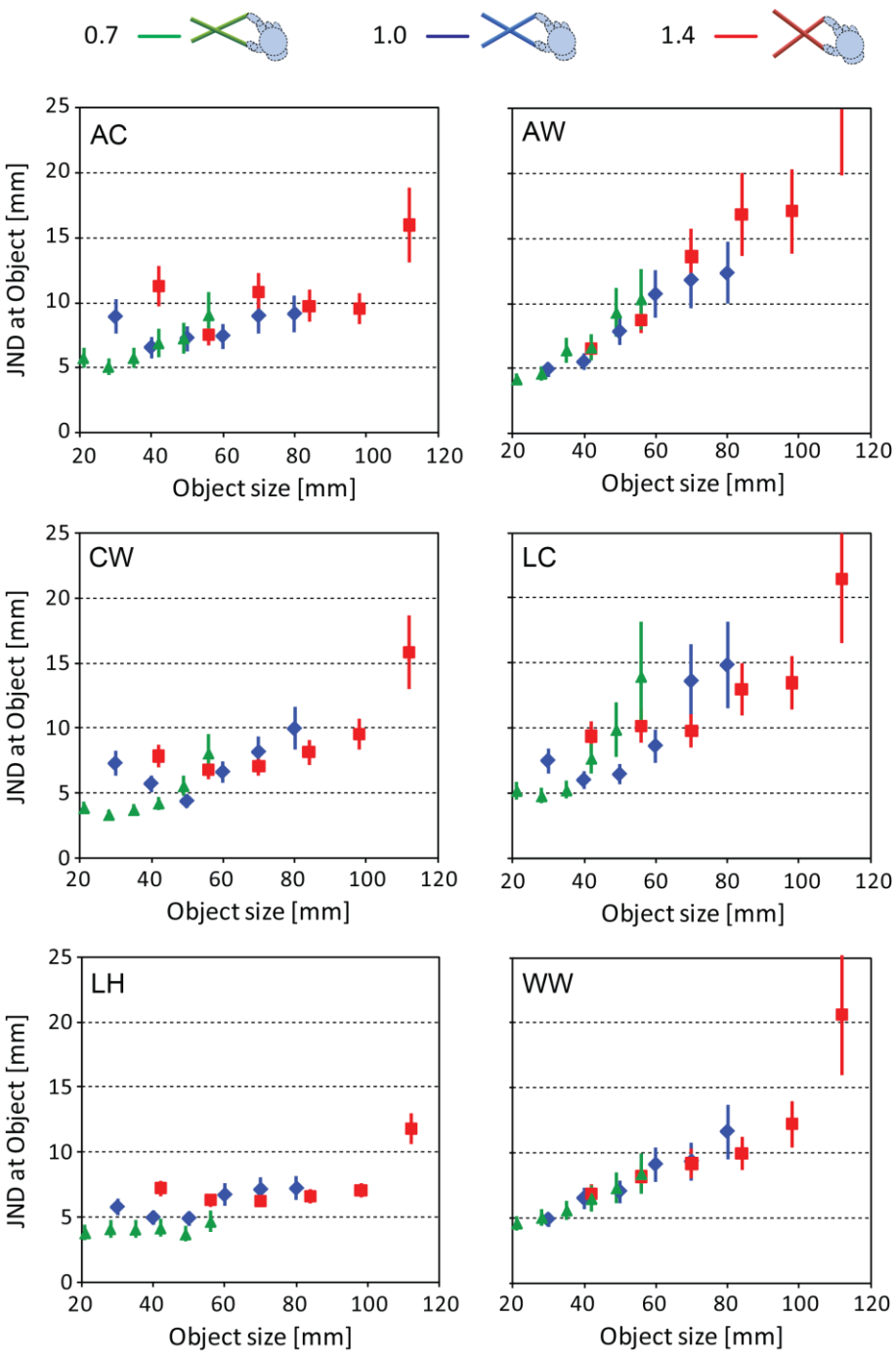


Figure A7. Individual experimental data for 6 observers. Each graph shows haptic sensitivities with three different gain tools across object sizes. Tools are [green=reduced (0.7:1), blue = normal (1:1), red = magnified (1.4:1)] from hand to object space. Object size ≠ Hand opening. Error bars represent +/- 1 SEM.

Appendix 8. Individual Performance (Experiment 7: cue weight change)

