

---

# Correspondence matching in long-range apparent motion precedes featural analysis

---

Nicolaas Prins

Department of Psychology, University of Mississippi, Oxford, MS 38677, USA;  
e-mail: nprins@olemiss.edu

Received 9 March 2007, in revised form 20 October 2007

---

**Abstract.** It has been suggested that correspondence matching in long-range motion is mediated by a perceptually high-level, ‘intelligent’ system. This suggestion is based on findings that long-range motion can be perceived between stimuli that could not be detected by lower-level motion mechanisms acting on Fourier motion energy, and that correspondence matching is affected by featural similarities between motion tokens that would be invisible to low-level (Fourier) motion detectors. Here, the effects of spatial-frequency content, color, and binocular disparity on correspondence matching are investigated. It is shown that the effects of featural matches between motion tokens develop only over time and lag behind the effects of the relative proximity between motion tokens in the retinal projection. This suggests that correspondence matching in long-range apparent motion is mediated by a mechanism which acts initially on the retinal coordinates of the motion tokens only, but may be biased to favor matching tokens that are featurally similar through a slower top–down influence by higher-level processes.

## 1 Introduction

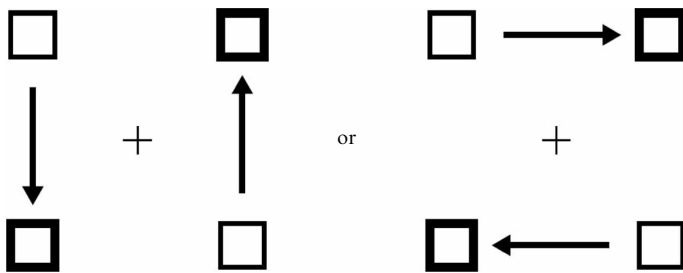
Braddick (1973, 1974) proposed that two distinct mechanisms exist by which visual motion is detected. The short-range system is presumably mediated by direction-selective neurons operating locally to extract the Fourier motion energy present in the spatio-temporal luminance distribution across the neuron’s receptive field (eg Reichardt 1969; Adelson and Bergen 1985). The long-range system, on the other hand, was proposed to be mediated by higher-level, ‘intelligent’, interpretative processes. The long-range system signals motion across much greater distances and temporal lags than the short-range system. Although the validity of Braddick’s proposal has since been contested (eg Cavanagh and Mather 1989), and more elaborate classifications have been proposed (eg Lu and Sperling 1995), the distinction has persisted in the motion literature. The present investigation concerns what would, by most, be considered long-range motion. The distance between motion tokens in the experiments reported here is on the order of degrees, beyond the range of the short-range process. Moreover, observers report perceiving motion that could not be signaled by detectors acting on Fourier motion energy. Rather, in Lu and Sperling’s (1995) motion-classification scheme, the motion reported by observers here relies on tracking of salient features which are first extracted from the image. Lu and Sperling’s ‘feature tracking’ and Braddick’s (1974) ‘long-range motion’ are considered by many (eg Scott-Samuel and Georgeson 1999) to be interchangeable terms.

The simplest example of a long-range motion perception is ‘phi motion’, which results when two stimuli are presented asynchronously at two different locations. Using appropriate spatiotemporal parameters, observers will report perceiving smooth motion of a single stimulus between the two locations (eg Wertheimer 1912). Apparently, the visual system assigns ‘correspondence’ between the two stimuli, in that it interprets the display as two successive glimpses of a single stimulus, rather than as the asynchronous appearance and disappearance of two separate stimuli.

Two stimuli presented in an apparent-motion sequence need not have the same physical appearance in order for motion to be perceived between them. Motion may

be perceived between two tokens that differ in color (eg Kolers and von Grünau 1976; Kolers and Green 1984; Caelli et al 1993; Oyama et al 1999), orientation (eg Burt and Sperling 1981; Green 1986; Mack et al 1989; Ullman 1980), shape (eg Kolers and Pomerantz 1971; Kolers and von Grünau 1976; Navon 1976; Shechter et al 1988; Mack et al 1989; Zhuo et al 2003), size (eg Mack et al 1989; Schechter and Hochstein 1989), spatial frequency (eg Green 1986), and luminance polarity (eg Shechter and Hochstein 1990). Moreover, long-range apparent motion may be perceived between two subjective figures (Mather 1988; Koriat 1994) and between figures defined solely by binocular disparity (He and Nakayama 1994; Prins and Juola 2001). The perception of long-range apparent motion is also influenced by intention (Suzuki and Peterson 2000). These results confirm the notion that long-range apparent motion is mediated by higher-order, interpretative or 'intelligent' processes rather than by low-level motion detectors acting on Fourier motion energy.

