

Visual-haptic cue weighting is independent of modality-specific attention

Hannah B. Helbig

Max Planck Institute for Biological Cybernetics,
Tübingen, Germany



Marc O. Ernst

Max Planck Institute for Biological Cybernetics,
Tübingen, Germany



Some object properties (e.g., size, shape, and depth information) are perceived through multiple sensory modalities. Such redundant sensory information is integrated into a unified percept. The integrated estimate is a weighted average of the sensory estimates, where higher weight is attributed to the more reliable sensory signal. Here we examine whether modality-specific attention can affect multisensory integration. Selectively reducing attention in one sensory channel can reduce the relative reliability of the estimate derived from this channel and might thus alter the weighting of the sensory estimates. In the present study, observers performed unimodal (visual and haptic) and bimodal (visual-haptic) size discrimination tasks. They either performed the primary task alone or they performed a secondary task simultaneously (dual task). The secondary task consisted of a same/different judgment of rapidly presented visual letter sequences, and so might be expected to withdraw attention predominantly from the visual rather than the haptic channel. Comparing size discrimination performance in single- and dual-task conditions, we found that vision-based estimates were more affected by the secondary task than the haptics-based estimates, indicating that indeed attention to vision was more reduced than attention to haptics. This attentional manipulation, however, did not affect the cue weighting in the bimodal task. Bimodal discrimination performance was better than unimodal performance in both single- and dual-task conditions, indicating that observers still integrate visual and haptic size information in the dual-task condition, when attention is withdrawn from vision. These findings indicate that visual-haptic cue weighting is independent of modality-specific attention.

Keywords: attention, multisensory integration, dual task, vision, touch

Citation: 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, <http://journalofvision.org/8/1/21/>, doi:10.1167/8.1.21.

Introduction

Multiple sensory systems provide information about objects we encounter in our environment. For example, we use our sense of touch as well as our visual sense to gather information about the shape or size of an object. The brain integrates such redundant information acquired from different senses into a coherent percept to come up with the most reliable (unbiased) estimate (e.g., for a review, see Ernst & Bühlhoff, 2004). That is, the nervous system integrates noisy sensory information such that the variance (σ^2) of the unified multimodal estimate is maximally reduced. Given that the noise distributions associated with the individual estimates are independent and Gaussian, the optimal combined estimate (maximum likelihood estimate, MLE) is a linear combination of the individual unimodal estimates that are weighted by their relative reliabilities ($1/\sigma^2$); more reliable cues are assigned a larger weight (cf. Method section). Recent research showed that the optimal cue combination rule (MLE) predicts observers' behavior for a variety of perceptual tasks and sensory modalities (e.g., audiovisual localization:

Alais & Burr, 2004; visual-haptic size perception: Ernst & Banks, 2002; Gepshtein & Banks, 2003; visual-haptic shape perception: Helbig & Ernst, 2007b; audio-tactile event perception: Bresciani, Dammeier, & Ernst, 2006; Shams, Ma, & Beierholm, 2005; visual-proprioceptive localization: van Beers, Wolpert, & Haggard, 2002).

In the present study, we investigated whether visual-haptic cue integration is affected by withdrawing attention selectively from one sensory channel. Attention was modulated by engaging observers in an attention-demanding secondary task while they performed the primary task (size discrimination). This distractor task exerted a selectively stronger effect on processes of the visual as compared to the haptic modality. Diverting attention away from a sensory modality is assumed to increase the variance of the information provided by this sensory channel (Prinzmetal, Amiri, Allen, & Edwards, 1998; Prinzmetal, Nwachuku, Bodanski, Blumenfeld, & Shimizu, 1997) and might therefore result in a reduction of weight attributed to this sensory modality. However, if multisensory integration takes place automatically at a pre-attentive level of processing such an attentional modulation should not affect cue weighting. There is evidence for both hypotheses reported in the literature.

In favor of the idea that attention can affect multi-sensory integration, it was found that observers report different percepts when asked to respond to one sensory modality versus another sensory modality (e.g., Bertelson & Radeau, 1981; Massaro, 1998; Warren, Welch, & McCarthy, 1981). For example, Massaro (1998) studied perception of bimodal (facial and vocal) emotional expressions and found increased influence of the modality to which attention is directed. Moreover, a study by Alsius, Navarra, Campbell, and Soto-Faraco (2005) was taken as evidence that multisensory integration depends on attentional resources. They presented observers with a videotape of a talking head. In some trials the word visually articulated by lip movements (visual stimulus) and the spoken word (auditory stimulus) mismatched, leading to an illusory McGurk percept as a result of audiovisual integration. Participants had to report what the speaker said. In some conditions, participants additionally had to perform an auditory or visual distractor task concurrently. The proportion of McGurk percepts, i.e., visually influenced percepts, was reduced in the dual-task conditions. Based on this finding, Alsius et al. concluded that audiovisual integration depends on attentional resources and breaks down under high attentional load.

In contrast, there is also a growing body of evidence for the automatic, pre-attentive nature of multisensory integration. Some studies found that multisensory integration can help attentional selection, and therefore it was concluded that multisensory integration occurs before attentional selection is completed (e.g., Driver, 1996; Soto-Faraco, Navarra, & Alsius, 2004; Spence & Driver, 2000). For example, Driver (1996) presented two speech sounds simultaneously, at the same location. In addition, a videotape of a talking head was shown. Lip movements were in sync with one of the two speech sounds. Observers were instructed to selectively listen to the target message (sound speech accompanied by the visual speech). He found less interference when the movie was presented at a location different from the auditory signals as opposed to when the visual speech was presented at the same location as the two auditory speech signals. He argues that visual–auditory integration (ventriloquism) pulls the apparent location of the target message away from the location of the distractor message and thereby facilitates attentional selection. Therefore, audiovisual integration must arise prior to the processes of attentional selection. However, these studies do not directly investigate the effects of attention on the process of multimodal integration. Further evidence for the automatic nature of multisensory integration comes from studies that focused on spatial attention to the location of the sensory input in one modality (e.g., Bertelson, Vroomen, Driver, & De Gelder, 2000; Vroomen, Bertelson, & de Gelder, 2001a). From these results, it was concluded that ventriloquism (visual bias on the perceived auditory location), i.e., audiovisual integration, does not depend on endogenous or exogenous spatial attention to the location of the visual

event. While most of the studies focused on audiovisual integration, Shore and Simic (2005) found that top-down influences do not affect multisensory integration in the visuo-tactile domain.

More directly related to the purpose of the present experiments are a number of studies that examined the effect of selective attention to one specific sensory modality (e.g., Bresciani et al., 2005, 2006; De Gelder & Vroomen, 2000; Helbig & Ernst, 2007a; Massaro, 1987a, 1987b; Shams, Kamitani, & Shimojo, 2000, 2002; Spence, Pavani, & Driver, 2004; Spence & Walton, 2005; Vroomen, Driver, & de Gelder, 2001b). In these studies, observers were presented with slightly conflicting multi-sensory stimuli and were instructed to respond to the stimulus in one sensory modality while ignoring the other. The reported percept was always strongly biased by the input in the task-irrelevant modality indicating that multisensory information is automatically integrated even when the sensory input from one modality is explicitly task irrelevant. However, by merely instructing observers to report input from one modality, it cannot be ensured that indeed more attention is directed to the to-be-attended modality. In addition, most of these experiments were conducted under conditions of low attentional load. It has been argued that under such conditions, it is particularly hard to ignore an irrelevant stimulus (e.g., Lavie, 1995). It is possible that remaining attentional resources were used to process the irrelevant stimulus that then interfered with the to-be-attended stimulus.

To summarize, on the one hand some studies observed attentional effects on multisensory integration. On the other hand, there is a large number of studies providing evidence for the automaticity of multisensory integration. However, none of the previous studies modulated the degree of attention directed the sensory modalities while quantitatively testing whether multisensory information is integrated.

Thus, we here apply a dual-task paradigm to quantitatively examine the effect of selectively attending to one over the other modality. In the primary task, participants performed a visual and/or haptic size discrimination task. Participants' performance in this task was compared to an ideal observer (MLE model) to test for optimal integration. In the distractor task, observers had to perform a same/different judgment of two rapidly presented visual letter sequences. As we show in the result section, performing this concurrent task withdraws selectively more attention from the visual than from the haptic primary task. That is, it selectively increases the variance of the visual estimate while the haptic estimate is less affected. Participants' performance in the primary task alone (single task) was compared to the performance obtained when the distractor task was carried out simultaneously (dual task). To test whether visual-haptic integration is susceptible to modality-specific attention or whether it is an early, automatic process independent of modality-specific attention, we propose two optimal integration

models (late and early integration model, see Figure 1) and test the experimental data against these models.

If visual-haptic integration is subject to modality-specific attentional cue weighting should be affected by adding the distractor task: Assuming that the distractor task withdraws relatively more attention from vision than touch, the variance of the visual estimate (probability density function) is increased to a larger extent than the variance of the haptic estimate (see Figures 1A and 1C). According to the maximum likelihood estimation (MLE) model, the relative visual and haptic weights are calculated based on the unimodal variances. Hence, if the attentional manipulation exerts an effect prior to integration, the relative cue weights are calculated based on the affected unimodal variances and should therefore be shifted

towards touch (*late integration model*; see Figure 1C). In the dual-task condition, the variance of the combined bimodal estimate should be increased compared to the variance in the single-task experiment due to the generally increased workload. If the estimates are still integrated, the bimodal variance should be lower than the variance of either unimodal estimate. Thus, according to the late integration model, the variance of the bimodal estimate in the dual-task condition can be predicted from the variances of the unimodal estimates using the MLE approach (cf. Equation 5).

