

Temporal aspects of cue combination

Christa M. van Mierlo

Faculty of Human Movement Sciences, Vrije Universiteit,
Amsterdam, The Netherlands



Eli Brenner

Faculty of Human Movement Sciences, Vrije Universiteit,
Amsterdam, The Netherlands



Jeroen B. J. Smeets

Faculty of Human Movement Sciences, Vrije Universiteit,
Amsterdam, The Netherlands



The human brain processes different kinds of information (or cues) independently with different neural latencies. How does the brain deal with these differences in neural latency when it combines cues into one estimate? To find out, we introduced artificial asynchronies between the moments that monocular and binocular cues indicated that the slant of a surface had suddenly changed. Subjects had to detect changes in slant or to indicate their direction. We found that the cues were combined to improve performance even when the artificial asynchrony between them was about 100 ms. We conclude that neural latency differences of tens of milliseconds between cues are irrelevant because of the low temporal resolution of neural processing.

Keywords: slant, cue combination, neural latencies, temporal resolution

Citation: van Mierlo, C. M., Brenner, E., & Smeets, J. B. J. (2007). Temporal aspects of cue combination. *Journal of Vision*, 7(7):8, 1–11, <http://journalofvision.org/7/7/8/>, doi:10.1167/7.7.8.

Introduction

People use various kinds of information to make sense of visual input from the external world. For example, they estimate the slant or orientation of a surface from texture gradients, motion parallax, retinal shape, binocular disparity, and so on. The brain is believed to process various kinds of information (cues) in different visual pathways in the brain, with neural latencies that can differ by tens of milliseconds (Schmolesky et al., 1998). After such independent processing, the brain combines different cues for the same property into a single estimate that is more reliable than any of the estimates based on the individual cues (a weighted average; Ernst & Banks, 2002; Hillis, Watt, Landy, & Banks, 2004; Jacobs, 1999; Knill & Saunders, 2003; Landy, Maloney, Johnston, & Young, 1995; van Beers, Sittig, & Gon, 1999). The contribution of each cue to this estimate is thought to primarily be determined by its reliability, but it could also be influenced by other factors such as the likelihood of the value indicated by the estimate occurring, the consistency between different cues, or the correlation between the errors of the two cues (Hogervorst & Brenner, 2004; Knill & Saunders, 2003; Landy et al., 1995; Oruc, Maloney, & Landy, 2003).

Differences in neural latency between cues about unrelated attributes (such as color and motion) may be responsible for the large systematic errors that subjects make when trying to synchronize changes within such

cues (Arnold & Clifford, 2002; Aymoz & Viviani, 2004; Moutoussis & Zeki, 1997a, 1997b; Nishida & Johnston, 2002; Viviani & Aymoz, 2001; Wu, Kanai, & Shimojo, 2004; Zeki & Moutoussis, 1997). However, differences in neural latency need not occur only for cues that provide information on unrelated attributes. Cues that describe the same property or attribute of a stimulus are also likely to have different neural latencies (Greenwald, Knill, & Saunders, 2005). Do these timing differences lead to systematic errors when such cues are combined, or are there special mechanisms in the brain for preventing this?

Aymoz and Viviani (2004) found that people tend to make smaller systematic errors in synchronizing changes in color and movement when the changes were the consequence of another person's actions. They speculated that the action of the other person activated a specialized system that reestablishes synchrony within the brain by compensating for the neural delay between the cues. Bartels and Zeki (2006) showed that people were better at synchronizing cues that describe the same attribute than at synchronizing cues that describe different attributes. Thus, there might be special mechanisms for dealing with timing differences between cues that are normally combined, such as cues for the same property.

Several cross-modal studies suggest that precise synchrony of different cues might be irrelevant due to the relatively low temporal resolution of neural processing. Munhall, Gribble, Sacco, and Ward (1996) showed that the McGurk effect is robust for lags of up to 180 ms. Shams, Kamitani, and Shimojo (2002) found that subjects

perceive a single visual flash as two flashes when it is accompanied by two auditory beeps. Similarly, auditory sequences of beeps have been found to modulate the tactile perception of sequences of taps (Bresciani et al., 2005). These effects persisted even when the flashes, beeps, and taps were separated by more than 100 ms. Hence, delays of up to 100 ms seem to be tolerated when integrating cues across modalities. Is this also the case for cues for the same attribute within a single modality? In particular, can the benefits of combining cues (e.g., through weighted averaging) be obtained without precise temporal synchrony?

Experiment 1

We conducted an experiment in which we explored the sensitivity of cue combination for asynchronies between binocular and monocular slant cues. Subjects had to detect changes in the slant of a plane. The orientation of the plane was evident from binocular disparity and from monocular information. The binocular cue could vary independently of the monocular cue, so that either one cue or both cues could indicate a change in slant. When both cues indicated a change, the timing of the change could differ for the two cues, thus creating different artificial cue asynchronies. We studied how subjects' detection of the changes in slant varied as a consequence of the asynchrony between the cues. To determine whether subjects were really combining the cues for detecting the change, we compared their performance when both cues changed with the performance predicted by probability summation based on their performance when only the binocular or only the monocular cue changed (Hillis, Ernst, Banks, & Landy, 2002; Poom, 2002; Wuerger, Hofbauer, & Meyer, 2003).

Methods

Subjects

Six subjects (three women and three men) participated in the experiment. One subject was an author; the other five were volunteers who were naive with respect to the purpose of the experiment. All subjects had normal binocular vision; their stereo acuity was better than 60 arcsec (tested with Randot plates).