The correspondence problem in long-range apparent motion arises when two or more motion tokens are presented in each stimulus frame (Ullman 1979). The visual system needs to assign correspondence between each token in one frame and one of a number of possible matches in the next frame. A straightforward example of a motion stimulus which presents the correspondence problem is shown in figure 1. This configuration, first used by Gengerelli (1948), consists of two alternating frames. Each frame contains two motion tokens which are diagonally opposed on the corners of a rectangle. Between frames, the motion tokens are positioned on alternating diagonals of the rectangle. For each motion token in one frame, two possible candidate matches exist in the alternate frame. Hence, one of two possible motion paths may be perceived. The two different possible interpretations of the scene compete for precedence (Burt and Sperling 1981) and the result is a bistable percept: only one of the two possible solutions to the correspondence problem will be perceived at any given moment and once a solution is generated it tends to persist with only occasional switches between solutions.



**Figure 1.** The motion correspondence problem. The motion tokens drawn with different borders are presented in alternate frames. For each motion token in any frame, two possible matches exist in the alternate frame. Observers may perceive vertical motion (left) or horizontal motion (right).

Much research has concentrated on determining the principles by which the visual system assigns correspondence between motion tokens in alternate frames. One important factor is the relative distance between a token and its potential matches: all else equal, correspondence between tokens will be assigned such that motion occurs along the shortest path (eg Ullman 1979; Burt and Sperling 1981). This is referred to as the nearest-neighbor principle (eg Dawson 1991). The nearest-neighbor principle acts on the relative distances in the 2-D retinal projection of the display and not on relative distances in an object-centered 3-D representation of the stimulus configuration (Prins and Juola 2001), although 3-D distances between tokens have been shown to affect the correspondence-matching process, especially when the retinal locations of the tokens are chosen such that neither percept is strongly preferred over the other (eg Green and Odom 1986; Prins and Juola 2001).

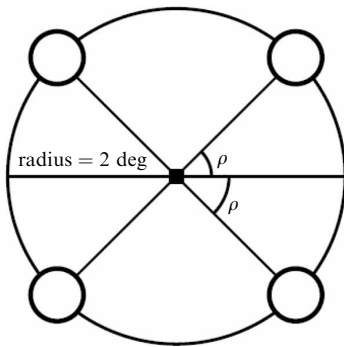
It was noted above that correspondence may be assigned between two tokens that are featurally dissimilar on any of a number of dimensions. How physical similarities between tokens affect the correspondence process has been investigated in many studies. Whereas abundant and rather uncontroversial evidence exists that featural properties affect the perceived quality of motion in non-competitive motion displays (eg Kolers and Pomerantz 1971; Ullman 1980; Caelli et al 1993), the evidence regarding the effect of featural properties on the correspondence problem in competitive motion displays is far less conclusive. One reason for the discrepant results obtained with competitive motion displays is that the effects of featural properties of the motion tokens may be overlooked unless one percept is not strongly favored over the other, on the basis of the retinal coordinates of the motion tokens alone.

Prins and Juola (2001) hypothesized that correspondence matching in long-range apparent motion is initially performed by an automatic (eg Schneider and Schiffrin 1977) low-level mechanism that acts on the retinal coordinates of the tokens only, and that higher-level, more 'intelligent', processes may influence the lower-level mechanism, through a controlled (eg Schneider and Schiffrin 1977) top-down process. Such a scheme would predict that the influence of featural properties of tokens, being mediated by a slow, controlled process, would require more time to develop than the effect of the retinal coordinates of the motion tokens. The purpose of the current series of experiments is to test this prediction. An ambiguous motion display was used for which observers could generate one of two possible solutions to the correspondence problem. In separate experiments the relative spatial-frequency content, color, or binocular disparity of the motion tokens were varied. For one of the two solutions to the correspondence problem the featural integrity of tokens remained intact whereas for the other the tokens changed appearance between frames. The actual motion sequence consisted of two frames, presented either immediately or following a static frame containing all motion tokens simultaneously, which presumably would allow higher-order processes enough time to exert an influence on the assumed low-level motion mechanism. In all experiments reported here the effect of featural similarity is quantified by its trade-off with spatial proximity of tokens.