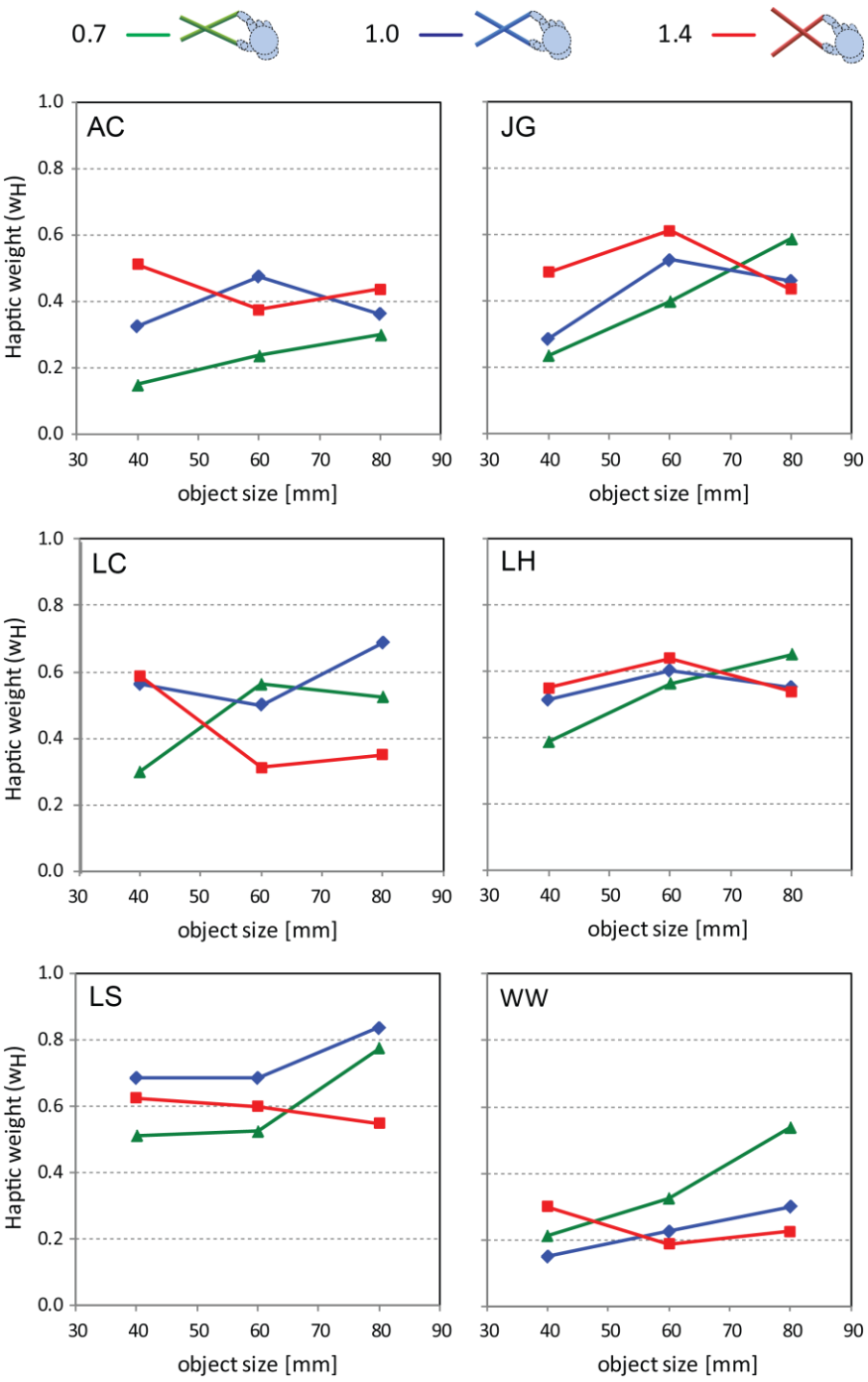


Figure A8. Individual experimental data for 6 observers. Each graph shows haptic cue weight changes across object sizes using three different gain tools. Tools are [green = reduced (0.7:1), blue = normal (1:1), red = magnified (1.4:1)] from hand to object space. Error bars represent +/- 1 SEM.

References

Adams, W.J., Banks, M.S., & van Ee, R. (2001). Adaptation to three-dimensional distortions in human vision. *Nature Neuroscience, 4* (11), 1063-1064.

Alais, D. & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology*, *14* (3), 257-262.

Angelaki, D.E., Gu, Y., & DeAngelis, G.C. (2009). Multisensory integration: psychophysics, neurophysiology, and computation. *Current Opinion in Neurobiology, 19*, 452-458.

Baber, C. (2003). *Cognition and tool use: Forms of engagement in human and animal use of tools.* LONDON: Taylor & Francis.

Bach, P., Knoblich, G., Gunter, T.C., Friederici, A.D., & Prinz, W. (2005). Action comprehension: Deriving spatial and functional relations. *Journal of Experimental Psychology: Human Perception and Performance, 31* (3), 465-479.

Battaglia, P.W., Di Luca, M., Ernst, M.O., Schrater, P.R., Machulla, T., & Kersten, D. (2010). Within- and Cross-modal distance information disambiguate visual size-change perception. *PLoS Computational Biology*, *6* (3), e1000697.

Battaglia, P.W., Jacobs, R.A., & Aslin, R.N. (2003). Bayesian integration of visual and auditory signals for spatial localization. Bayesian integration of visual and auditory signals for spatial localization. *Journal of the Optical of America A, 20* (7), 1391-1397.

Beauchamp, M.S., Lee, K.E., Haxby, J.V., & Martin, A. (2003). fMRI responses to video and point-light displays of moving humans and manipulable objects. *Journal of Cognitive Neuroscience, 17* (7), 991-1001.

Beierholm, U.R., Körding, K.P., Shams, L., & Ma, W.J. (2008). Comparing Bayesian models for multisensory cue combination without mandatory integration. *Advances in Neural Information Processing Systems, 20,* 81-88.

Beierholm, U.R., Quartz, S.R., & Shams, L. (2009). Bayesian priors are encoded independently from likelihoods in human multisensory perception. *Journal of Vision, 9* (5):23, 1-9.