Apparatus and stimuli

A Silicon Graphics Onyx Reality Engine was used to present the stimuli on a CRT monitor (120 Hz; horizontal size: 39.2 cm, 815 pixels; vertical size: 29.3 cm, 611 pixels; spatial resolution refined with antialiasing techniques). The subject sat 40 cm from the monitor, resting his or her head on a chin rest. The subject wore liquid crystal shutter

spectacles that successively blocked each eye in synchrony with the refresh rate of the monitor (120 Hz) so that different images were shown to the left and right eye in rapid alternation. A new image was presented to each eye every 16.7 ms (60 Hz). The individual's interocular distance was taken into account when creating the image presented to each eye. As a result, both the subject's ocular convergence and the retinal images were appropriate for the stimulus at the simulated distance.

Stimuli

The stimuli were designed so that monocular and binocular information on changes in slant could be independently manipulated. The stimulus was a simulated red ring within which 10 dots were randomly distributed. The ring had an outer radius of 70 mm and a width of 10.5 mm. The dots had a diameter of 5 pixels. The dots were added to increase the strength of the binocular disparity cue. There were very few dots within the ring, so that the contribution of their distribution to the slant percept was probably negligible. Every 16.7 ms, the ring changed its position to a new random position within 20 mm of the center of the simulated surface (which coincided with the center of the screen). At the same time, new random positions were chosen for the dots. This prevented subjects from detecting slant changes on the basis of motion in the image. The ring and dots were presented to the left eye first and then to the right eye, before being replaced by a similar ring and dots at a slightly different position. Subjects perceived this stimulus as several rings that jittered on a plane. The slant of this plane was defined by the binocular disparities, the shape of the ring, and the distribution of the dots.

Because the ring's shape and the dots' distribution always indicated the same slant, we will refer to them together as a monocular cue. To change the slant indicated by the monocular and binocular cues independently, we determined how a surface with a slant defined by the monocular cue would look to a single (cyclopean) eye and then rendered images for the two eyes that, on average, provide this retinal image, while having the binocular slant that we wanted (Knill, 1998; Landy et al., 1995).

Most of the time, both the binocular and monocular cues suggested that the slant of the plane in which the rings seemed to jitter was 10° (base slant), with a positive angle meaning that the top is further away than the bottom. The slant could increase abruptly by 5°, 10°, 15°, 20°, or 25°, which means that the top always tilted further away. These slant changes could occur in the binocular cue alone, in the monocular cue alone, or in both cues. When both cues changed their slant, they could do this simultaneously or asynchronously with eight different timings. The change in the binocular disparity cue could occur 400, 200, 100, or 50 ms before or after the change in the monocular cue. The next slant change occurred between 4 and 6 s after the slant had returned to its baseline value for

both cues. The plane regained its base slant gradually within 400 ms; it returned slowly so that the subjects would not perceive this as a second change (see Figure 1).

Procedure

Subjects saw a set of rings jittering on a plane that occasionally changed its slant. They had to respond to any change in slant by pressing the right mouse button. No feedback was given. In total, there were 55 conditions: For each of the five amplitudes, there were nine 2-cue conditions with various asynchronies (including the 0-ms asynchrony) and two single-cue conditions. Subjects performed 20 trials per condition, 1,100 trials in total, distributed over several sessions. The slant change for each trial was randomly selected from these 55 conditions.

Data analysis

We considered subjects to have detected the change if they responded between 150 ms after the first cue changed and 1 s after the last cue changed (response interval). We determined the fraction of detected slant changes for each cue asynchrony. If the binocular and monocular cues are processed completely independently of one another and independently give rise to responses within the allocated time, the asynchrony between them should be irrelevant, and the probability of detecting a slant change when both cues change (P_{both}) is the chance of not missing the slant change in both cues, which can be calculated on the basis of the subject's performance for the single-cue conditions ($P_{\text{binocular}}$ and $P_{\text{monocular}}$):

$$P_{\text{both}} = 1 - (1 - P_{\text{binocular}})(1 - P_{\text{monocular}}). \quad (1)$$

If we find that performance for a particular two-cue condition is better than predicted by probability summation (P_{both} in Equation 1), we can conclude that the subjects

detected the corresponding slant changes better than was to be expected on the basis of simply having two chances to react. This better performance could be due to a clever cue combination because a combined estimate of the slant change can be more reliable than the estimates on the basis of the single cues. If such a better performance is found, then a comparison of the different cue asynchronies might reveal the temporal sensitivity of the cue combination.

For each cue asynchrony and amplitude, we used a paired t test to examine whether performance was better than predicted by probability summation (Equation 1). We also tested with a paired t test per cue asynchrony whether performance was poorer on the asynchronous than on the synchronous two-cue condition. Using t tests in this manner is a conservative way of determining whether performance differs between the conditions, because the benefit we can expect from cue combination depends on the relative resolution of the two cues, which is likely to differ between subjects. If we find that performance on the two-cue conditions is systematically better than predicted by probability summation and depends on the timing difference between the cues, we will have a strong indication that subjects combined the two cues into one estimate of the change in slant. Differences between the asynchronies will then reveal the temporal resolution of combining the cues.

We estimated uncertainty bounds for each subject (standard error of the mean of the binomial distribution) from the observed fraction of slant changes detected (P) and the number of samples in each condition ($n = 20$):

$$SEM = \frac{\sqrt{P(1-P)}}{\sqrt{n}}. \quad (2)$$

Results

Our subjects' average performance is displayed in Figure 2. On average, subjects detected changes in the binocular cue better than changes in the monocular cue