If instead integration occurs at an early level of processing, prior to attentional effects, we would expect that the process of cue weighting is not affected by the withdrawal of attention from vision because the relative visual and haptic

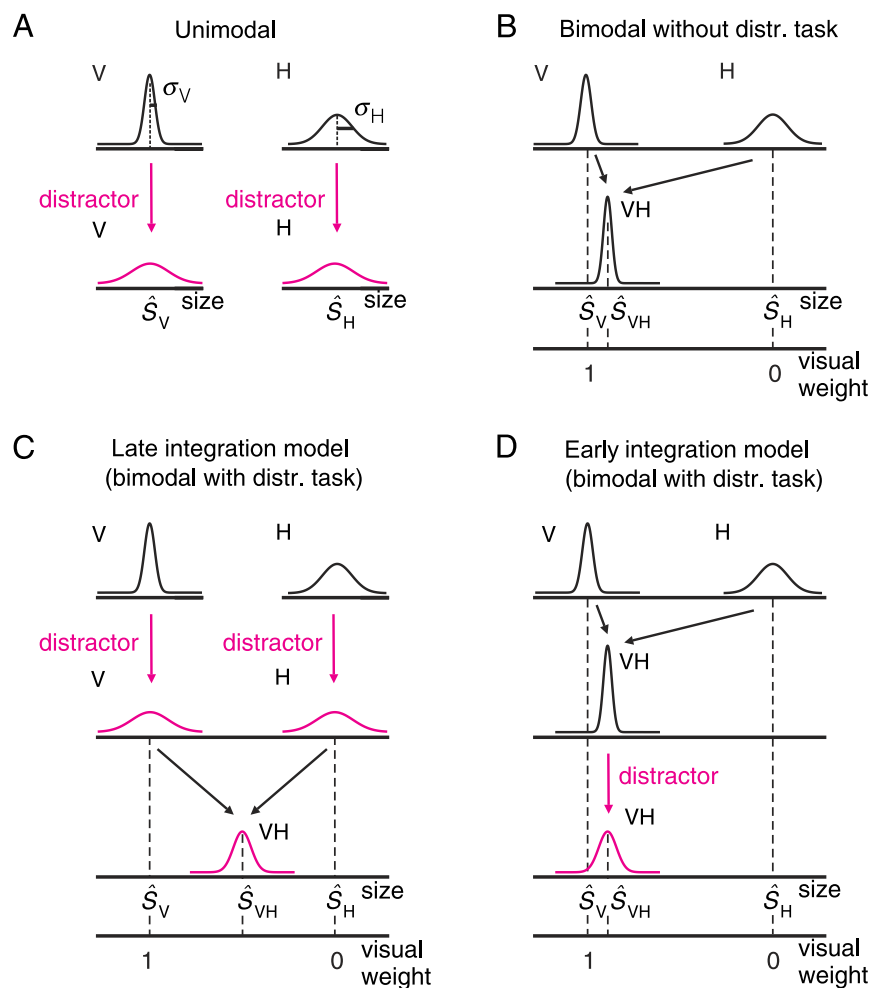


Figure 1. (A) The effect of the distractor task on the variances of the unimodal visual and haptic size estimates (σ_V^2 , σ_H^2) (in both the late and early integration models). The variance of the vision-based estimate is more increased than the variance of the haptics-based estimate. The curves are Gaussian probability density functions of the estimates. (B) Integration of visual and haptic size information (single-task condition). Higher relative cue weight is attributed to the more reliable cue. The variance of the bimodal estimate is lower than the variance of either unimodal estimate (MLE; Ernst & Banks, 2002). (C) Relative cue weights and bimodal variance are calculated from the variances that are affected by the distractor task. The distractor task exerts a larger effect (increase in variance) on the visual estimate; thus, less weight is attributed to vision (as compared to the situation without distractor). (D) Visual and haptic size information is integrated prior to the effect of the distractor task. Hence, the relative visual cue weight is not affected.

weights are calculated from the unimodal variances which are not yet affected by the attentional manipulation. Unlike the weights, the variance of the bimodal estimate is expected to be increased as a result of the interference with the distractor task at a later stage of processing (*early integration model*; see Figure 1D). It is, however, not possible to predict the amount by which the JND increases because it is not known how exactly the distractor interferes with the main task. Note that according to the early integration model, the distractor task can only have a differential effect on the two unimodal estimates if the distractor interacts with the estimates in a non-linear fashion (we return to this point in the [General discussion](#)). Note further that the early integration model is also an optimal model because the information available at the integration stage is used in an optimal way according to the MLE model.

Methods

Participants

Twelve female experienced psychophysical observers with normal or corrected-to-normal vision participated in the experiment for payment. All except two were right-handed. All participants were naive to the purpose of the experiment. The average age was 23 years (range 17–27). Participants gave their informed consent before taking part in the experiment, which was performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki.

Stimuli

Stimuli and experimental procedure were similar to those used by Ernst and Banks (2002). The visual and haptic stimuli of the main task were horizontal bars raised 30 mm above a plane with the upper surface being oriented perpendicular to the line of sight of the observer (see Figure 2A). Participants looked at the bar and/or grasped it with index finger and thumb of the dominant hand. The head was stabilized with a head-and-chin rest.

The visual stimuli of the main task were random dot stereograms representing a bar. The dots comprised a visual angle of 8×8 arcmin at a viewing distance of 50 cm and were displayed with a density of roughly 9 dots per deg^2 . The visual scene is displayed on a 21-in. RGB-computer monitor (SONY, GDM-F500R) with a resolution of 1280×1024 pixels (refresh rate 100 Hz). Participants viewed the mirror image of the visual scene via liquid-crystal shutter glasses (CrystalEyes™) providing binocular disparity (see Figure 2A).

The reliability of the visual stimulus was manipulated by adding a random component to the depth of the dots (parallel to the line of sight). There were two visual quality conditions. In the high visual quality condition, the random component was zero, and thus all dots that represent the upper surface of the bar or the background plane were located on one plane in the stereogram. In the low visual quality condition, the dots were randomly displaced in depth following a uniform distribution with a range of ± 1.5 cm (see Figure 2B, right panel).

The stimuli of the secondary task were visually presented uppercase letters (height of letters: 7 mm which corresponds to a visual angle of 48 arcmin; width:

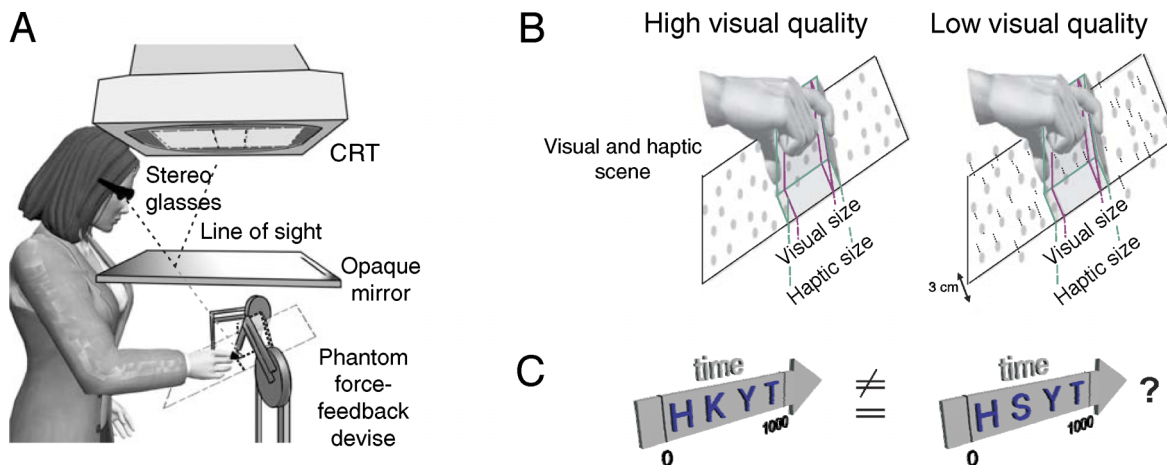


Figure 2. Setup and stimuli. (A) The visual scene is displayed on a cathode ray tube (CRT). Participants binocularly view the mirror image of the visual scene via shutter glasses that are used to present disparity. A head and chin rest limited head movements. PHANTOM force-feedback devices are used to present the haptic stimuli (adapted from Ernst & Banks, 2002). (B) The visual and haptic stimuli are horizontal bars raised 30 mm above a plane. Visual and haptic scene could be controlled independently such that stimuli could be generated that differ in visually and haptically specified height. The reliability of the visual scene was manipulated by displacing the dots that represent the bar randomly following a uniform distribution with a 3-cm depth step (right panel). (C) The stimuli of the distractor task were visually displayed letters. Letters were presented successively, 250 ms each letter.

approximately 10–32 arcmin). The letters were displayed with the same apparatus, on top of the bar, at the position of the fixation dot (center of the upper surface of the bar). In both the high and low visual quality conditions, the letters were presented 1 mm above the dots representing the bar to avoid that the letters are covered by “noise” (the dots). The haptic stimuli were presented with two PHANToM™ force-feedback devices, one each for the index finger and thumb. Using these devices, one can apply forces to the participants’ fingers to simulate the feel of touching 3-dimensional objects. Observers could not see their hands. The tips of index finger and thumb were visually represented by small spherical markers.