Beisert, M., Massen, C., & Prinz, W. (2010). Embodied rules in tool use: A tool-switching study. *Journal of Experimental Psychology: Human Perception and Performance, 36* (2), 359-372.

Bello, F., Coles, T.R., Gould, D.A., Hughes, C.J., John, N.W., Vidal, F.P., & Watt, S. (2010). The need to touch medical virtual environments? *At Workshop on Medical Virtual Environments, IEEE Virtual Reality, 2010,* Waltham, MA, USA.

Berniker, M., & Körding, K. (2011). Bayesian approaches to sensory integration for motor control. *Wiley Interdisciplinary Reviews: Cognitive Science, 2* (4), 419-428.

Berrymann, L.J., Yau, J.M., & Hsiao, S.S. (2006). Representation of object size in the somatosensory system. *Journal of Neurophysiology, 96,* 27-39.

Berti, A., & Frassinetti, F. (2000). When far becomes near: Remapping of space by tool use. *Journal of Cognitive Neuroscience, 12* (3), 415-420.

Bonifazi, S., Farnè, A., Rinaldesi, L., & Làdavas, E. (2007). Dynamic size-change of peri-hand space through tool-use: Spatial extension or shift of the multi-sensory area. *Journal of Neuropsychology, 1,* 101-114.

Botvinick, M.M., Buxbaum, L.J., Bylsma, L.M., & Jax, S.A. (2009). Toward an integrated account of object and action selection: A computational analysis and empirical findings from reaching-to-grasp and tool-use. *Neuropsychologia, 47,* 671-683.

Bresciani, J., Dammeier, F., & Ernst, M.O. (2006). Vision and touch are automatically integrated for the perception of sequences of events. *Journal of Vision, 6*, 554-564.

Bril., B., Rein, R., Nonaka, T., Weban-Smith, F., & Dietrich, G. (2010). The role of expertise in tool use: skill differences in functional action adaptations to task constraints. *Journal of Experimental Psychology: Human Perception and Performance, 36* (4), 825-839.

Bruno, N., &Bertamini, M. (2010). Haptic perception after a change in hand size. *Brief communication in Neuropsychologia, 48*, 1853-1856.

Burge, J., Fowlkes, C.C., & Banks, M.S. (2010). Natural-scene statistics predict how the figure-ground cue of convexity affects human depth perception. *The Journal of Neuroscience, 30* (21), 7269-7280.

Calvert, G.A., Brammer, M.J., & Iversen, S.D. (1998). Crossmodal identification. *Trends in Cognitive Sciences, 2* (7), 247-253.

Calvert, G.A., & Thesen, T. (2004). Multisensory integration: methodological approaches and emerging principles in the human brain. *Journal of Physiology – Paris, 98*, 191-205.

Cannon, P.R., Hayes, A.E., & Tipper, S.P. (2010). Sensorimotor fluency influences affect: Evidence from electromyography. *Cognition & Emotion*, 24 (4), 681-691.

Cardinali, L., Frassinetti, F., Brozzoli, C., Urquizar, C., Roy, A.C., & Farnè, A. (2009). Tool-use induces morphological updating of the body schema. *Current Biology, 19* (12), R478-R479.

Chao, L.L., & Martin, A. (2000). Representation of manipulable man-made objects in the dorsal stream. *NeuroImage, 12*, 478-484.

Clark, J.J., & Yuille, A.L. (1990). *Data fusion for sensory information processing systems*. Boston: Kluwer Academic Publishers.

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

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

Craddock, M., & Lawson, R. (2009a). The effects of size changes on haptic object recognition. *Attention, Perception, & Psychophysics, 71* (4), 910-923.

Craddock, M., & Lawson, R. (2009b). Size-sensitive perceptual representations underlie visual and haptic object recognition. *PLoS ONE, 4* (11), e8009, 1-10.

Creem-Regehr, S.H., & Lee, J.N. (2005). Neural representations of graspable objects: are tools special? *Cognitive Brain Research, 22*, 457-469.

De Gelder, B., & Bertelson, P. (2003). Multisensory integration, perception and ecological validity. *Trends in Cognitive Sciences, 7* (10), 460-467.

De Preester, H., & Tsakiris, M. (2009). Body-extension versus body-incorporation: Is there a need for a body-model? *Phenomenology and the Cognitive Science, 8*, 307-319.

Debats, N.B., van de Langenberg, R.W., Kingma, I., Smeets, J.B.J., & Beek, P.J. (2010). Exploratory movements determine cue weighting in haptic length perception of handheld rods. *Journal of neurophysiology, 104*, 2821-2830.

Deneve, S. (2008). Bayesian spiking neurons I: Inference. *Neural Computation, 20*, 91-117.

Deneve, S., Latham, P.E., & Pouget, A. (2001). Efficient computation and cue integration with noisy population codes. *Nature Neuroscience, 4* (8), 826-831.

Deneve, S., & Pouget, A. (2004). Bayesian multisensory integration and cross-modal spatial links. *Journal of Physiology – Paris, 98*, 249-258.

Di Luca, M., Ernst, M.O., & Adams, W. (2008). A modal multimodal integration. *Journal of Vision, 8* (6), 526.