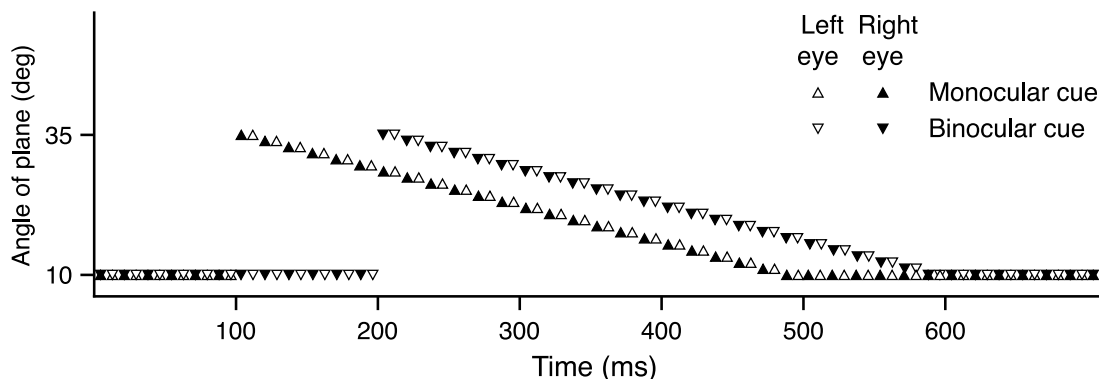


Figure 1. Schematic representation of one of the conditions of Experiment 1 in which both cues change by 25°, with a 100-ms delay between the changes. Each frame on the screen is represented by two symbols: one indicating the value of the monocular cue (the upward-pointing triangles) and another indicating the value of the binocular cue (the downward-pointing triangles).

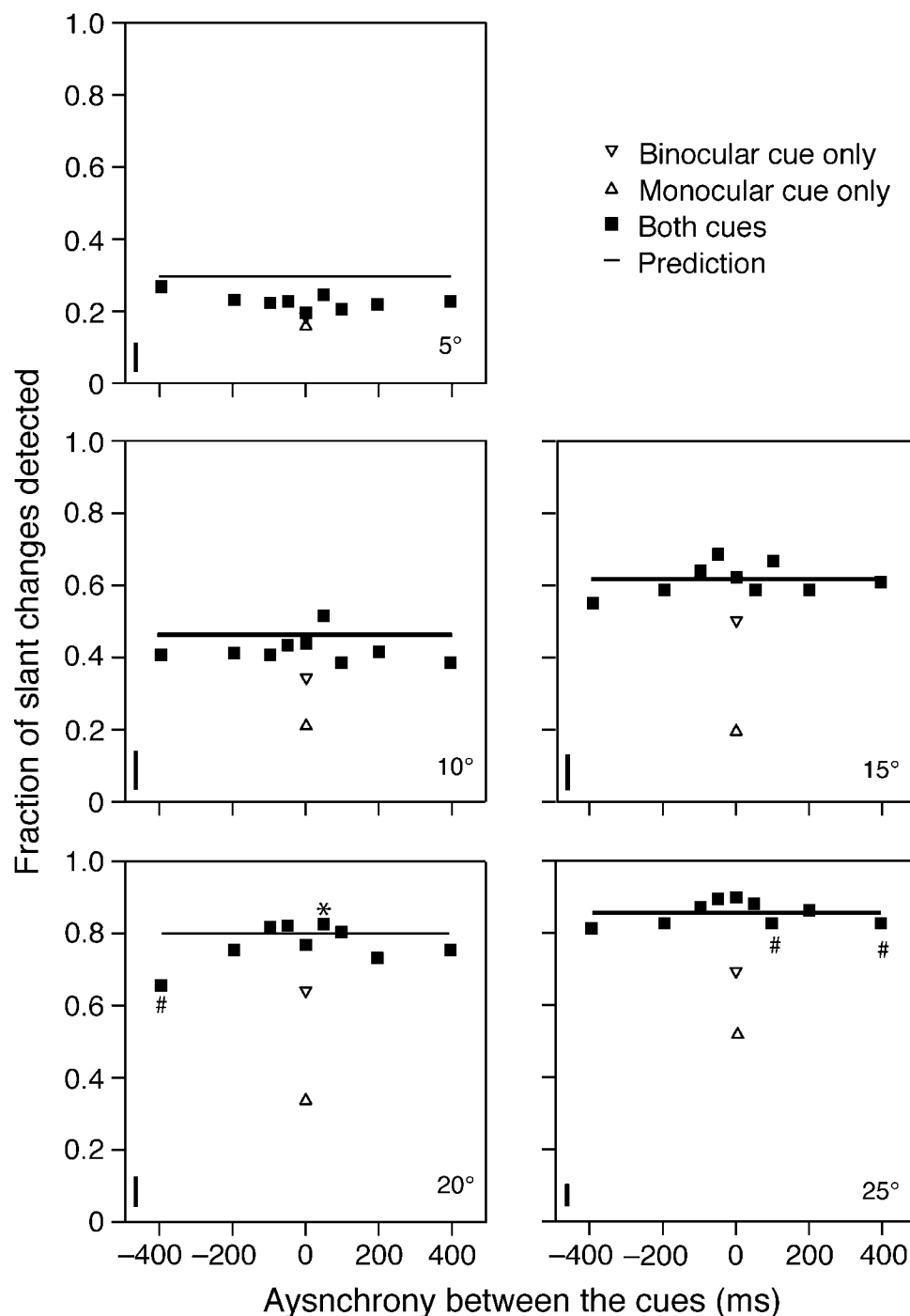


Figure 2. Average performance in Experiment 1 for the five different amplitudes of change. Positive values of the asynchrony indicate that the monocular cue changed after the binocular cue. The data for changes in a single cue are plotted at an asynchrony of 0 ms. The error bar at the bottom left of each graph is an estimate of the within-subjects standard error for the two-cue performance (averaged across asynchronies). *Significantly better performance than predicted by probability summation. #Significantly worse performance than for the 0-ms asynchrony.

(compare the upward and downward triangles in each panel). That is, subjects responded to 39% to 56% of the changes in the binocular cue and to 8% to 55% of the changes in the monocular cue. For slant changes in both cues, only detection of 20° slant changes with an asynchrony of +50 ms between the cues was significantly

better than predicted by probability summation, $t(5) = 3.08$, $p = .027$. One out of 45 comparisons being significant is fewer than one would expect by chance alone. For the small changes in slant (amplitudes of 5° and 10°), performance seemed to be systematically worse than predicted by probability summation.