Visual and haptic scenes were spatially aligned. Participants have the convincing impression that they are visually and haptically exploring the same 3-dimensional scene. Both scenes could be controlled independently which allowed us to introduce size conflicts between the visual and haptic stimuli. The bar appeared at a randomly chosen horizontal position in center of the workspace. Thus projected area above or below the bar was not a useful cue to object size.

Procedure

In the main task, observers were asked to estimate the size (height) of a raised bar, either visually (V) or haptically (H) alone or by using information from both sensory modalities simultaneously (VH). The bar’s width spanned the entire workspace; the bar’s depth was always 3 cm. The only dimension that was varied was the height of the bar. Throughout this manuscript, we use size and height of the bar synonymously. On each trial, observers were presented with a standard stimulus whose height was constant at 55.0 mm and a comparison stimulus whose height was 41.0, 47.0, 49.0, 51.0, 53.0, 57.0, 59.0, 61.0, 63.0, or 69.0 mm. Standard and comparison stimuli were displayed in two subsequent intervals in counterbalanced order. Observers made a 2-interval forced-choice (2-IFC) response indicating which of the two bars was larger.

In the bimodal conditions (VH), visual and haptic height of the standard stimulus could be in conflict ($S_V - S_H = \delta$). The average was always 55.0 mm. We used conflicts of $\delta = \pm 6$ mm and ± 3 mm in the high visual quality condition and $\delta = \pm 6$ mm and 0 mm in the low visual quality condition. Conflicts ($S_V > S_H$, $S_V < S_H$) were counterbalanced to prevent adaptation. Five participants took part in the high visual quality condition. A second group of seven observers participated in the low visual quality condition. In both conditions, we measured haptic-alone (H), visual-alone (V), and visual-haptic (VH) size discrimination performance both with and without distractor task (V + D, H + D, VH + D, V, H, VH).

The distractor task consisted of comparing two series of letters using a same/different paradigm. Random series of four different uppercase letters (A–Z) were presented centered on top of the bar at the same location as the

fixation dot (in the single-task conditions) to prevent participants from directing their eyes away from the main task stimulus when they had to perform the distractor task concurrently. The fixation dot was not shown in the dual-task conditions but instead replaced by the letters. Letters rapidly succeeded one another (rapid serial visual presentation). Each letter was displayed for 250 ms. One series of letters was presented in each interval (synchronized with the presentation of the main task stimuli). Fifty percent of the trials in the distractor task were “different” trials. In those trials, one of the four letters (in one interval) was randomly replaced by another letter. Letters were chosen such that none of the letters occurred twice within one series.

Before each trial, the observers saw a start button. They were informed that they can initiate the presentation of the next trial by pressing the button. A fixation point indicated the position of the bar. In the haptics-alone and visual-haptic conditions, this fixation point helped participants to grasp the invisible bar. The bar was invisible before the beginning of the trial to prevent observers from making an initial visual judgment. They were instructed to fixate on this point and grasp the bar with index finger and thumb. The bar was presented as soon as both fingers made contact with the bar. Furthermore, the spherical markers representing the fingertips disappeared when the fingers were in contact with the bar. The stimulus was presented for 1000 ms. In the vision-alone condition, the procedure was the same except that participants did not grasp the bar. The stimulus appeared when the start button was pressed. The stimulus was extinguished after 1000 ms. Observers again pressed the start button to initiate the presentation of the stimulus in the second interval. Thereafter, response buttons appeared. When two judgments were required (in dual-task conditions), four response buttons were available (larger/smaller for the primary task and same/different for the distractor task). Participants were free to decide which response to provide first. The next trial was released as soon as the response was provided. No feedback was given.

Before the experiment, participants were administered ten practice trials in each single-task condition (V, H, VH) and forty practice trials in each dual-task condition (V + D, H + D, VH + D). Participants performed 400 trials per condition. (V, H, VH at four ($\delta = \pm 6$ and ± 3 mm) or three ($\delta = \pm 6$ and 0 mm) different conflict conditions, with and without distractor). Thus, observers performed a total of either $6 \cdot 2 \cdot 400 = 4800$ trials (in the high visual quality condition) or $5 \cdot 2 \cdot 400 = 4000$ trials (in the low visual quality condition). Conditions were blocked (100 trials per block) and presented in counterbalanced order. Within a block, each of the 10 comparison stimuli occurred 10 times (5 times in the first and 5 times in the second interval, presented randomly). In the bimodal conditions (VH, VH + D), all four or three conflicts occurred within one block. Combining 10 times 10 comparison stimuli with either four or three different types of standard stimuli

yielded blocks of 400 or 300 trials. These 400-trial or 300-trial blocks were randomized and subdivided in blocks of 100 trials.

Data analysis

Figure 3 illustrates how the point of subjective equality (PSE) and the discrimination threshold (JND) were derived from the data. The proportion of trials in which the comparison stimulus is perceived as being taller than the standard is plotted versus the size of the comparison stimulus. To obtain psychometric functions, the data were fitted with cumulative Gaussians free to vary in position (PSE) and slope (JND) using the software package *psignifit* (see <http://bootstrap-software.org/psignifit/>; Wichmann & Hill, 2001). The point of subjective equality (PSE) is the point at which the stimulus is judged as being taller than the standard 50% of the time, i.e., it corresponds to the size that is perceived as being equal to the standard stimulus. The discrimination threshold (just-noticeable difference, JND) is defined as the standard deviation of the underlying Gaussian (see Figure 3A). The JND corresponds to the size of a comparison stimulus that can be reliably discriminated from the standard stimulus (discrimination threshold). PSE and JND were determined separately for each of the conditions and for each participant.

Experimental testing for optimal cue integration

The optimal integration strategy is the maximum likelihood estimate. To test whether human behavior corresponds to such an ideal observer, we applied an experimental procedure that has become common practice in the recent past (e.g., Alais & Burr, 2004; Ernst & Banks, 2002; Helbig & Ernst, 2007b; Knill & Saunders, 2003). We compare the observers' performance measured in the experiment to the predictions of a Maximum Likelihood Estimator:

In a situation in which there are two (conditionally independent) cues to size (visual and haptic), the statistically optimal strategy for cue integration is a weighted average of the unimodal estimates (\hat{S}_V , \hat{S}_H):

$$\hat{S}_{VH} = w_V \hat{S}_V + w_H \hat{S}_H, \quad (1)$$

where the relative visual and haptic weights (w_V , w_H) are inversely proportional to the variances of the individual estimates (σ_V , σ_H)

$$w_V = \frac{1}{\frac{1}{\sigma_V^2} + \frac{1}{\sigma_H^2}} = \frac{\sigma_H^2}{\sigma_H^2 + \sigma_V^2} \quad (2)$$

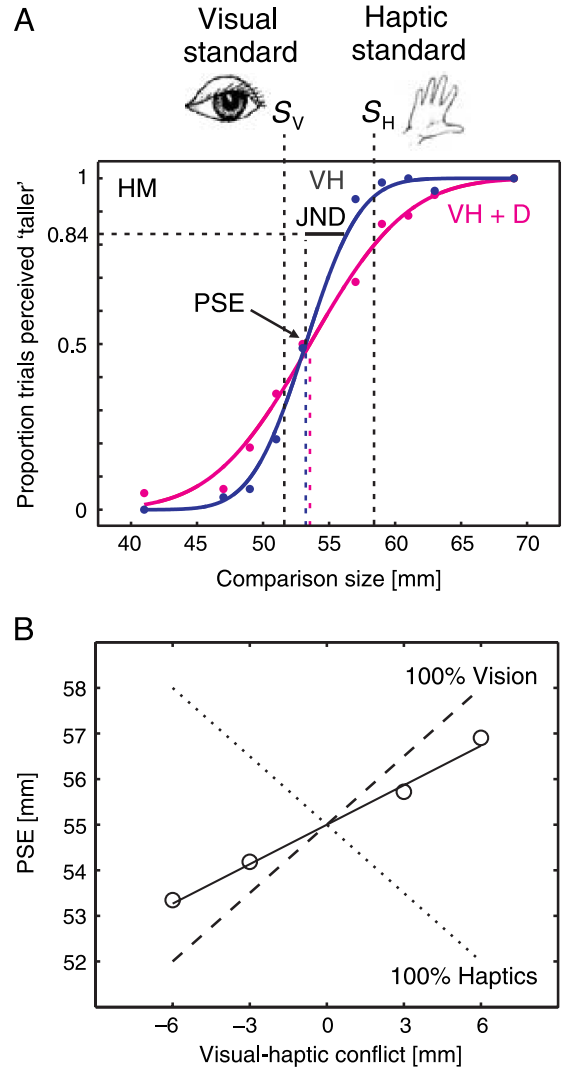


Figure 3. (A) Psychometric functions for size estimation. The ordinate represents the proportion of trials in which the comparison stimulus was perceived as 'taller' than the standard stimulus as a function of the size of the comparison stimulus. Data were fitted with cumulative Gaussians to derive the PSE (50% level) and the JND (84% level). Data were exemplarily from one participant (HM) obtained for bimodal size discrimination in the single-task (blue line) and dual-task (red line) experiments. Visually and haptically specified sizes of the standard stimulus were in conflict: $S_V = 52$ mm, $S_H = 58$ mm (visual-haptic conflict $S_V - S_H = -6$ mm). The JND increases when the distractor task is conducted concurrently (red line), whereas the PSE and therefore the relative visual and haptic weights are not affected. (B) PSE for bimodal size discrimination as a function of the visual-haptic conflict. The continuous line is a linear fit to the data. The slope of this line is a measure of the relative visual and haptic cue weights. PSEs lie along a line of slope 0.5 if vision dominates completely (visual weight $w_V = 1.0$), and along a line of slope -0.5 if haptics is the dominant modality ($w_H = 1.0$).

and

$$w_H = \frac{\sigma_V^2}{\sigma_H^2 + \sigma_V^2} \quad (3)$$

and sum to one:

$$w_V + w_H = 1 \quad (4)$$

(see Clark & Yuille, 1990; Yuille & Bülthoff, 1996).