Diedrichsen, J., Hashambhoy, Y., Rane, T., & Shadmehr, R. (2005). Neural correlates of reach errors. *The Journal of Neuroscience, 25* (43), 9919-9931.

Downing, P.E., Chen, A.W., Peelen, M.V., Dodds, C.M., & Kanwisher, N. (2005). Domain specificity in visual cortex. *Cerebral Cortex, 16*, 1453-61.

Driver, J., & Spence, C. (1998). Attention and the crossmodal construction of space. *Trends in Cognitive Sciences, 2* (7), 254-262.

Driver, J., & Spence, C. (1999). Cross-modal links in spatial attention. In G.W. Humphreys, J. Duncan, & A. Treisman (Eds.), *Attention, space, and action: studies in cognitive neuroscience* (*p.130-149*). Oxford: Oxford University Press.

Durlach, N.I., Delhorne, L.A., Wong, A., Ko, W.Y., Rabinowitz, W.M., & Hollerbach, J. (1989). Manual discrimination and identification of length by the finger-span method. *Perception & Psychophysics, 46* (1), 29-38.

Ernst, M.O. (2005). A Bayesian view on multimodal cue integration (chapter 6). In G. Knoblich, I.M. Thornton, M. Grosjean, M. Shiffrar (Eds.), *Human body perception from the inside out* (*p.105-131*). New York: Oxford University Press.

Ernst, M.O. (2007). Learning to integrate arbitrary signals from vision and touch. *Journal of Vision, 7* (5):7, 1-14.

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

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

Farnè, A., Bonifazi, S. & Làdavas, E. (2005a). The role played by tool-use and tool-length on the plastic elongation of peri-hand space: A single case study. *Cognitive Neuropsychology, 22* (3/4), 408-418.

Farnè, A., Iriki, A., & Làdavas, E. (2005b). Shaping multisensory action-space with tools: evidence from patients with cross-modal extinction. *Neuropsychologia, 43*, 238-248.

Farnè, A. & Làdavas, E. (2000). Dynamic size-change of hand peripersonal space following tool use. *NeuroReport, Cognitive Neuroscience, 11* (8), 1645-1649.

Farnè, A., Serino, A., & Làdavas, E. (2007). Dynamic size-change of peri-hand space following tool-use: determinants and spatial characteristics revealed through cross-modal extinction. *Cortex, 43*, 436-443.

Fetsch, C.R., Pouget, A., DeAngelis, G.C., & Angelaki, D.E. (2012). Neural correlates of reliability-based cue weighting during multisensory integration. *Nature Neuroscience, 15* (1), 146-154.

Ferrari, P.F., Rozzi, S., & Fogassi,L. (2005). Mirror neurons responding to observation of actions made with tools in monkey ventral premotor cortex. *Journal of Cognitive Neuroscience, 17* (2), 212-226.

Fine, M.S., & Thoroughman, K.A. (2007). Trial-by-trial transformation of error into sensorimotor adaptation changes with environmental dynamics. *Journal of Neurophysiology, 98*, 1392-1404.

Fogassi, L., Gallese, V., Fadiga, L., Luppino, G., Matelli, M., Rizzolatti, G. (1996). Coding of peripersonal space in inferior premotor cortex (Area F4). *Journal of Neurophysiology, 76* (1), 141-157.

Gaiβert, N., Wallraven, C., & Bülthoff, H.H. (2010). Visual and haptic perceptual spaces show high similarity in humans. *Journal of Vision, 10* (11):2, 1-20.

Gepshtein, S., & Banks, M.S. (2003). Viewing geometry determines how vision and haptics combine in size perception. *Current Biology, 13*, 483-488.

Gepshtein, S., Burge, J., Ernst, M.O., & Banks, M.S. (2005). The combination of vision and touch depends on spatial proximity. *Journal of Vision, 5* (11), 7, 1013-1023.

Ghahramani, Z., Wolpert, D.M., & Jordan, M.I. (1996). Generalization to local remappings of the visuomotor coordinate transformation. *The Journal of Neuroscience, 16* (21), 7085-7096.

Ghazanfar, A.A., & Schroeder, C.E. (2006). Is neocortex essentially multisensory? *TRENDs in Cognitive Sciences, 10* (6), 278-285.

Girshick, A.R., & Banks, M.O. (2009). Probabilistic combination of slant information: Weighted averaging and robustness as optimal percepts. *Journal of Vision, 9 (9):8*, 1-20.

Goldenberg, G., & Hagmann, S. (1998). Tool use and mechanical problem solving in apraxia. *Neuropsychologia, 36* (7), 581-589.

Gu, Y., Angelaki, D.E., & DeAngelis, G.C. (2008). Neural correlates of multisensory cue integration in macaque MSTd. *Nature Neuroscience, 11* (10), 1201-1210.

Haggard, P., & Jundi, S. (2009). Rubber hand illusions and size-weight illusions: Self-representation modulates representation of external objects. *Perception, 38,* 1796-1803.

Hayes, A.E., Paul, M.A., Beuger, B., & Tipper, S.P. (2008). Self produced and observed actions influence emotion: The roles of action fluency and eye gaze. *Psychological Research, 72* (4), 461-472.