For the larger amplitudes of change (15° , 20° , and 25°), a broad performance peak around the smaller cue asynchronies was apparent. The 20° slant changes with an asynchrony of -400 ms between the cues and the 25° slant changes with either a $+100$ - or a $+400$ -ms asynchrony between the cues were significantly less likely to be detected than synchronous slant changes of the same amplitude, $t(5) > -2.02$, $p < .05$. Three significant comparisons out of 40 is only one more than what one would expect by chance alone.

Discussion

Performance was significantly better than probability summation for only one amplitude–asynchrony condition. Performance often even appeared to be worse than predicted by Equation 1. A possible reason for this might be that Equation 1 does not consider false positives: correct responses that are independent of the actual change. In our analysis, we assumed that all responses that people made resulted from them really detecting the change. However, subjects sometimes seemed to misinterpret the jitter in the ring as a change in slant. We know that such false-positive responses occurred because we regularly observed responses long (>2 s) after the change had occurred. Presumably, these responses also occur when a change has taken place but was not detected. Because people are likely to make as many false-positive responses in the two-cue conditions as in each single-cue condition, Equation 1 will overestimate the predicted performance for the two-cue conditions because it incorporates the false-positive responses twice: once in $P_{\text{binocular}}$ and once in $P_{\text{monocular}}$. The conditions with larger asynchronies can be expected to contain slightly more false positives than the synchronous conditions because of their longer response intervals. Moreover, people are more likely to respond when they do not detect the target because once they have detected it, there will temporarily be no need to respond; thus, the number of false positives will depend on the subject's performance. We therefore propose that, as a result of ignoring false-positive responses in Equation 1, the predictions in Figure 2 lie higher than they should. This would explain why performance was no better than predicted by probability summation despite the apparent peak at small cue asynchronies for the larger amplitudes of slant change.

Experiment 2

If our impression that there is a peak in performance for small cue asynchronies is correct, then the peak's width suggests that an asynchrony of up to 100 ms between the cues hardly influences the benefit that is obtained from

combining the cues. However, this proposal rests on the assumption that we overestimated two-cue performance in Experiment 1 as a consequence of not accounting for false positives. In our second experiment, we therefore asked subjects to perform a task that allowed us to determine the number of false positives they made. They now had to indicate the direction of any slant change that they saw. Because subjects could respond both incorrectly and correctly when not responding to an actual change in slant, the term “false positives” no longer adequately describes such responses. Henceforth, we will refer to these responses as “guesses”. When subjects guess, about half of their guesses will be correct and half will be incorrect. The number of guesses is therefore twice the number of errors in indicating the direction of the change. By removing subjects' guesses from their responses before applying Equation 1, we can calculate predictions for probability summation in which guesses are considered.

Methods

Subjects

The same six subjects participated in the second experiment.

Apparatus, stimuli, and procedure

We used the same setup as in the previous experiment but made a few changes to the stimuli and procedure. In the previous experiment, our subjects detected changes in binocular disparity more easily than changes in the monocular cue. We therefore used a larger base slant (25° relative to frontal) to increase the reliability of the monocular cues (Knill, 1998). The change in slant always had an amplitude of 20° , but it could be in either direction. The combination of a 25° base slant and a $\pm 20^\circ$ change ensured that the surface never crossed the frontoparallel plane, which is important because doing so could make the changes in the monocular cue ambiguous or at least less clear. As in the previous experiment, slant changes could occur in binocular disparity, in the monocular cue, or in both, with time intervals ranging up to 400 ms. Subjects were instructed to indicate the direction of any slant changes that they detected. They pushed the left mouse button for “backward” slant changes and the right mouse button for “forward” slant changes (with the direction referring to the movement of the top of the surface).

Data analysis

Because choice reaction times are known to be longer than simple reaction times, we gave subjects slightly more time to respond. We determined the fraction of detected slant changes (both incorrect and correct responses) within

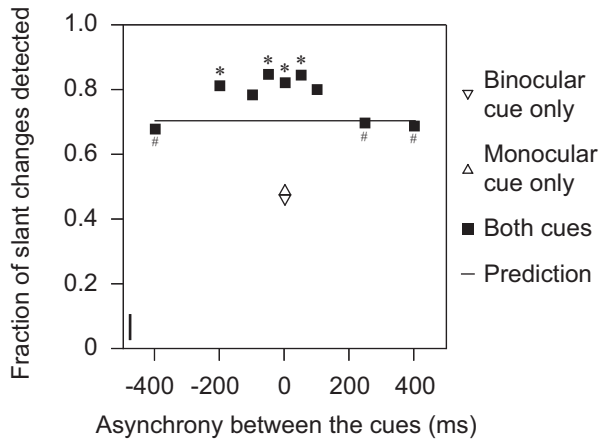


Figure 3. Average performance in Experiment 2. *Significantly better performance than predicted by probability summation. #Significantly worse performance than for the 0-ms asynchrony. Other details are as described in Figure 2.

an interval starting from 150 ms after the change in the first cue up to 1.2 s after the change in the last cue (when only one cue changed, it was both the first and the last). We assume that the responses that the subjects make consist of a number of real detections and a number of guesses. Because these guesses are as likely to be correct as incorrect, we assume that the number of guesses is twice the number of incorrect responses. Equation 1 only applies to the number of slant changes that subjects detected, not to their guesses. Thus, before we use this equation to predict two-cue performance, the fraction of guessed responses has to be removed from all the P values. The fraction of changes that were detected (P_d) can be estimated from the total fraction of trials in which the subject responded (P_r) and the fraction in which the subject responded incorrectly (P_e):

$$P_d = P_r - 2P_e. \quad (3)$$

Equation 3 holds independently for each condition; thus, substituting P_d for the P values in Equation 1 gives

$$P_{r_{\text{both}}} - 2P_{e_{\text{both}}} = 1 - (1 - (P_{r_{\text{monocular}}} - 2P_{e_{\text{monocular}}})) \times (1 - (P_{r_{\text{binocular}}} - 2P_{e_{\text{binocular}}})) \quad (4)$$

Equation 4 can be used to take guesses into account when predicting the fraction of responses for presentations with two cues ($P_{r_{\text{both}}}$) on the basis of single-cue performances ($P_{r_{\text{monocular}}}$ and $P_{r_{\text{binocular}}}$).