The variance of the statistically optimal bimodal estimate is

$$\sigma_{VH}^2 = \frac{\sigma_V^2 \cdot \sigma_H^2}{\sigma_V^2 + \sigma_H^2}, \quad (5)$$

where σ_V , σ_H , and σ_{VH} are the standard deviations of the probability density functions of the visual, haptic, and combined size estimates. The variance of the combined estimate is always less than the variance of either estimate

$$\sigma_{VH}^2 \leq \min(\sigma_V^2, \sigma_H^2). \quad (6)$$

Estimates of σ_V^2 and σ_H^2 are obtained by fitting psychometric functions to the data of the unimodal visual and haptic estimates. The just-noticeable difference (JND) is derived from the psychometric functions and corresponds to

$$\text{JND} = \sqrt{2} \cdot \sigma \quad (7)$$

That is, from the unimodal visual and haptic JNDs (JND_V , JND_H), we can calculate (1) the optimal JNDs (Equations 5 and 7) for bimodal, visual-haptic size estimation and (2) the predicted weights for optimal integration behavior (Equations 2, 3, and 7).

Empirically, the visual and haptic weights can be obtained by introducing small conflicts between visually and haptically specified size of the standard stimulus (see Figure 3B). A shift of the PSE towards the actual visual or haptic input is a measure of the relative visual and haptic weight:

$$w_V = \frac{|\text{PSE} - S_H|}{|S_V - S_H|}. \quad (8)$$

We measured the PSEs with different size conflicts (± 6 and ± 3 mm in the high visual quality condition and ± 6 and 0 mm in the low visual quality condition) and plotted the PSEs as a function of the visual-haptic conflict (see Figure 3B). In order to get a more reliable estimate of the relative visual and haptic weights, we fitted a line to these data. The slope of this line is a measure of the relative visual and haptic cue weight. If vision were to dominate completely (visual weight $w_V = 1.0$), the points of subjective equality (PSE_{VH}) are consistent with the visually specified standard stimulus.

Thus, they would lie along a line with a slope of 0.5. If the haptic modality dominates ($w_H = 1.0$), the bimodal PSEs lie along a line of slope -0.5 .

Results and discussion

The experiment aimed at testing whether withdrawing relatively more attention from one sensory modality (vision) than from the other modality (haptics) has an effect on the multisensory integration process. In particular, we asked whether there is an effect of modality-specific attention on the weighting of information from different sensory channels.

Differential influence of the distractor task on unimodal estimates

First, we tested whether the applied distractor task does indeed withdraw selectively more attention from the visual main task. Withdrawing attention reduces behavioral performance. Specifically, it increases the variability of the observers' responses (Prinzmetal et al., 1997, 1998). Therefore, we first tested whether the distractor task has a stronger influence on the vision-based estimates (V) than on haptics-based estimates (H) (relatively larger increase in variability of observers' responses). Such a differential effect would show that selectively more attention was drawn away from the visual sensory channel.

To this end, we compared size discrimination performance (JND) of the unimodal main tasks (V, H) to the performance obtained when the distractor task was carried out simultaneously (V + D, H + D). Note that in the high visual quality condition, the performance on the visual task was clearly better than the performance on the haptic main task. Therefore, comparing the loss in performance of the visual and haptic tasks could be problematic because increasing the JND by the same amount has a relatively stronger effect on the more reliable estimate. We therefore added the low visual quality condition in which performance in the visual and haptic single tasks is about equal. As expected, we found that the performance in the visual as well as the haptic task drops when a demanding distractor task is carried out concurrently. However, the increase in variability of the responses in the visual size estimation task was relatively stronger than in the haptic task (see Figure 4). While the JND of the haptic size estimate increased by only 24.2%, the JND of the visual size estimate increased by 85.2% (high visual quality) and 51.7% (low visual quality), respectively. For each visual quality condition, JND data were averaged across $n = 5$ and 7 observers, respectively. The haptic JND data were averaged across the $n = 12$ observers because the unimodal haptic task is obviously not affected by the

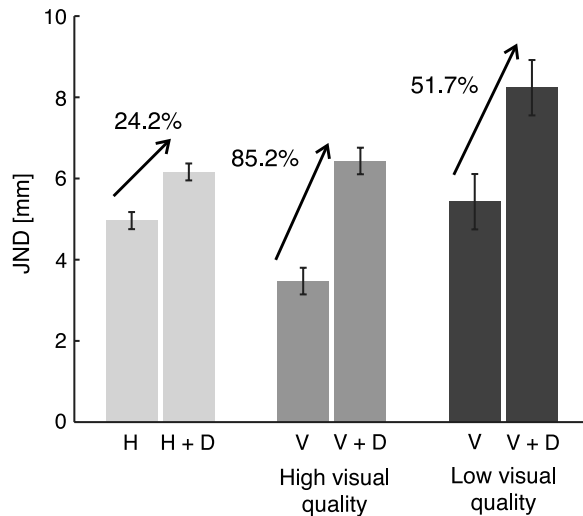


Figure 4. Just noticeable differences (JNDs) for unimodal size discrimination without and with distractor task (+D). Light grey bars represented the haptic condition, dark grey represents the high visual quality condition, and black represents the low visual quality condition. The error bars depict the standard error of the mean across observers.

visual quality. In order to verify that there is a differential effect of the distractor task on the visual and the haptic main task, a repeated measures ANOVA was performed on discrimination performance (JND) with the within-subjects factors modality (visual, haptic) and load (single task, dual task) and the between-subjects factor visual quality (high, low). The analysis revealed a significant main effect of load ($F(1,10) = 29.998$, $MSE = 1.717$, $p < .001$).

That is, the applied distractor task has significantly reduced the performance of the main task. Importantly, the interaction between modality and load was significant ($p < .025$). Hence, the distractor task had a significantly stronger effect on the vision-based estimates (see Figure 4). This differential effect on the variability of responses indicates that the attentional manipulation was effective.

This finding is consistent with previous research that demonstrated that dual-task interference depends on the similarity of input modalities (Allport, Antonis, & Reynolds, 1972; Wickens, 1980, 1984). Note that in order to perform both visual size estimation and distractor task, observers had to divide attention between visually presented letters, and the visually or haptically presented bar during the whole 1-s interval of stimulus presentation as the four visual letters of the distractor task were presented briefly and successively for 250 ms each. Besides the visual perceptual load, the distractor task is likely to affect processes common to both the visual and the haptic size discrimination task, such as for example memory components, and thus impairs performance not only in the visual but also in the haptic main task to some extent.

Performance on the distractor task

We analyzed performance on the distractor task to ensure that the selectively stronger interference of the distractor task with the visual main task (see previous section) was not simply due to a tradeoff between main and distractor task performance. We conducted a repeated measures ANOVA on distractor performance (percent correct responses) with within-subjects factor Condition (H + D, V + D, VH + D) and between-subjects factor visual quality (high, low). The analysis revealed a significant main effect of condition ($F(2,20) = 10.668$, $MSE = 4.203$, $p = .001$). The performance on the distractor task did not differ significantly between the high and the low visual quality conditions ($F(1,10) = 0.945$, $p > .35$). The interaction of condition and visual quality ($p > .79$) was not significant. Therefore, we collapsed the data across visual quality conditions and conducted post hoc *t*-tests. The α level was adjusted using a Bonferroni correction (so that the α level was set to .017 instead of .05). The results are shown in Figure 5. A two-tailed paired-sample *t*-test revealed significant differences of distractor performance between the conditions H + D and V + D (H + D: mean p.c. = 89.3%; V + D: mean p.c. = 85.6%; $t(11) = 4.199$, $p < .0015$) as well as V + D and VH + D (V + D: mean p.c. = 85.6%; VH + D: mean p.c. = 88.6%; $t(11) = -4.062$, $p < .0019$). H + D and VH + D did not differ significantly ($t(11) = 0.836$, $p > .42$).

The results show that the distractor performance is lowest when the distractor task is combined with the visual main task. That is, performance on the distractor task as well as on the size discrimination task were lowest in the visual-alone condition. Therefore, we can conclude that the differential effect of the distractor on the main tasks (V, H) is not merely due to a tradeoff between distractor performance and main task performance.

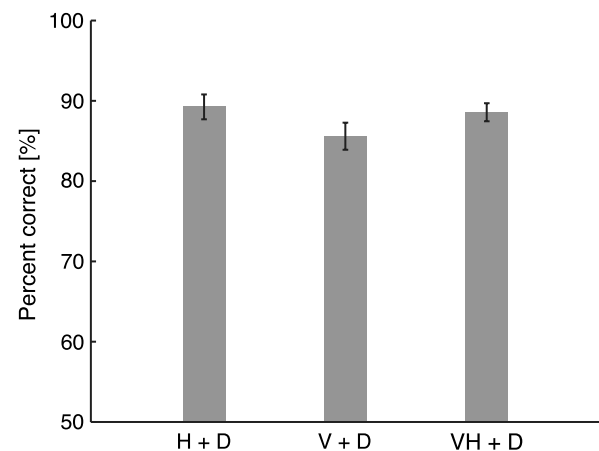


Figure 5. Performance on the distractor task D (percent correct responses) when combined with different main tasks (haptic-alone H + D, visual-alone V + D, visual-haptic VH + D). The error bar depicts the standard error of the mean across observers.