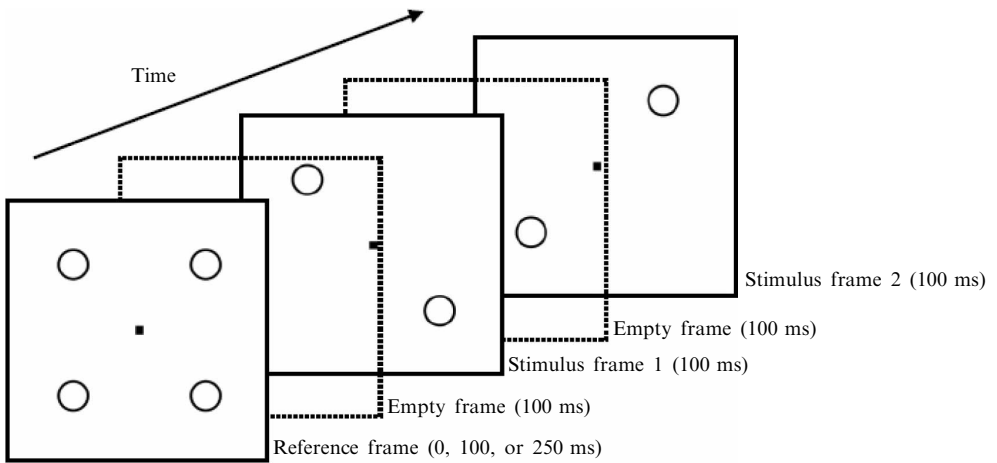
The main finding is that no evidence is found that either the relative spatial frequency content, color, or binocular disparity of motion tokens exerts any effect on which solution to the correspondence problem is generated when little time is allowed for these features to exert their influence. The finding that featural properties affect correspondence matching is replicated only when the information regarding the featural properties is available some time before correspondence needs to be assigned.

## 2 General methods

In all three experiments, the motion tokens were centered on the corners of an imaginary rectangle (figure 2). The height and width of the rectangle varied from trial to trial under the constraint that the corners of the rectangle (and hence the motion tokens) were located on an imaginary circle of radius 2 deg. The relative height and width of the rectangle were quantified here as the absolute angle between the horizon and the imaginary lines connecting diagonally opposed tokens ( $\rho$  in figure 2). The positions of the motion tokens could be varied with a resolution of 0.1 pixels (0.17 min of arc), except where noted. Each trial consisted of the following sequence of frames (schematically illustrated in figure 3): a frame containing all four motion tokens simultaneously ('reference frame', variable period of time: 0, 100, or 250 ms), empty frame (100 ms), stimulus frame (100 ms), empty frame (100 ms), stimulus frame (100 ms), empty frame (until response). These spatiotemporal parameters led to the perception of smooth motion. Each of the stimulus frames contained two motion tokens that differed either in spatial-frequency content (experiment 1), chromaticity (experiment 2), or binocular



**Figure 2.** Display geometry for all experiments. Tokens were positioned on the perimeter of an imaginary circle (radius = 2 deg) as well as on the corners of an imaginary rectangle. Angle  $\rho$  was varied according to a staircase procedure to determine the equivalence angle—the angle  $\rho$  at which horizontal and vertical motion were perceived with equal probability.



**Figure 3.** Schematic diagram of a trial. A trial started with the presentation of a reference frame containing all four motion tokens for a variable period of time (reference frame duration = 0, 100, or 250 ms), followed by an empty frame (100 ms), the first stimulus frame containing two diagonally opposed motion tokens (100 ms), an empty frame (100 ms), and the second stimulus frame containing two motion tokens diagonally opposed on the other diagonal (100 ms). The fixation mark was presented throughout.

disparity (experiment 3) on opposite corners of the diagonal of the rectangle defined by the positions of the tokens. The alternate stimulus frame again contained two motion tokens differing in appearance, this time diagonally opposed on the alternate diagonal. Featural properties of motion tokens between frames matched either across the horizontal or vertical sides of the imaginary rectangle defined by the token positions. The featural properties of motion tokens in the reference frame matched those in the stimulus frames.