Heath, M., Almeida, Q.J., Roy, E.A., Black, S.E. & Westwood, D. (2003). Selective dysfunction of tool-use: A failure to integrate somatosensation and action. *Neurocase, 9* (2), 156-163.

Helbig, H.B., & Ernst, M.O. (2007a). Optimal integration of shape information from vision and touch. *Experimental Brain Research, 179* (4), 595-606.

Helbig, H.B., & Ernst, M.O. (2007b). Knowledge about a common source can promote visual-haptic integration. *Perception, 36* (10), 1523-1533.

Helbig, H.B., & Ernst, M.O. (2008). Visual-haptic cue weighting is independent of modality-specific attention. *Journal of Vision, 8* (1):21, 1-16.

Higuchi, S., Imamizu, H., & Kawato, M. (2007). Cerebeller activity evoked by common tool-use execution and imagery tasks: an fMRI study. *Cortex, 43*, 350-358.

Hillis, J.M., Ernst, M.O., Banks, M.S., & Landy, M.S. (2002). Combining sensory information: Mandatory fusion within, but not between, senses. *Science, 298*, 1627-1630.

Hillis, J.M., Watt, S.J., Landy, M.S., & Banks, M.S. (2004). Slant from texture and disparity cues: Optimal cue combination. *Journal of Vision, 4*, 967-992.

Hogervorst, M., & Brenner, E. (2004). Combining cues while avoiding perceptual conflicts. *Perception, 33*, 1155-1172.

Holmes, N.P., Calvert, G.A., & Spence, C. (2004b). Extending or projecting peripersonal space with tools? Multisensory interactions highlight only the distal and proximal ends of tools. *Neuroscience Letters, 372*, 62-67.

Holmes, N.P., Calvert, G.A., Spence, C. (2007a). Tool use changes multisensory interactions in seconds: evidence from the Crossmodal congruency task. *Experimental Brain Research, 183*, 465-476.

Holmes, N.P., Sanabria, D., Calvert, G.A., Spence, C. (2007b). Tool-use: Capturing multisensory spatial attention or extending multisensory peripersonal space? *Cortex, 43*, 469-489.

Holmes, N.P., & Spence, C. (2004a). The body schema and multisensory representation(s) of peripersonal space. *Cognitive Processing, 5*, 94-105.

Holmes, N.P., Spence, C., Hansen, P.C., Mackay, C.E., & Calvert, G.A. (2008). The multisensory attentional consequences of tool use: A functional magnetic resonance imaging study. *PLoS ONE, 3* (10), e3502.

Imamizu, H., Kuroda, T., Miyauchi, Y., Yoshida, T., & Kawato, M. (2003). Modular organization of internal models of tools in human cerebellum. *Proceedings of the National Academy of Sciences, 100* (9), 5461-5466.

Imamizu, H., Miyauchi, S., Tamada, T., Sasaki, Y., Takino, R., Pülz, B., Yoshioka, T., & Kawato, M. (2000). Human cerebellar activity reflecting an acquired internal model of a new tool. *Nature, 403*, 192-195.

Ingram, J.N., Howard, I.S., Flanagan, J.R., & Wolpert, D.M. (2010). Multiple grasp-specific representations of tool dynamics mediate skilful manipulation. *Current Biology, 20* (7), 618-623.

Ingram, J.N., Körding, K.P., Howard, I.S., & Wolpert, D.M. (2008). The statistics of natural hand movements. *Experimental Brain Research, 188*, 223-236.

Iriki, A. (2006). The neural origins and implications of imitation, mirror neurons and tool use. *Current Opinion in Neurobiology, 16*, 660-667.

Iriki, A., Tanaka, M., & Iwamura, Y. (1996). Coding of modified body schema during tool use by macaque postcentral neurones. *NeuroReport, 7*, 2325-2330.

Jackson, C.V. (1953). Visual factors in auditory localization. *The Quarterly Journal of Experimental Psychology, 5* (2), 52-65.

Järveläinen, J., Schürmann, M., & Hari, R. (2004). Activation of the human primary motor cortex during observation of tool use. *NeuroImage, 23*, 187-192.

Jazayeri, M., & Movshon, J.A. (2006). Optimal representation of sensory information by neural populations. *Nature Neuroscience, 9* (5), 690-696.

Johnson-Frey, S.H. (2003). What's so special about human tool use? *Neuron, 39*, 201-204.

Johnson-Frey, S.H. (2004). The neural bases of complex tool use in humans. *Trends in Cognitive Sciences, 8* (2), 71-78.

Kluzik, J., Diedrichsen, J., Shadmehr, R., Bastian, A.J. (2008). Reach adaptation: What determines whether we learn an internal model of the tool or adapt the model of our arm. *Journal of Neurophysiology, 100*, 1455-1464.

Knill, D.C. (1998). Ideal observer perturbation analysis reveals human strategies for inferring surface orientation from texture. *Vision Research, 38*, 2635-2656.

Knill, D.C. (2003). Mixture models and the probabilistic structure of depth cues. *Vision Research, 43* (7), 831-854.

Knill, D.C. (2007a). Robust cue integration: A Bayesian model and evidence from cue-conflict studies with stereoscopic and figure cues to slant. *Journal of Vision, 7* (7):5, 1-24.

Knill, D.C. (2007b). Learning Bayesian priors for depth perception. *Journal of Vision, 7* (8):13, 1-20.