Size discrimination performance with different size conflicts

Gepshtein, Burge, Ernst, and Banks (2005) observed that JNDs increase at larger amounts of intersensory conflict, indicating that observers do not integrate multi-sensory information in a statistically optimal manner when cues are not consistent, but instead integration seems to decline with increasing amounts of conflict. In our study, repeated measures ANOVAs revealed that bimodal JNDs did not vary across conflicts (high visual quality: visual-haptic conflicts $-6, -3, +3, +6$; low visual quality: visual-haptic conflicts $-6, 0, +6$). This holds for single-task conditions (high quality: $F(3,12) = 0.486$, $MSE = 0.035$, $p > .697$; low quality: $F(2,12) = 0.142$, $MSE = 0.383$, $p > .868$) as well as for dual-task conditions (high quality: $F(3,12) = 0.825$, $MSE = 0.127$, $p > .504$; low quality: $F(2,12) = 1.285$, $MSE = 0.647$, $p > .311$). Hence, the conflicts introduced between the visually and haptically specified sizes were still small enough for integration to occur. To increase the statistical power, for further analysis, we collapsed the data across conflicts.

Bimodal size discrimination performance without distractor task (VH)

To verify that human observers do indeed integrate visual and haptic size information in a statistically optimal way (Ernst & Banks, 2002) in the single-task condition, we examined whether the bimodal size discrimination thresholds (JNDs) correspond to the predictions of the MLE model (see [Methods](#)). The model predicts a reduction of the variance for the bimodal as opposed to the unimodal visual and haptic size estimate ([Equations 5 and 6](#)). Such a reduction of variance is the signature of integration.

[Figure 6A](#) shows observed versus predicted bimodal JNDs (single-task condition) for all 12 individual observers (left panel). Predictions were derived from the single cue JNDs without distractor task. The data agree with the models' predictions (dashed line). JNDs obtained in the low visual quality condition (black squares) are in general higher than JNDs in the high visual quality condition (grey squares). A repeated measures ANOVA with within-subjects factor prediction (predicted, observed) and between-subjects factor visual quality (high, low) revealed that this difference between the high and low visual quality condition is significant ($F(1,10) = 5.049$, $MSE = 0.582$, $p < .048$). Importantly, the observed JNDs did not differ significantly from the prediction ($F(1,10) = 3.864$, $MSE = 0.289$, $p > .077$; the interaction of prediction and visual quality was not significant, $p > .62$). We further tested whether the variance of the bimodal estimate is indeed lower than the variance of either unimodal size estimate ([Equation 6](#)). To this end, we conducted two one-tailed paired-sample t -tests comparing the JND obtained for bimodal estimates with the unimodal visual and haptic

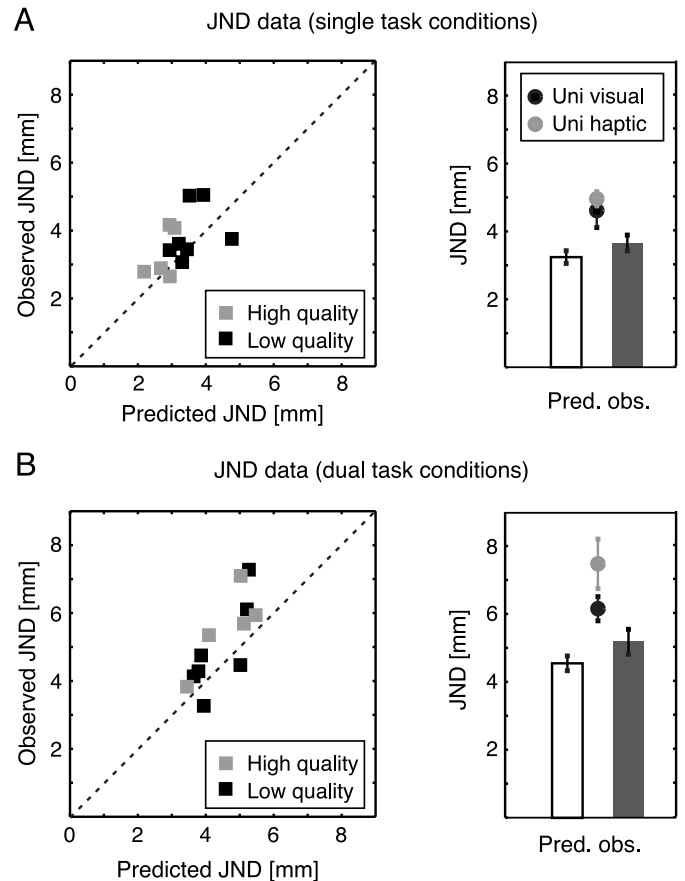


Figure 6. Bimodal size discrimination performance. Left panels: predicted versus observed just-noticeable differences (JNDs). The dashed diagonal line represents equality. Data are for 12 individual observers. Black squares represent the low visual quality condition. Grey squares represent the high visual quality condition. Right panels: Data averaged across the 12 participants. The white and grey bars show the predicted and observed bimodal JNDs. The black and grey circles show the unimodal visual and haptic JNDs. The error bars represent the standard error of the mean across participants. (A) Single-task experiment. (B) Dual-task experiment.

JND, respectively. These tests revealed that the bimodal JND is significantly lower than the JND of the unimodal visual ($t(11) = 2.002$, $p < .035$) and haptic ($t(11) = 4.584$, $p < .001$) estimate (see [Figure 6A](#), right panel). Together, these findings confirm that in the single-task conditions visual and haptic size information is integrated according to the MLE scheme.

Bimodal size discrimination performance with distractor task (VH + D)

[Figure 6B](#) shows JND data of the dual-task experiment. The left panel shows observed versus predicted bimodal JNDs for individual participants. The white and grey bars in [Figure 6B](#) summarize these data (right panel). The

predictions are derived from the MLE model (Equation 5) and were calculated from the unimodal visual and haptic JNDs obtained when observers perform a concurrent distractor task (V + D, H + D). These predictions should be fulfilled in case attentional effects (withdrawing attention from vision) occur prior to the integration process because in this case the integration process and thus the computation of the optimal performance (JND) are based on the unimodal estimates whose variance is increased by the distractor task (V + D, H + D) (late integration model). As mentioned already in the [Introduction](#), it is not possible to predict the bimodal JND (JND_{VH+D}) when the interference of the distractor task occurs after the integration process (early integration model) because it is unknown how exactly the distractor task would affect the JND of the bimodal estimate.

A repeated measures ANOVA was conducted on the JNDs with within-subjects factor prediction (predicted, observed) and between-subjects factor visual quality (high, low). Bimodal JNDs obtained in the high and low visual quality conditions did not differ significantly in this experiment ($F(1,10) = 0.653$, $MSE = 1.898$, $p > .437$). Interestingly, the analysis revealed that the observed bimodal JNDs were significantly higher than values predicted by an optimal integrator from the V + D and H + D unimodal estimates ($F(1,10) = 7.366$, $MSE = 0.355$, $p < .022$). We further tested whether the variance of the bimodal size estimate is reduced also when observers perform a concurrent distractor task. One-tailed, paired-sample *t*-tests revealed that the bimodal JND_{VH+D} is significantly lower than the unimodal visual JND_{V+D} ($t(11) = 4,020$, $p < .001$) and the unimodal haptic JND_{H+D} ($t(11) = 2,601$, $p < .013$) observed in the dual-task condition (see [Figure 6B](#), right panel). In summary, these results indicate that visual and haptic shape information is still integrated when attention is selectively withdrawn from one sensory channel. That is, multisensory integration does not break down under such conditions of high attentional load (in contrast to [Alsus et al., 2005](#)).

With regard to the integration models two possible interpretations remain: Either integration takes place after effects of the distractor occur (late integration model) but becomes slightly sub-optimal with the introduction of the distractor task. Alternatively, it may be that the effects of the distractor task occurred at a later stage of processing, after integration, so that the bimodal JNDs are slightly increased relative to the predictions for optimal behavior (early integration model; [Figure 1D](#)). To distinguish these two hypotheses, we next analyzed the relative cue weights in single-task and dual-task conditions.

Cue weighting with and without distractor task

The relative visual and haptic weights are inversely proportional to their variances (Equations 2 and 3). That is, higher weight is attributed to the more reliable sensory

channel. Hence, the visual weight is expected to decrease when the visual reliability is reduced. Similarly, when the variance of one sensory modality (here the visual channel) is increased in relation to the other modality due to increased attentional load, the weight associated with this channel should be reduced if that attentional mechanisms operate prior to the process of multisensory integration (late integration model). If, however, visual-haptic integration is an early, pre-attentive process, cue weights should be derived from the variance of the unimodal estimate that are not yet affected by the increased attentional load (V, H not V + D, H + D) and should thus not be affected by such attentive processes.