For each of the nine 2-cue conditions, we used paired t tests to examine whether the observed two-cue performance

was significantly better than the value predicted using Equation 4. We also used eight paired t tests to examine whether performance for each asynchrony was poorer than that for the synchronous slant changes.

Results

We first determined the fraction of incorrect responses (P_e) for each subject and each condition. An analysis of variance on P_e with Condition and Subject as factors revealed that P_e did not differ significantly between the different conditions ($F(10) = 0.784$, $p = .644$). Because it is important to get a reliable estimate of P_e , and the number of guesses is quite modest, we determined a single value for each subject and used this value for $P_{e_{\text{both}}}$, $P_{e_{\text{monocular}}}$, and $P_{e_{\text{binocular}}}$ in Equation 4.

Figure 3 shows the average fraction of slant changes that was detected. Subjects responded to 65.6% of the forward slant changes and 78.7% of the backward slant changes. For four of the nine asynchronies (−200, −50, 0, and 50 ms), the paired t tests revealed that performance was better than predicted by Equation 4, $t(5) > 2.038$, $p < .05$. Performance for the −400-, +200-, and +400-ms cue asynchronies was significantly poorer than that for the synchronous condition, $t(5) > 2.276$, $p < .05$.

Discussion

For asynchronies up to about 100 ms, performance with both cues was clearly better than predicted by probability summation. Probability summation did reliably predict performance for the largest cue asynchronies (± 400 ms), confirming that our analysis now addresses all major issues. Thus, the findings of Experiment 2 seem to support the weak evidence provided by Experiment 1 that subjects combine the cues even when the timing of the changes differs slightly between the cues. In addition, there appears to be a shift in the optimal delay toward negative asynchronies, which is consistent with binocular slant cues being processed faster than monocular slant cues (Greenwald et al., 2005). Cross-modal cue combination is also known to persist with asynchronies of slightly more than 100 ms (Munhall et al., 1996; Shams et al., 2002); hence, cues for the same property (slant) within a single modality (vision) do not appear to be treated in a special manner.

Experiment 3

In the third experiment, we examined the validity of two assumptions that we made when interpreting the data

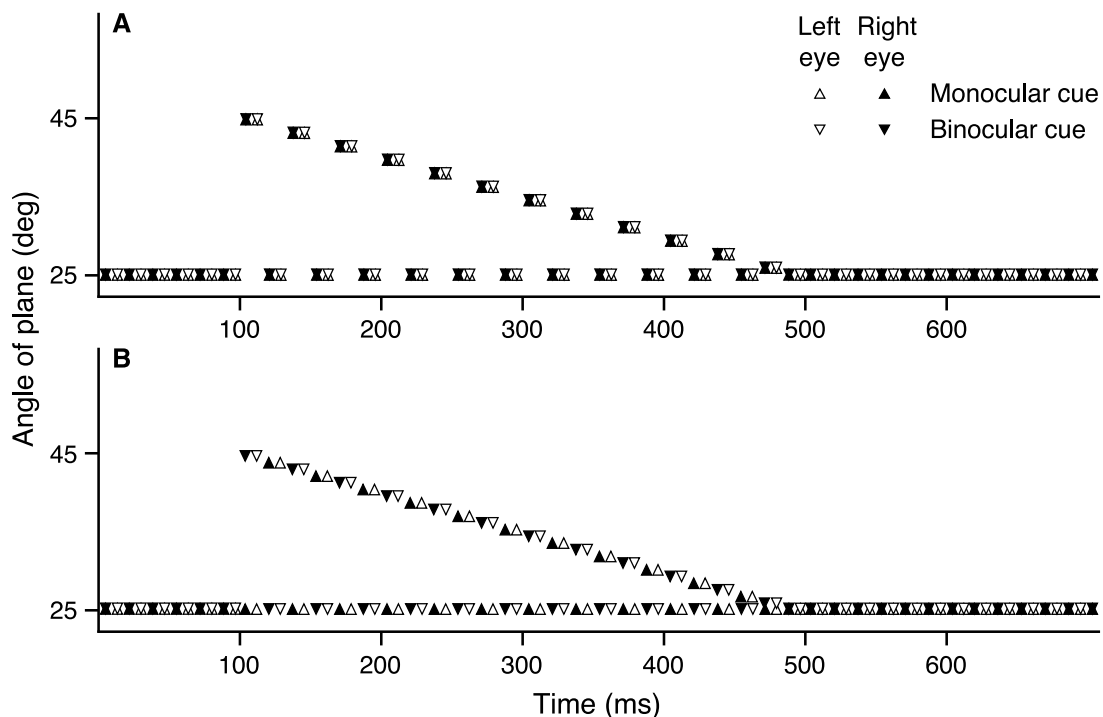


Figure 4. Schematic representation of the synchronous (A) and asynchronous (B) two-cue conditions of [Experiment 3](#). Both panels show a 20° slant change from a 25° base slant. In the synchronous condition, the cues are never in conflict, whereas in the asynchronous condition, they are in conflict whenever the surface is not at the base slant.

of [Experiments 1](#) and [2](#). The first is that the lack of change in one cue does not influence the detection of a slant change in the other cue. The second is that binocular disparities and retinal shape do not interact before providing estimates of slant. Further assumptions are discussed in the [General discussion](#) section.