Knill, D.C., & Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends in Neurosciences, 27* (12), 712-719.

Knill, D.C., & Richards, W. (Eds.). (1996). *Perception as Bayesian Inference.* Cambridge University Press.

Knill, D.C., & Saunders, J.A. (2003). Do humans optimally integrate stereo and texture information for judgments of surface slant? *Vision Research, 4*3, 2539-2558.

Körding, K.P., Beierholm, U., Ma, W.J., Quartz, S., Tenenbaum, J.B., & Shams, L. (2007a). Causal inference in multisensory perception. *PLoS ONE, 9*, e943.

Körding, K.P., & Tenenbaum, J.B. (2007b). Causal inference in sensorymotor integration. *Neural Information Processing Systems, 19*, 737-744.

Körding, K.P., & Wolpert, D.M. (2006). Bayesian decision theory in sensorimotor control. *Trends in Cognitive Sciences, 10* (7), 319-326.

Landy, M.S., & Brenner, E. (2001). Motion-disparity interaction and the scaling of stereoscopic disparity. In M.R.M. Jenkin, & L.R. Harris (Eds.), *Vision and Attention*, (*p.131-152*). New York: Springer.

Landy, M.S., & Kojima, H. (2001). Ideal cue combination for localizing texture-defined edges. *Journal of the Optical Society of America, A, 18*, 2307-2320.

Landy, M.S., Maloney, L.T., Johnston, E.B., & Young, M. (1995). Measurement and modelling of depth cue combination: in defense of weak fusion. *Vision Research, 35* (3), 389-412.

Lawson, R. (2009). A comparison of the effects of depth rotation on visual and haptic three-dimensional object recognition. *Journal of Experimental Psychology: Human Perception and Performance, 35* (4), 911-930.

Lee, S., Kong, Y., Lowe, B.D., & Song, S. (2009). Handle grip span for optimising finger-specific force capability as a function of hand size. *Ergonomics, 52*, 5, 601-608.

Lewis, J.W. (2006). Cortical networks related to human use of tools. *The Neuroscientist, 12* (3), 211-231.

Ley, I., Haggard, P., & Yarrow, K. (2009). Optimal integration of auditory and vibrotactile information for judgements of temporal order. *Journal of Experimental Psychology: Human Perception & Performance, 35*, 1005-1019.

Ma, W.J., Beck, J.M., Latham, P.E., & Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nature Neuroscience, 9* (11), 1432-1438.

Ma, W.J., Beck, J.M., & Pouget, A. (2008a). Spiking networks for Bayesian inference and choice. *Current Opinion in Neurobiology, 18*, 1-6.

Ma, W.J., & Pouget, A. (2008b). Linking neurons to behaviour in multisensory perception: A computational review. *Brain Research, 1242*, 4-12.

Macaluso, E., & Driver, J. (2001). Spatial attention and crossmodal interactions between vision and touch. *Neuropsychologia, 39*, 1304-1316.

Magosso, E., Ursino, M., di Pellegrino, G., Làdavas, E., & Serino, A. (2010). Neural bases of peri-hand space plasticity through tool-use: Insights from a combined computational-experimental approach. *Neuropsychologia, 48*, 812-830.

Mamassian, P., Landy, M., & Maloney, L.T. (2002). Bayesian Modelling of Visual Perception. In R.P.N. Rao, B.A. Olshausen, & M.S. Lewicki (Eds.).

Probabilistic Models of the Brain: Perception and Neural Function, Chapter 1 (p. 13-36). Cambridge, MA: MIT Press.

Maravita, A., Clarke, K., Husain, M., & Driver, J. (2002a). Active tool use with the contralesional hand can reduce cross-modal extinction of touch on that hand. *Neurocase, 8*, 411-416.

Maravita, A., Husain, M., Clarke, K., Driver, J. (2001). Reaching with a tool extends visual-tactile interactions into far space: evidence from cross-modal extinction. *Neuropsychologia, 39*, 580-585.

Maravita,A., & Iriki, A. (2004). Tools for the body (schema). *Trends in Cognitive Sciences, 8* (2), 79-86.

Maravita, A., Spence, C., & Driver, J. (2003). Multisensory integration and the body schema: Close to hand and within reach. *Current Biology, 13*, R531-R539.

Maravita, A., Spence, C., & Kennett, S., & Driver, J. (2002b). Tool-use changes multimodal spatial interactions between vision and touch in normal humans. *Cognition, 83*, B25-B34.

Martin, T.A., Keating, J.G., Goodkin, H.P., Bastian, A.J., & Thach, W.T. (1996). Throwing while looking through prisms. I. Focal olivocerebellar lesions impair adaptation. *Brain, 119*, 1183-1198.

Massen, C., & Prinz, W. (2007). Programming tool-use actions. *Journal of Experimental Psychology: Human Perception and Performance, 33* (3), 682-704.

Mehta, B., & Schaal, S. (2002). Forward models in visuomotor control. *Journal of Neurophysiology, 88*, 942-953.

Meredith, M.A., & Stein, B.E. (1983). Interactions among converging sensory inputs in the superior colliculus. *Science, 221*, 389-391.

Meredith, M.A., & Stein, B.E. (1986). Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *Journal of Neurophysiology, 56* (3).