First, we analyzed the results of the single-task experiment. [Figure 7A](#) shows observed versus predicted relative visual weights without distractor task. Data for all 12 observers in the high and low visual quality conditions are depicted in the left panel of [Figure 7A](#). The right panel summarizes these data (averaged across visual quality conditions). To test whether the predictions of the MLE model hold for visual-haptic size estimation in the single-task conditions, we conducted a repeated measures ANOVA on relative visual cue weights with within-participants factor prediction (predicted, observed) and between participants factor visual quality (high, low). The analysis revealed a significant main effect of visual quality ($F(1,10) = 23.965$, $MSE = 0.024$, $p < .001$). The relative visual weight is lower in the low visual quality conditions (black squares). Most importantly, the predicted visual weights do not differ from the observed visual weights (predicted: 0.55; observed: 0.54; $F(1,10) = 0.036$, $MSE = 0.015$, $p > .851$). There was a significant interaction between factors prediction and visual quality. Therefore, we conducted post hoc *t*-tests comparing predicted and observed visual weights for the high and low visual quality conditions separately. In neither of the visual quality conditions did the observed weight deviate significantly from the predicted value (high quality: $t(11) = -1,829$, $p > .14$; low quality: $t(11) = 1,806$, $p > .20$). To conclude, we successfully replicated the results of [Ernst and Banks \(2002\)](#): As predicted by the MLE rule, visual weights decrease when the visual quality is reduced. Overall, we found that the observed visual weights do not differ from the predicted weights. Together with the reduction in variance for bimodal estimates (see previous section), these results show that observers integrate visual and haptic size information in a statistically optimal manner when there is no attention demanding secondary task.

Secondly, we analyzed the weight data of the dual-task experiment ([Figure 7B](#)). We compared the observed visual weight (in the dual-task condition) to the predictions of the optimal integration model (MLE). To distinguish between the late and early integration models, we first compared the observed cue weights to the predictions derived from the unimodal visual and haptic JNDs with distractor task (JND_{V+D} and JND_{H+D} ; [Figure 7B](#), upper panel). This set of predictions should be met if attentional

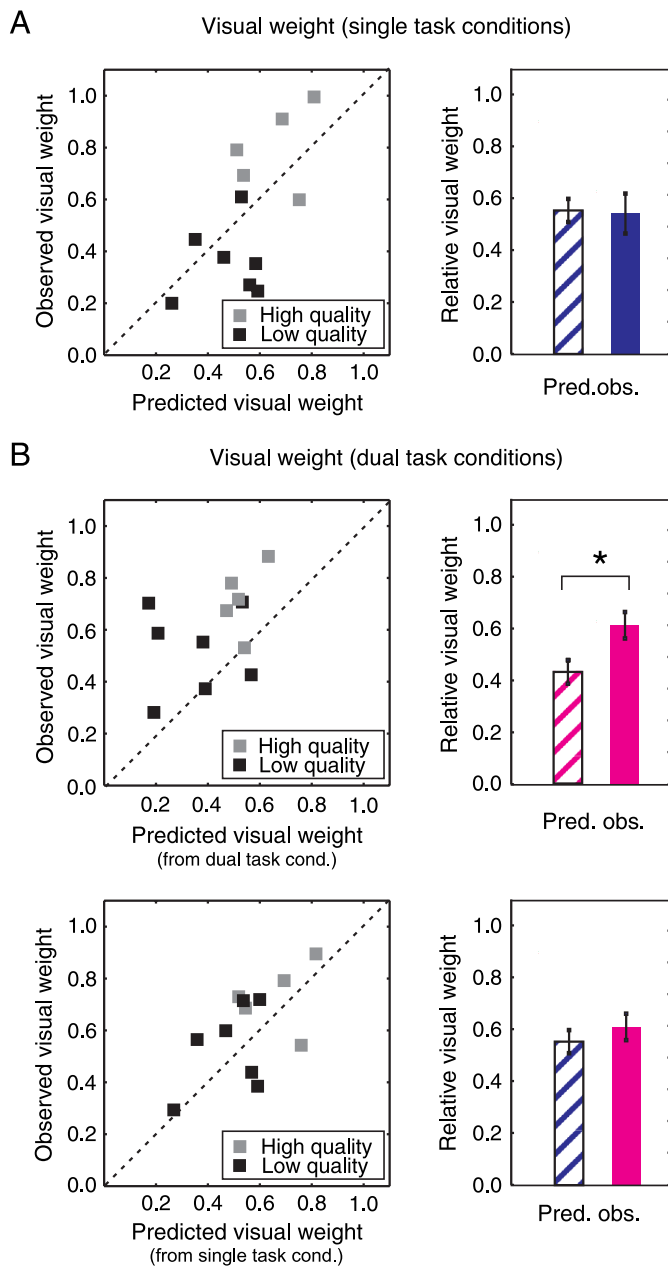


Figure 7. Observed versus predicted relative visual weight for individual subjects (left panels, dashed diagonal lines represent equality) and averaged across observers (right panels). Blue bars represent visual weights measured in the single-task condition, and red bars represent the dual-task condition. Blue and red striped bars represent the predicted relative visual weights calculated from the unimodal JNDs in the single-task (blue) and dual-task (red) conditions, respectively. (A) Without distractor. (B) With distractor. Upper panels: predictions calculated from the dual-task condition (JND_{V+D} , JND_{H+D}). Lower panels: predictions calculated from the single-task condition (JND_V , JND_H) compared to empirical visual weight in the dual-task experiment.

processes (withdrawing attention from vision) occur prior to multisensory integration (late integration model) because in this case, computation of the relative cue weights is based on the unimodal estimates that are affected by the distractor ($V + D$, $H + D$).

We then compared the observed weights to the predictions derived from the unimodal visual and haptic JNDs without distractor task (JND_V and JND_H ; Figure 7B, lower panel). This set of predictions should be met if multisensory integration occurs prior to the attentional effects (early integration model) because in this case, the unimodal estimates from which the cue weights are derived are not affected by the distractor (V , H). The left half of Figure 7B shows the predicted versus observed visual weights for the individual observers; the right half shows these data averaged across subjects and visual quality conditions.

When comparing the weight data against the predictions derived from the unimodal JNDs with distractor task (predictions from JND_{V+D} , JND_{H+D} ; Figure 7B, upper panel), it is clear that these predictions fall short. This can be seen both when looking at the individual observers data (left panel) as well as the average data (right panel). The observed visual weight was higher than the predicted value. A repeated measures ANOVA with within-participants factor prediction (predicted, observed) and between participants factor visual quality (high quality, low quality) confirmed this observation. The difference between empirical data and predictions (derived from the unimodal data in the dual-task conditions) was statistically significant (predicted: 0.43; observed: 0.61; $F(1,10) = 10.437$, $MSE = 0.018$, $p < .01$). If multisensory integration is affected by withdrawing attention from one sensory channel, we would have expected the visual weight to be reduced in this experiment (because the variance of the visual channel increased more than that of the haptic channel).

We next compared the visual weights obtained in the dual-task conditions to the predictions derived from unimodal variances without distractor task (JND_V and JND_H). The data are shown in the lower panel of Figure 7B. The visual weights do not differ significantly from these predictions ($F(1,10) = 1.519$, $MSE = 0.013$, $p > .245$). That is, they are not affected by the secondary task which is withdrawing attention from vision. This suggests that multisensory integration precedes attentive processes, i.e., multisensory estimates are integrated into a unified percept before the reliability of to the unimodal estimate, is changed due to attentional effects (early integration model).

Non-optimal models that may account for the data are discussed in the next section (cf. General discussion).

General discussion

The goal of this study was to examine whether selectively attending to one sensory modality modulates

the weighting of cues from different sensory systems. In accordance with the early integration model, we observed that the visual weight is not decreased when attention is drawn away from the visual channel (by adding a ‘visual’ distractor task), even though this manipulation increased the variability of the visual channel selectively more than that of the haptic channel. Moreover, the JND of the bimodal estimate increased when the demanding distractor task was performed but was still significantly lower than the unimodal JNDs, indicating that observers integrate multisensory information even under dual-task conditions. Together, these results demonstrate that modality-specific attention does not interfere with the process of multi-sensory integration.

We observed that the secondary task exerts a stronger effect on the visual than the haptic main task. At the first glance, this might seem puzzling. How is it possible that the secondary task exerts different effects on vision and haptics even though visual and haptic information seems to be integrated prior to attentive processes (such as withdrawing attention from one sense) and thus the same amount of noise should be added to the two channels? Possibly, the distractor task affects the variance of the estimates in a non-linear manner. In the following, we discuss one possible scenario how the distractor could affect the unisensory estimates differently even though the unisensory signals are fused prior to the occurrence of attentional effects: In this context, it is important to note that by means of psychometric functions we can only measure the signal-to-noise ratio of the underlying estimate:

$$\frac{1}{\text{JND}} \propto \frac{\text{signal}}{\text{noise}}. \quad (9)$$

If, for example, the JND would be inversely proportional to the signal-to-noise ratio, then the increase in JND would depend on the signal strength. That is, two estimates that appear equally reliable (same JND) can have different signal strengths. Increasing the noise of these two estimates by the same amount (or decreasing the signal strengths) has consequently a stronger effect on one than on the other estimate even though the same amount of noise is added (or the signal strength is decreased by the same amount).

We here tested human performance against two optimal models of cue integration (the early and the late integration model). The empirical data were consistent with the early integration model suggesting that visual and haptic information is integrated prior to the occurrence of attentional effects (withdrawing selectively more attention from one sensory channel by means of a distractor task). In the following, we discuss alternative non-optimal models that may also account for the data. For example, it might be that the cue weights are learned under conditions of low attentional load and that these weights are applied also under conditions of high attentional load.