Cue conflict in single-cue conditions

In the single-cue conditions, one cue changed its slant whereas the other remained in base slant, thus creating a cue conflict during the slant change. Our analysis was based on the assumption that for these conditions, the detection of a slant change in one cue was not affected by the unchanging slant of the other cue. On the basis of this assumption, we concluded from the findings of [Experiment 2](#) that detection of slant changes is better than predicted by probability summation when the two cues change in close temporal proximity. In the third experiment, we compared performance in two conditions that were identical to the single-cue conditions of [Experiments 1](#) and [2](#) with performance in two new single-cue conditions in which the slant conflict during the slant change was reduced. If performance in the new conditions is better than that in the original conditions, we would have to consider the possibility that performance in the previous experiments was not better when changes in

two cues were combined, but worse when one cue indicated that there was no change.

Independency of processing

Up until now, we have assumed that the two cues are processed independently before the brain combines them into one estimate of slant. Tittle and Braunstein (1993) suggested that this assumption might not hold for all cues within the visual system. For shape judgments from binocular disparity and motion parallax, they found that the presence of motion in a stereo display helped solve the binocular–correspondence problem. Thus motion helped to establish the binocular estimate of shape as well as providing an independent estimate of shape. Adams and Mamassian (2004) showed that texture information can also modulate shape from disparity in a way that is inconsistent with simple linear cue combination. We investigated whether the cues in our experiment interacted before they each supplied an independent estimate of slant with the help of two new two-cue conditions. In the first condition, the two cues signaled a slant change on alternate pairs of frames in rapid sequence (a pair of frames means one frame per eye). In the second condition, the two cues signaled the change simultaneously once every two pairs of frames (see [Figure 4](#)). In both cases, each cue alternates rapidly between the new slant and the

base slant, but in the asynchronous condition, the two cues are always in conflict, whereas in the synchronous condition, the cues always agree. If the cues do not interact before providing estimates of slant (and the cue combination process is not very sensitive to the precise timing of the estimates, as we have already seen), then performance in the two conditions should be the same. If we find better performance when the two cues change simultaneously, we would have evidence that the cues interact before they each generate an estimate of slant.

Methods

Subjects

Nine subjects (five women and four men) participated in this experiment. Six of the subjects had also participated in the former two experiments. All subjects had normal binocular vision; their stereo acuity was better than 60 arcsec (tested with Randot plates).

Apparatus, stimuli, and procedure

The same setup as in the former two experiments was used. We repeated the single-cue conditions from the second experiment for our modified procedure (see below) and added four new conditions. In a new monocular single-cue condition, the change in the monocular cue was presented to only one of the eyes (by simply not drawing the images for the other eye) so that there was no conflicting binocular disparity cue. In a new binocular

single-cue condition, only the dots that were previously used to fill the ring were visible. Omitting the ring practically eliminated the monocular cue so that the conflict was very much reduced, while leaving the binocular cue largely intact.

We introduced two additional conditions: an asynchronous and a synchronous two-cue condition. In the asynchronous condition, the slant changes were specified by both cues in rapid sequence. That is, in one frame, the monocular cue specifies base slant while the binocular cue indicates a changed slant, whereas in the next frame, the monocular cue indicates a changed slant and the binocular cue is in base slant. In the synchronous two-cue condition, both cues specify a changed slant simultaneously once every two pairs of frames (see Figure 4), with the other frame specifying the base slant.

As in Experiment 2, the base slant of the surface was 25° , and the ring could change its slant by $\pm 20^\circ$. To simplify the analysis, we changed our paradigm to a forced-choice procedure. Our subjects had to indicate the direction of the slant change after an auditory signal indicated that a slant change had occurred.

Data analysis

Due to the simplified procedure, we could just compare the proportion of correct responses between the different conditions. We used one-sided chi-square tests to examine whether there were more correct responses in the new binocular single-cue condition than

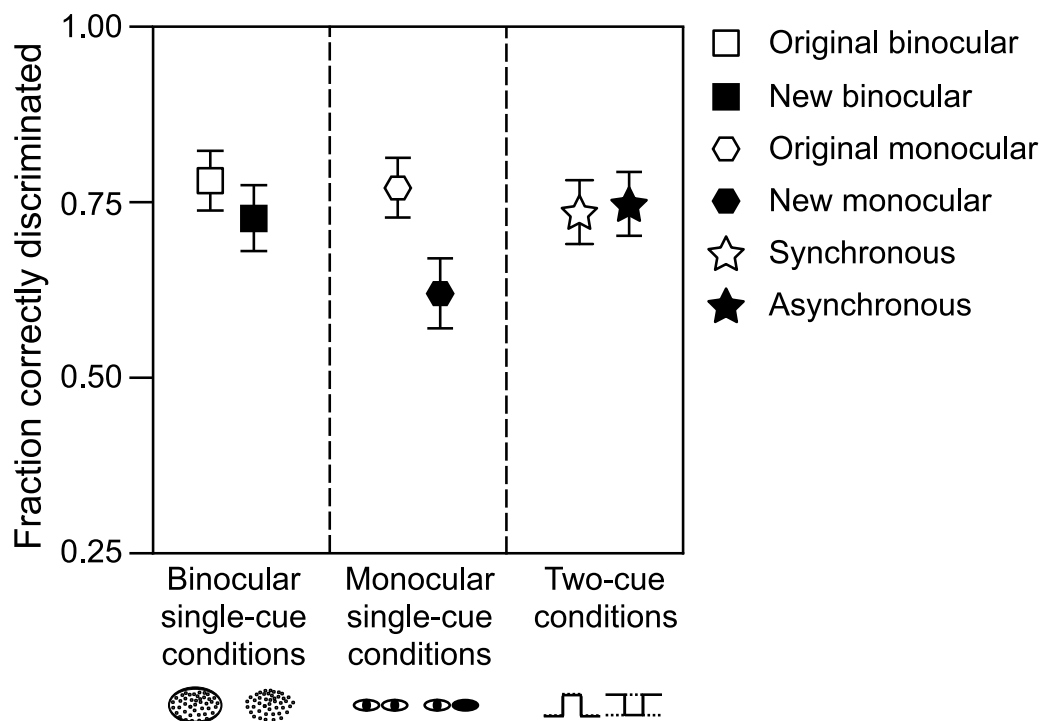


Figure 5. Average performance in Experiment 3. Error bars show 95% confidence intervals across subjects.

in the original binocular single-cue condition, in the new monocular single-cue condition than in the original monocular single-cue condition, and in the synchronous than in the asynchronous two-cue condition.