Meredith, M.A., & Stein, B.E. (1996). Spatial determinants of multisensory integration in cat superior colliculus neurons. *Journal of Neurophysiology, 75* (5), 1843-1857.

Meyer, G.F., Wuerger, S.M., Röhrbein, F., & Zetzsche, C. (2005). Low-level integration of auditory and visual motion signals requires spatial co-localisation. *Experimental Brain Research, 166*, 538-547.

Murescaux, J., Leroy, J., Gagner, M., Rubino, F., Mutter, D., Vix, M., Butner, S.E., & Smith, M.K. (2001).Transatlantic robot-assisted telesurgery. *Nature, 413*, 379-380.

Natarajan, R., Murray, I., Shams, L., & Zemel, R.S. (2009). Characterizing response behaviour in multisensory perception with conflicting cues. *In Advances in Neural Information Processing Systems, 21*. Cambridge, MA: MIT Press.

Obayashi, S., Suhara, T., Kawabe, K., Okauchi, T., Maeda, J., Akine, Y., Onoe, H., & Iriki, A. (2001). Functional brain mapping of monkey tool use. *NeuroImage, 14*, 853-861.

Oruç, İ., Maloney, L.T., & Landy, M.S. (2003). Weighted linear cue combination with possibly correlated error. *Vision Research, 43*, 2451-2468.

Osiurak, F., Jarry, C., & Le Gall, D. (2010). Grasping the affordances, understanding the reasoning: toward a dialectical theory of human tool use. *Psychological Review, 117* (2), 517-540.

Peeters, R., Simone, L., Nelissen, K., Fabbri-Destro, M., Vanduffel, W., Rizzolatti, G., & Orban, G.A. (2009). The representation of tool use in humans and monkeys: common and uniquely human features. *The Journal of Neuroscience, 29* (37), 11523-11539.

Pettypiece, C.E., Goodale, M.A., & Culham, J.C. (2010). Integration of haptic and visual size cues in perception and action revealed through cross-modal conflict. *Experimental Brain Research, 201*, 863-873.

Popescu, V.G., Burdea, G.C., Bouzit, M., & Hentz, V.R. (2000). A virtual-reality-based telerehabilitation systemwith force feedback. *IEEE Transactions on information technology in biomedicine, 4* (1), 45-51.

Povinelli, D.J., Reaux, J.E., & Frey, S.H. (2010). Chimpanzees' context-dependent tool use provides evidence for separable representations of hand and tool even during active use within peripersonal space. *Neuropsychologia, 48*, 243-247.

Redding, G.M., & Wallace, B. (1997). *Adaptive spatial alignment.* Mahwah, NJ: Erlbaum.

Rizzolatti , G. & Craighero, L. (2004). The mirror-neuron system. *Annual Review Neuroscience, 27*, 169-192.

Roach, N.W., Heron, J., & McGraw, P.V. (2006). Resolving multisensory conflict: a strategy for balancing the costs and benefits of audio-visual integration. *Proceedings of the Royal Society, 273*, 2159-2168.

Robles-De-La-Torre, G. (2006). The importance of the sense of touch in virtual and real environments. *Multimedia, IEEE Computer Society, 13* (3), 24-30.

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

Rosas, P., Wagemans, J., Ernst, M.O., & Wichmann, F.A. (2005). Texture and haptic cues in slant discrimination: reliability-based cue weighting without statistically optimal cue combination. *Journal of the Optical Society of America A, 22* (5), 801-809.

Rosas, P., Wichmann, F.A., & Wagemans, J. (2007). Texture and object motion in slant discrimination: Failure of reliability-based weighting of cues may be evidence for strong fusion. *Journal of Vision, 7* (6), 3, 1-21.

Rowland, B., Stanford, T., Stein, B. (2007). A Bayesian model unifies multisensory spatial localization with the physiological properties of the superior colliculus. *Experimental Brain Research, 180*, 153-161.

Sato, Y., Toyoizumi, T., & Aihara, K. (2007). Bayesian inference explains perception of unity and ventriloquism aftereffect: identification of common sources of audiovisual stimuli. *Neural Computation, 19* (12), 3335-3355.

Schindel. R., & Arnold, D.H. (2010). Visual sensitivity can scale with illusory size changes. *Current Biology, 20*, 841-844.

Schwartz, D.L., & Holton, D.L. (2000). Tool use and the effect of action on the imagination. *Journal of Experimental Psychology, Learning, Memory, and Cognition, 26* (6), 1655-1665.

Scott, S., & Gray, R. (2010). Switching tools: perceptual-motor recalibration to weight changes. *Experimental Brain Research, 201*, 177-189.

Shams, L., & Beierholm, U. (2010). Causal inference in perception. *Trends in Cognitive Science, 14* (9), 425-432.

Shams, L., Kamitani, Y., & Shimojo, S. (2002). Visual illusion induced by sound. *Cognitive Brain Research, 14*, 147-152.

Shams, L., Ma, W.J., & Beierholm, U. (2005). Sound-induced flash illusion as an optimal percept. *NeuroReport, 16* (17), 1923-1927.

Smeets, J.B.J., van den Dobbelsteen, J.J., de Grave, D.D.J., van Beers, R.J., & Brenner, E. (2006). *PNAS (Proceedings of the National Academy of Sciences of the United States of America), 103* (49), 18781-18786.