Such a model would also account for the data (cue weights) we observed. However, this alternative account seems rather unlikely. If the weights were learned, how would the system decide which set of weights to apply in which situation? Note that in our experiment the order of conditions was randomized. How would the system decide in which situation (condition) to learn the cue weights in the first place? Furthermore, why should a system that is optimal in condition decide to use a sub-optimal strategy in another condition? In addition, there is a growing body of evidence speaking against learned weights and in favor of an online estimation of the weights from the variances of the current stimuli (Alais & Burr, 2004; Ernst & Banks, 2002; Hillis, Ernst, Banks, & Landy, 2002, optimal integration papers). In sum, it seems rather unlikely that the system first learns the variance and computes the cue weights on non-distractor trials and then further uses these cue weights under dual-task conditions.

The results of the present study are broadly consistent with previous work on multimodal perception that provided evidence for the automatic, pre-attentive nature of multisensory integration, mostly in the auditory–visual domain (e.g., Bertelson et al., 2000; Driver, 1996; Vroomen et al., 2001a). Evidence in support of the view that multisensory integration is an early, pre-attentive process comes also from neurophysiological (ERP, fMRI) studies. Recent event-related potential (ERP) studies revealed that auditory mismatch negativity (MMN) can be evoked (Stekelenburg, Vroomen, & de Gelder, 2004) or eliminated (Colin, Radeau, Soquet, Dachy, & Deltenre, 2002) by illusory sound shifts induced by ventriloquism. MMN is typically evoked by an occasional deviant in a homogenous sequence of auditory stimuli and is thought to reflect pre-attentive processes (for a review, see Näätänen, 1992). Therefore, it can be inferred that ventriloquism, i.e., audiovisual integration, occurs at an early, pre-attentive processing stage. Similarly, MMN can be elicited by illusory McGurk percepts (Colin, Radeau, Soquet, Demolin, et al., 2002), indicating that audiovisual integration mechanisms in speech occur early during the perceptual processes. Furthermore, in accordance with the notion that multisensory integration is an early process, a large number of functional magnetic resonance imaging studies (fMRI) observed interactions of inputs from different sensory modalities in primary sensory areas typically thought of as being strictly modality specific (e.g., Calvert et al., 1997; Kayser, Petkov, Augath, & Logothetis, 2005; Macaluso, Frith, & Driver, 2000; Pekkola et al., 2005; van Atteveldt, Formisano, Goebel, & Blomert, 2004).

Along this line, a number of studies observed that even when observers were explicitly instructed to ignore one sensory modality, the bimodal percept was strongly influenced by the task-irrelevant sensory signal (e.g., Bresciani et al., 2005, 2006; De Gelder & Vroomen, 2000; Helbig & Ernst, 2007a; Massaro, 1987a, 1987b; Shams et al., 2000, 2002; Spence et al., 2004, Spence &

Walton, 2005), indicating that cross-modal integration occurs in an automatic fashion.

Paradoxically, some studies that were using exactly the same paradigm, i.e., studies that presented participants with conflicting bimodal information while asking them to report the percept from either one or the other modality (e.g., Bertelson & Radeau, 1981; Massaro, 1998; Warren et al., 1981), were taken as evidence by other researchers that attending to one sensory modality can influence cue integration. In these studies, it was found that the reported percept depended on which sensory modality participants were instructed to attend to (but still there was an influence of the task-irrelevant stimulus). In the following, we attempt to reconcile these apparently conflicting interpretations.

Recent research suggested a continuum of multisensory interactions ranging from complete fusion to independence between the cues (Bresciani et al., 2006; Ernst, 2005; Roach, Heron, & McGraw, 2006; Shams et al., 2005). Single-cue information from different sensory modalities is not necessarily lost completely when bimodal information is presented (Hillis et al., 2002). In line with this, Gepshtein et al. (2005) demonstrated that with increasing spatial conflicts between the multisensory stimuli, cues are no longer fused completely. With these considerations in mind, we come back to the studies that yielded different responses depending on which sensory modality was asked for. For example, in Massaro's (1998, p. 246) study, observers were presented with a talking head that says the word 'please' with different emotions (happy, angry, neutral) in the face and in the voice. When the emotional expression of face and voice were in conflict (e.g., happy face but angry voice), instructions to identify the emotion as happy or angry based on what observers hear as opposed to what they see, had a significant impact on the probability of happy judgments. This was taken as a hint that attending to one modality can alter the percept. However, if observers based their decision on both, what they saw and what they heard, equal proportions of happy and angry judgments were reported indicating that seen and heard information is not entirely fused. This is probably the reason why participants could independently report visual and auditory information. Similarly, in the studies conducted by Bertelson and Radeau (1981) and Warren et al. (1981), stronger biasing effects of the unattended modality were observed in conditions in which sensory information is fused. In conditions in which it was likely that the sensory estimates were not completely fused, the reported percept was closer to the actual input of the to-be-attended modality. To be more specific, in Bertelson and Radeau's study, the signal in the to-be-ignored modality exerted stronger biasing effects in fusion trials than in non-fusion trials (trials in which participants became aware of the intersensory conflict). In Warren et al.'s study, it was found that in a highly compelling stimulus situation, in which in addition participants were instructed that both sensory signals come from a common source, the reported percept was hardly influenced by the

attention condition. Whereas when observers were instructed that visual and auditory information may come from different sources, responses when asked to report vision as opposed to audition differed. To summarize, it seems that instructions to attend to one sensory modality while ignoring the other one has higher impact on the reported percept when multisensory information is not fused completely, so that single cue information is more readily accessible. Thus, we think that when studying effects of modality-specific attention it is indispensable to control whether cues are indeed fused.

A further concern with studies merely instructing observers to attend to one modality might be that this paradigm does not provide any means of monitoring whether indeed more attention is paid to one sensory modality versus the other. Possibly, sufficient attentional resources were available to process the task-irrelevant stimulus as well, which then interfered with the target-stimulus. Therefore, we quantitatively tested whether multisensory information is integrated, while modulating the degree of modality-selective attention within a dual-task paradigm that allows us to test whether indeed attention is withdrawn from one modality.

Alsuis and colleagues (2005) observed a reduction of McGurk percepts (i.e., fused percepts) under dual-task conditions. That is, in contrast to our study, they apparently observed that integration breaks down under conditions of high attentional load. From these findings, they concluded that multisensory integration depends on attentional resources. However, there is an alternative explanation of these results (illustrated in Figure 8).

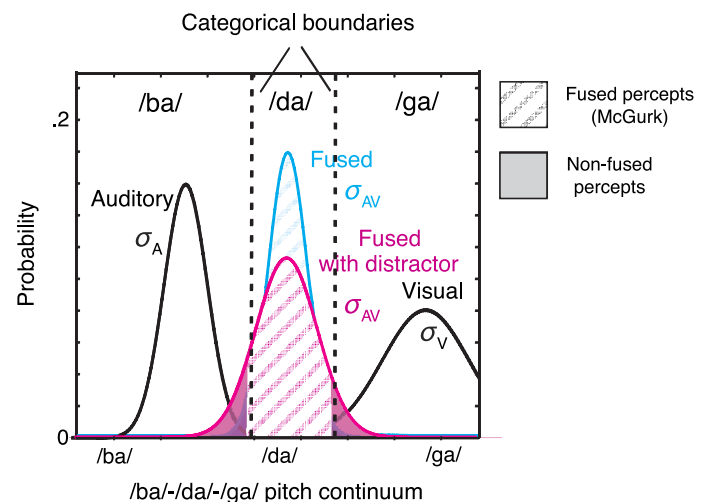


Figure 8. Illustrates a possible explanation for the reduced number of fusion responses observed under dual-task conditions (Alsuis et al., 2005). The variance of the probability density function associated with the fused percept (McGurk response, /da/) is increased when a distractor task is added (red versus blue). Therefore, a larger proportion falls beyond the categorical boundaries (filled area). Hence, the proportion of fused responses is reduced (striped area).

Consider, for example, a case in which an auditory syllable /ba/ and a visual syllable /ga/ are fused to produce /da/. If—consistent with our study—the distractor interferes with the main task stimuli at a post-integration stage and thus affects the fused percept (i.e., increases the variance of the probability density function (pdf) associated with the fused percept, here /da/), then the pdf of the fused percept is broadened to exceed the categorical boundaries, and consequently the probability of perceiving non-fused percepts (i.e., /ba/ or /ga/) increases. According to this alternative account, our results would be consistent with Alsius et al.'s study. An important difference between ours and their study, which can account for the apparently different results, is that they studied categorical perception (phoneme perception) whereas we studied continuous perception (size perception).

In conclusion, we here have quantitatively demonstrated that multisensory visual-haptic integration of size information is an automatic process which is unaffected by modality-selective attention.

Acknowledgments

This work was supported by the Max Planck Society, by the 5th and 6th Framework IST Program of the EU (IST-2001-38040 TOUCH-HapSys & IST-2006-027141 ImmerSence), and by the Deutsche Forschungsgemeinschaft DFG (Sonderforschungsbereich 550).

Commercial relationships: none.

Corresponding author: Hannah B. Helbig.

Email: helbig@tuebingen.mpg.de.

Address: Spemannstr. 41, D-72076 Tübingen, Germany.