Results

Our subjects' average performance is displayed in [Figure 5](#). None of the three chi-square tests were significant. The average performance in the new single-cue conditions even seems to be worse than in the corresponding original single-cue conditions.

Discussion

Reducing the cue conflict in the single-cue conditions did not improve performance. It even decreased the number of correct responses that subjects made, especially in the monocular single-cue condition. This is probably because presenting the slant change to only one of the eyes doubled the interval between the frames in which it was present. Similarly, removing the ring in the new binocular single-cue condition reduced the amount of binocular information a bit and practically eliminated the monocular information. Thus, subjects had slightly less of the relevant information available in the new single-cue conditions than in the original single-cue (and two-cue) conditions of [Experiments 1](#) and [2](#). This questions the validity of our control experiment to some extent, but it is clear that the benefits of not having a cue that does not change (while the other does) do not outweigh the costs of reducing the amount of information in the new single-cue conditions.

Subjects' performance in the new asynchronous condition was no worse than in the new synchronous condition. We interpret this finding as indicating that there was no interaction between the cues before each provided an estimate of the slant change. Thus, there was, for instance, no interaction at the level of individual points, as when solving the correspondence problem.

General discussion

In this study, we examined how the detection of changes in surface slant was affected by artificial delays between binocular and monocular cues. We found a benefit for detecting two-cue slant changes beyond that predicted by probability summation even when the two cues changed at moments that differed by tens of milliseconds. This implies that neural latency differences between visual cues will seldom be an issue for the brain when it combines cues into one estimate. Apparently,

either the processing of the cues themselves or combining them into one estimate has quite a poor temporal resolution. However, this conclusion rests on one final assumption that needs to be discussed.

We assumed throughout the study that the cues are used independently to detect changes in slant and that if there is evidence from more than one cue that the slant of the surface in which the ring jittered changed, this evidence is combined to obtain a more reliable estimate of the change. Considering only information on changes in slant seems reasonable to us because the visual system is generally most sensitive to transients. However, there are alternative ways to combine the information provided by the two cues. Evidence for a change in one cue might be combined with evidence for no change in the other cue when the cues do not change simultaneously (for evidence against this option, see [Experiment 3](#)). It is also possible that the cues are continuously combined to give a single estimate of slant, and subjects detect changes in this combined estimate. Finally, subjects might notice a difference only between the changed combined estimate of slant and the baseline value. Would any of these three alternatives influence our conclusion that the temporal resolution of processing and combining changes in visual slant cues is poor?

If evidence for a change in one cue is combined with evidence for a lack of change from the other cue, the first of these two changes would be equivalent to a change in a single cue. When the second cue changes, the value of the first cue is different than it was in the single-cue conditions, but the change of the cue in question is identical to the change of that cue alone. The fact that the first cue is slowly changing back to its original baseline value might even slightly decrease the probability of detecting the change because it is in the opposite direction from the change that is to be detected. Thus, according to the first alternative, there is no reason to expect performance to be any better than predicted by probability summation. Our finding of an increased probability of detecting the slant changes when the two cues changed within tens of milliseconds of each other can only be consistent with this alternative if the changes are considered to overlap in time. [Experiment 3](#) provides additional evidence against this alternative.

The second alternative is that subjects detect changes in a combined estimate of slant. If so, asynchronous slant changes would be equivalent to two smaller slant changes. The first change is identical to a change in a single cue. However, the second change is not because it starts from a different perceived slant. It is not evident that this should make the changes easier (or more difficult) to detect because the change within the single relevant cue is exactly equivalent, even if the perceived initial and final slants of the second change are different. Moreover, if the perceived slant at the time of the second change influences performance, why should detection improve in a similar manner both when the perceived slant is higher than the baseline value of 10° in [Experiment 1](#) and when it is

sometimes higher and sometimes lower than 25° in [Experiment 2](#)? Again, the simplest explanation would be that the large range of asynchronies for which performance is better than probability summation arises because the temporal resolution of the processing underlying the judgment of slant is so low that the two changes merge into one larger slant change.

The third alternative is that subjects do not respond to transients (changes in the perceived slant) at all but sometimes notice that the slant is no longer at the baseline value. Because each cue's slant changed back gradually, the combined value would also change gradually, which could explain why subjects' performance was quite insensitive to timing differences between changes in the two cues without having to rely on a poor temporal resolution of visual processing or cue combination. We do not find this very likely because the visual system is generally most sensitive to transients, the task was to detect slant changes, and none of the subjects ever reported seeing two changes in rapid sequence. However, our findings do not rule out this possibility. In particular, performance for approximately synchronous changes in both cues is not evidently or systematically different from performance for twice the amplitude of the change in a single cue (see [Figure 2](#)). Thus, if subjects detect the change by noticing that the slant is no longer at baseline, rather than noticing the change itself, we cannot be certain from our study that the processes involved have a low temporal resolution. However, such a mechanism would make timing differences between cues a much less relevant issue because a temporal error of tens of milliseconds only introduces large differences between cues at moments at which the image suddenly changes, such as when one makes saccades or when there are quickly moving objects. Thus, asynchronies might be tolerated because they are only present for short periods.