Spence, C., & Driver, J. (2000). Attracting attention to the illusory location of a sound: reflexive crossmodal orienting and ventriloquism. *Neuroreport, 11* (9), 2057-2061.

Spence, C., & McDonald, J. (2004). The cross-modal consequences of the exogenous spatial orientation of attention. In G.A. Calvert, C. Spence, & B.E. Stein (Eds.), *The handbook of multisensory processes. (p.3-26).* The MIT Press.

Stein, B.E., Huneycutt, W.S., & Meredith, M.A. (1988). Neurons and behaviour: the same rules of multisensory integration apply. *Brain Research, 448,* 355-358.

Stein, B.E., Meredith, M.A., Huneycutt, W.S., & McDade, L. (1989). Behavioral indices of multisensory integration: Orientation to visual cues is affected by auditory stimuli. *Journal of Cognitive Neuroscience, 1* (1), 12- 24.

Stein, B.E., & Stanford, T.R. (2008). Multisensory integration: current issues from the perspective of the single neuron. *Nature Neuroscience, 9,* 255-256.

Stevens, S.S. (1957). On the psychophysical law. *The Psychological Review, 64* (3), 153-181.

Stevens, S.S., & Stone, G. (1959). Finger span: ratio scale, category scale, and JND scale. *Journal of Experimental Psychology, 57* (2), 91-95.

Takahashi, C., Diedrichsen, J., & Watt, S.J. (2009). Integration of vision and haptics during tool use. *Journal of Vision, 9* (6):3, 1-13.

Thoroughman, K.A., Fine, M.S., & Taylor, J.A. (2007). Trial-by-trial motor adaptation: a window into elemental neural computation. In P. Cisek, T. Drew, & J.F. Kalaska (Eds.), *Progress in Brain Research, 165,* 373-382.

Trommershäuser, J., Körding, K.P., & Landy, M.S. (Eds.). (2011). *Sensory Cue Integration.* New York: Oxford University Press.

Trommershäuser, J., Maloney, L.T., & Landy, M.S. (2003). Statistical decision theory and the selection of rapid, goal-directed movements. *Journal of the Optical Society of America A, 20* (7), 1419-1433.

Trommershäuser, J., Maloney, L.T., & Landy, M.S. (2008). Decision making, movement planning and statistical decision theory. *Trends in Cognitive Sciences, 12* (8), 291-297.

Van Doren, C.L. (1995). Cross-modality matches of finger span and line length. *Perception & Psychophysics, 57* (4), 555-568.

van Leeuwen, L., Smitsman, A., & van Leeuwen, C. (1994). Affordances, perceptual complexity, and the development of tool use. *Journal of Experimental Psychology: Human Perception and Performance, 20* (1), 174-191.

van Wassenhove, V., Grant, K.W., & Poeppel, D. (2007). Temporal window of integration in auditory-visual speech perception. *Neuropsychologia, 45,* 598-607.

Wallace, M.T., Meredith, M.A., & Stein, B.E. (1998). Multisensory integration in the superior colliculus of the alert cat. *Journal of Neurophysiology, 80*, 1006-1010.

Warren, D.H. & Cleavers, W.T. (1971).Visual-proprioceptive interaction under large amounts of conflict. *Journal of Experimental Psychology, 90* (2), 206-214.

Wei, K., Körding, K.P. (2011). Causal inference in sensorimotor learning and control. In J. Trommershäuser, K.P. Körding, & M.S. Landy (Eds.), *Sensory Cue Integration. Chapter 2 (p. 30-45).* New York: Oxford University Press.

Welchman, A.E., Deubelius, A., Conrad, V., Bülthoff, H.H., & Kourtzi, Z. (2005). 3D shape perception from combined depth cues in human visual cortex. *Nature Neuroscience, 8* (6), 820-827.

Wichmann, F.A., & Hill, N.J. (2001a). The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception & Psychophysics, 63* (8), 1293-1313.

Wichmann, F.A., & Hill, N.J. (2001b). The psychometric function: II. Bootstrap-based confidence intervals and sampling. *Perception & Psychophysics, 63* (8), 1314-1329.

Witkin, H.A., Wapner, S., & Leventhal, T. (1952). Sound localization with conflicting visual and auditory cues. *Journal of Experimental Psychology, 43,* 58-67.

Wolpert, D.M., & Ghahramani, Z. (2000). Computational principles of movement neuroscience. *Nature Neuroscience, 3*, 1212-1217.

Wolpert, D.M., Ghahramani, Z., & Jordan, M.I. (1995). An internal model for sensorimotor integration. *Science, 269*, 1880-1882.

Woods, T.M., & Recanzone, G.H. (2004). Visually induced plasticity of auditory spatial perception in Macaques. *Current Biology, 14*, 1559-1564.

Young, M.J., Landy, M.S., & Maloney, L.T. (1993). A perturbation analysis of depth perception from combination of texture and motion cues. *Vision Research, 33* (18), 2685-2696.

Yuille, A.L., & Bülthoff, H.H. (1996). Bayesian decision theory and psychophysics. In D. Knill, & W. Richards (Eds.), *Bayesian Approaches to Perception* (*p.123-161*). Cambridge University Press.