References

- Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology*, *14*, 257–262. [PubMed] [Article]
- Allport, D. A., Antonis, B., & Reynolds, P. (1972). On the division of attention: A disproof of the single channel hypothesis. *Quarterly Journal of Experimental Psychology*, *24*, 225–235. [PubMed]
- Alsius, A., Navarra, J., Campbell, R., & Soto-Faraco, S. (2005). Audiovisual integration of speech falters under high attention demands. *Current Biology*, *15*, 839–843. [PubMed] [Article]
- Bertelson, P., & Radeau, M. (1981). Crossmodal bias and perceptual fusion with auditory–visual discordance. *Perception & Psychophysics*, *29*, 578–584. [PubMed]
- Bertelson, P., Vroomen, J., De Gelder, B., & Driver, J. (2000). The ventriloquist effect does not depend on the direction of deliberate visual attention. *Perception & Psychophysics*, *62*, 321–332. [PubMed]
- Bresciani, J.-P., Dammeier, F., & Ernst, M. O. (2006). Vision and touch are automatically integrated for the perception of sequences of events. *Journal of Vision*, *6*(5):2, 554–564, <http://journalofvision.org/6/5/2/>, doi:10.1167/6.5.2. [PubMed] [Article]
- Bresciani, J. P., Ernst, M. O., Drewing, K., Bouyer, G., Maury, V., & Kheddar, A. (2005). Feeling what you hear: Auditory signals can modulate tactile tap perception. *Experimental Brain Research*, *162*, 172–180. [PubMed]
- Calvert, G. A., Bullmore, E. T., Brammer, M. J., Campbell, R., Williams, S. C., McGuire, P. K., et al. (1997). Activation of auditory cortex during silent lipreading. *Science*, *276*, 593–596. [PubMed]
- Clark, J. J., & Yuille, A. L. (1990). *Data fusion for sensory information processing systems*. Boston: Kluwer Academic Publishers.
- Colin, C., Radeau, M., Soquet, A., Dachy, B., & Deltenre, P. (2002). Electrophysiology of spatial scene analysis: The mismatch negativity (MMN) is sensitive to the ventriloquism illusion. *Clinical Neurophysiology*, *113*, 507–518. [PubMed]
- Colin, C., Radeau, M., Soquet, A., Demolin, D., Colin, F., & Deltenre, P. (2002). Mismatch negativity evoked by the McGurk–MacDonald effect: A phonetic representation within short-term memory. *Clinical Neurophysiology*, *113*, 495–506. [PubMed]
- De Gelder, B., & Vroomen, J. (2000). The perception of emotions by ear and by eye. *Cognition & Emotion*, *14*, 289–311.
- Driver, J. (1996). Enhancement of selective listening by illusory mislocation of speech sound due to lip-reading. *Nature*, *381*, 66–68. [PubMed]
- Ernst, M. (2005). *A Bayesian view on multimodal cue integration* (chap. 6, pp. 105–131). New York: Oxford University Press.
- Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*, 429–433. [PubMed]
- Ernst, M. O., & Bühlhoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences*, *8*, 162–169. [PubMed]
- Gepshtein, S., & Banks, M. S. (2003). Viewing geometry determines how vision and haptics combine in size perception. *Current Biology*, *13*, 483–488. [PubMed] [Article]
- 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. <http://journalofvision.org/5/11/7/>, doi:10.1167/5.11.7. [PubMed] [Article]
- Helbig, H. B., & Ernst, M. O. (2007a). Knowledge about a common source can promote visual haptic integration. *Perception* 36, 1523–1533.
- Helbig, H. B., & Ernst, M. O. (2007b). Optimal integration of shape information from vision and touch. *Experimental Brain Research*, 179, 595–606. [PubMed]
- Hillis, J., Ernst, M., Banks, M., & Landy, M. (2002). Combining sensory information: Mandatory fusion within, but not between, senses. *Science*, 298, 1627–1630. [PubMed]
- Kayser, C., Petkov, C., Augath, M., & Logothetis, N. (2005). Integration of touch and sound in auditory cortex. *Neuron*, 48, 373–384. [PubMed] [Article]
- Knill, D. C., & Saunders, J. A. (2003). Do humans optimally integrate stereo and texture information for judgments of surface slant? *Vision Research*, 43, 2539–2558. [PubMed]
- Lavie, N. (1995). Perceptual load as a necessary condition for selective attention. *Journal of Experimental Psychology: Human Perception and Performance*, 21, 451–468. [PubMed]
- Macaluso, E., Frith, C. D., & Driver, J. (2000). Modulation of human visual cortex by crossmodal spatial attention. *Science*, 289, 1206–1208. [PubMed]
- Massaro, D. W. (1987a). Information-processing theory and strong inference: A paradigm for psychological inquiry. In H. Heuer & A. F. Sanders (Eds.), *Perspectives on perception and action* (pp. 273–299). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Massaro, D. W. (1987b). *Speech perception by ear and eye: A paradigm for psychological inquiry*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Massaro, D. W. (1998). *Perceiving talking faces: From speech perception to a behavioral principle*. Cambridge, MA: MIT Press.
- Näätänen, R. (1992). *Attention and brain function*. Hillsdale, NJ: Lawrence Erlbaum.
- Pekkola, J., Ojanen, V., Autti, T., Jaaskelainen, I. P., Mottonen, R., Tarkiainen, A., et al. (2005). Primary auditory cortex activation by visual speech: An fMRI Study at 3 t. *Neuroreport*, 16, 125–128. [PubMed]
- Prinzmetal, W., Amiri, H., Allen, K., & Edwards, T. (1998). Phenomenology of attention: I. Color, location, orientation, and spatial frequency. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 1–22.
- Prinzmetal, W., Nwachuku, I., Bodanski, L., Blumenfeld, L., & Shimizu, N. (1997). The phenomenology of attention. 2. brightness and contrast. *Consciousness and Cognition*, 6, 372–412. [PubMed]
- 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 of London B: Biological Sciences*, 273, 2159–2168. [PubMed] [Article]
- Shams, L., Kamitani, Y., & Shimojo, S. (2000). Illusions. What you see is what you hear. *Nature*, 408, 788. [PubMed]
- Shams, L., Kamitani, Y., & Shimojo, S. (2002). Visual illusion induced by sound. *Cognitive Brain Research*, 14, 147–152. [PubMed]
- Shams, L., Ma, W. J., & Beierholm, U. (2005). Sound-induced flash illusion as an optimal percept. *Neuroreport*, 16, 1923–1927. [PubMed]
- Shore, D. I., & Simic, N. (2005). Integration of visual and tactile stimuli: Top-down influences require time. *Experimental Brain Research*, 166, 509–517. [PubMed]
- Soto-Faraco, S., Navarra, J., & Alsius, A. (2004). Assessing automaticity in audiovisual speech integration: Evidence from the speeded classification task. *Cognition & Emotion*, 92, 13–23. [PubMed]
- Spence, C., & Driver, J. (2000). Attracting attention to the illusory location of a sound: Reflexive cross-modal orienting and ventriloquism. *Neuroreport*, 11, 2057–2061. [PubMed]
- Spence, C., Pavani, F., & Driver, J. (2004). Spatial constraints on visual-tactile cross-modal distractor congruency effects. *Cognitive, Affective, and Behavioral Neuroscience*, 4, 148–169. [PubMed]
- Spence, C., & Walton, M. (2005). On the inability to ignore touch when responding to vision in the crossmodal congruency task. *Acta Psychologica*, 118, 47–70. [PubMed]
- Stekelenburg, J. J., Vroomen, J., & de Gelder, B. (2004). Illusory sound shifts induced by the ventriloquist illusion evoke the mismatch negativity. *Neuroscience Letters*, 357, 163–166. [PubMed]
- van Atteveldt, N., Formisano, E., Goebel, R., & Blomert, L. (2004). Integration of letters and speech sounds in the human brain. *Neuron*, 43, 271–282. [PubMed] [Article]
- van Beers, R. J., Wolpert, D. M., & Haggard, P. (2002). When feeling is more important than seeing in sensorimotor adaptation. *Current Biology*, 12, 834–837. [PubMed] [Article]
- Vroomen, J., Bertelson, P., & de Gelder, B. (2001a). The ventriloquist effect does not depend on the direction of automatic visual attention. *Perception & Psychophysics*, 63, 651–659. [PubMed]
- Vroomen, J., Driver, J., & de Gelder, B. (2001b). Is cross-modal integration of emotional expressions independent

- of attentional resources? *Cognitive, Affective, & Behavioral Neuroscience*, *1*, 382–387. [[PubMed](#)]
- Warren, D. H., Welch, R. B., & McCarthy, T. J. (1981). The role of visual–auditory “compellingness” in the ventriloquism effect: Implications for transitivity among the spatial senses. *Perception & Psychophysics*, *30*, 557–564. [[PubMed](#)]
- Wichmann, F. A., & Hill, N. J. (2001). The psychometric function: I. fitting, sampling, and goodness of fit. *Perception & Psychophysics*, *63*, 1293–1313. [[PubMed](#)]
- Wickens, C. D. (1980). The structure of attentional resources. In R. Nickerson (Ed.), *Attention and performance* (vol. VIII, pp. 239–257). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Wickens, C. D. (1984). Processing resources in attention. In R. Parasuraman & R. Davies (Eds.), *Varieties of attention* (pp. 63–101). New York: Academic Press.
- Yuille, A. L., & Bülthoff, H. H. (1996). Bayesian theory and psychophysics. In D. Knill & W. Richards (Eds.), *Perception as Bayesian inference*. Cambridge, MA: Cambridge University Press.