Thus, we conclude that timing differences between cues are unlikely to be an important issue in human slant change perception. Probably, the temporal resolution of the processing of individual cues is so low that differences in timing can be ignored when combining cues. In daily life, external events will always cause cues to change in synchrony, so, differences in timing between visual cues within the brain only arise from differences in neural processing time. Because the reported differences in visual processing time (Schmolesky et al., 1998) are modest when compared with the low temporal resolution that we find for cue combination, it is unlikely that these differences in processing time have much influence on our perceptual judgments or need to be compensated for.

Acknowledgments

This study was conducted at the Neuroscience Department of the Erasmus MC in Rotterdam, The Netherlands.

This research was supported by The Netherlands Organisation for Scientific Research (NWO; MaGW Grant 452-02-007).

Commercial relationships: none.

Corresponding author: Christa M. van Mierlo.

Email: c.vanmierlo@fbw.vu.nl.

Address: Faculteit der Bewegingswetenschappen/Faculty of Human Movement Sciences, Vrije Universiteit, Van der Boechorststraat 9, 1081 BT Amsterdam, The Netherlands.

References

- Adams, W. J., & Mamassian, P. (2004). Bayesian combination of ambiguous shape cues. *Journal of Vision*, 4(10):7, 921–929, <http://journalofvision.org/4/10/7/>, doi:10.1167/4.10.7. [[PubMed](#)] [[Article](#)]
- Arnold, D. H., & Clifford, C. W. (2002). Determinants of asynchronous processing in vision. *Proceedings of the Royal Society of London Series B: Biological Sciences*, 269, 579–583. [[PubMed](#)] [[Article](#)]
- Aymoz, C., & Viviani, P. (2004). Perceptual asynchronies for biological and non-biological visual events. *Vision Research*, 44, 1547–1563. [[PubMed](#)]
- Bartels, A., & Zeki, S. (2006). The temporal order of binding visual attributes. *Vision Research*, 46, 2280–2286. [[PubMed](#)]
- 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](#)]
- Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415, 429–433. [[PubMed](#)]
- Greenwald, H. S., Knill, D. C., & Saunders, J. A. (2005). Integrating visual cues for motor control: A matter of time. *Vision Research*, 45, 1975–1989. [[PubMed](#)]
- 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. [[PubMed](#)]
- 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(12):1, 967–992, <http://journalofvision.org/4/12/1/>, doi:10.1167/4.12.1. [[PubMed](#)] [[Article](#)]
- Hogervorst, M. A., & Brenner, E. (2004). Combining cues while avoiding perceptual conflicts. *Perception*, 33, 1155–1172. [[PubMed](#)]
- Jacobs, R. A. (1999). Optimal integration of texture and motion cues to depth. *Vision Research*, 39, 3621–3629. [[PubMed](#)]

- Knill, D. C. (1998). Discrimination of planar surface slant from texture: Human and ideal observers compared. *Vision Research*, 38, 1683–1711. [[PubMed](#)]
- 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](#)]
- Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. (1995). Measurement and modeling of depth cue combination: In defense of weak fusion. *Vision Research*, 35, 389–412. [[PubMed](#)]
- Moutoussis, K., & Zeki, S. (1997a). A direct demonstration of perceptual asynchrony in vision. *Proceedings of the Royal Society of London Series B: Biological Sciences*, 264, 393–399. [[PubMed](#)] [[Article](#)]
- Moutoussis, K., & Zeki, S. (1997b). Functional segregation and temporal hierarchy of the visual perceptive systems. *Proceedings of the Royal Society of London Series B: Biological Sciences*, 264, 1407–1414. [[PubMed](#)] [[Article](#)]
- Munhall, K. G., Gribble, P., Sacco, L., & Ward, M. (1996). Temporal constraints on the McGurk effect. *Perception & Psychophysics*, 58, 351–362. [[PubMed](#)]
- Nishida, S., & Johnston, A. (2002). Marker correspondence, not processing latency, determines temporal binding of visual attributes. *Current Biology*, 12, 359–368. [[PubMed](#)] [[Article](#)]
- Oruc, I., Maloney, L. T., & Landy, M. S. (2003). Weighted linear cue combination with possibly correlated error. *Vision Research*, 43, 2451–2468. [[PubMed](#)]
- Poom, L. (2002). Visual binding of luminance, motion and disparity edges. *Vision Research*, 42, 2577–2591. [[PubMed](#)]
- Schmolesky, M. T., Wang, Y., Hanes, D. P., Thompson, K. G., Leutgeb, S., Schall, J. D., et al. (1998). Signal timing across the macaque visual system. *Journal of Neurophysiology*, 79, 3272–3278. [[PubMed](#)] [[Article](#)]
- Shams, L., Kamitani, Y., & Shimojo, S. (2002). Visual illusion induced by sound. *Cognitive Brain Research*, 14, 147–152. [[PubMed](#)]
- Tittle, J. S., & Braunstein, M. L. (1993). Recovery of 3-D shape from binocular disparity and structure from motion. *Perception & Psychophysics*, 54, 157–169. [[PubMed](#)]
- van Beers, R. J., Sittig, A. C., & Gon, J. J. (1999). Integration of proprioceptive and visual position-information: An experimentally supported model. *Journal of Neurophysiology*, 81, 1355–1364. [[PubMed](#)] [[Article](#)]
- Viviani, P., & Aymoz, C. (2001). Colour, form, and movement are not perceived simultaneously. *Vision Research*, 41, 2909–2918. [[PubMed](#)]
- Wu, D. A., Kanai, R., & Shimojo, S. (2004). Vision: Steady-state misbinding of colour and motion. *Nature*, 429, 262. [[PubMed](#)]
- Wuerger, S. M., Hofbauer, M., & Meyer, G. F. (2003). The integration of auditory and visual motion signals at threshold. *Perception & Psychophysics*, 65, 1188–1196. [[PubMed](#)]
- Zeki, S., & Moutoussis, K. (1997). Temporal hierarchy of the visual perceptive systems in the Mondrian world. *Proceedings of the Royal Society of London Series B: Biological Sciences*, 264, 1415–1419. [[PubMed](#)] [[Article](#)]