Whether the motion tokens were positioned on the left or right diagonal in the first stimulus frame was determined randomly for each trial. In the horizontal-match condition, it was determined randomly for each trial which of the two possible token forms appeared on the top side of the rectangle. Similarly, in the vertical-match condition, it was determined randomly for each trial which of the two possible token forms appeared on the left side of the rectangle. In each of the experiments, there were six coded experimental conditions: 3 reference frame durations (RFD; 0, 100, or 250 ms)  $\times$  2 possible token distributions (TD; identities match across horizontal or vertical sides).

Trials were presented in blocks of 120 trials; 20 repetitions of each of the six conditions. The dependent variable in all experiments is the 'equivalence angle': the angle  $\rho$  (figure 2) at which horizontal motion and vertical motion were equally likely to be perceived. The underlying idea is that any biasing effect resulting from TD will reveal itself by trading off with relative proximity of the tokens (eg Burt and Sperling 1981;

Green and Odom 1986; Shechter et al 1988; Shechter and Hochstein 1989; He and Nakayama 1994). The angle determining the spatial positions of the motion tokens was varied for each condition individually by a staircase procedure (the BEST Pest; Pentland 1980). The six staircases were randomly interleaved.

The author (NP) and five naive observers participated as observers. Stimuli were generated with a custom-written Matlab program, and presented on a Mitsubishi Diamond-pro 2070SB monitor, running at  $800 \times 600$  pixel resolution and frame rate 120 Hz, driven by a Cambridge Research Systems VSG 2/5 graphics board. Viewing distance was 1 m.

For each of the observers, the trials for each condition were combined across blocks. Logistic functions relating the probability of perceiving motion along horizontal paths to angle  $\rho$  were then fit using a maximum-likelihood criterion to estimate the 'equivalence angle' (the angle  $\rho$  at which horizontal motion and vertical motion are equally likely to be perceived) and the slope of the logistic function. All estimates are based on a minimum of 80 trials per condition. A parametric bootstrap procedure (eg Efron and Tibshirani 1993;  $B = 400$ ) was then used to derive standard errors of the equivalence angles. The statistical reliability of the effects of TD, RFD, and their interaction (X) was assessed by bootstrap simulations (the details of which are presented in the appendix).

### 3 Experiment 1: Spatial frequency

#### 3.1 Stimuli

The two different motion tokens were created by either low-pass or high-pass filtering a bright disk with a radius of 0.4 deg. The luminance of the disk (before filtering) was  $87.3 \text{ cd m}^{-2}$  set against a gray background ( $53.5 \text{ cd m}^{-2}$ ). Subpixel resolution was achieved by defining the edge of the disk with a logistic function. Specifically, the disk's luminance profile (pre-filtering) in  $\text{cd m}^{-1}$  was defined by:

$$\text{Lum}(r) = 53.5 + 33.8 \left( 1 + \exp \frac{r - 0.4}{0.028} \right)^{-1},$$

where  $r$  is the distance from the disk's center (in deg).

The filter profile of the low-pass filter was defined as:

$$L_{\text{LP}}(f) = \exp \left[ - \left( \frac{f}{1.45} \right)^4 \right],$$

where  $f$  is spatial frequency in  $\text{cycles deg}^{-1}$ . The filter profile of the high-pass filter was the complement of the low-pass-filter profile:

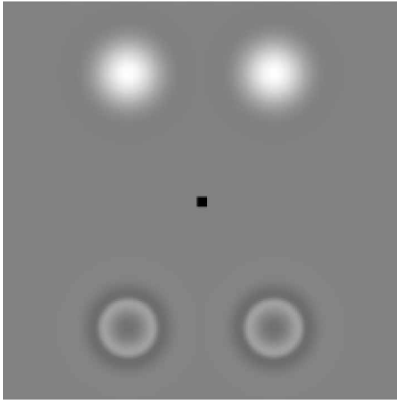
$$L_{\text{HP}}(f) = 1 - \exp \left[ - \left( \frac{f}{1.45} \right)^4 \right].$$

The 50% cut-off frequency for both filters is  $1.32 \text{ cycles deg}^{-1}$ . Example motion tokens thus defined are shown in figure 4.

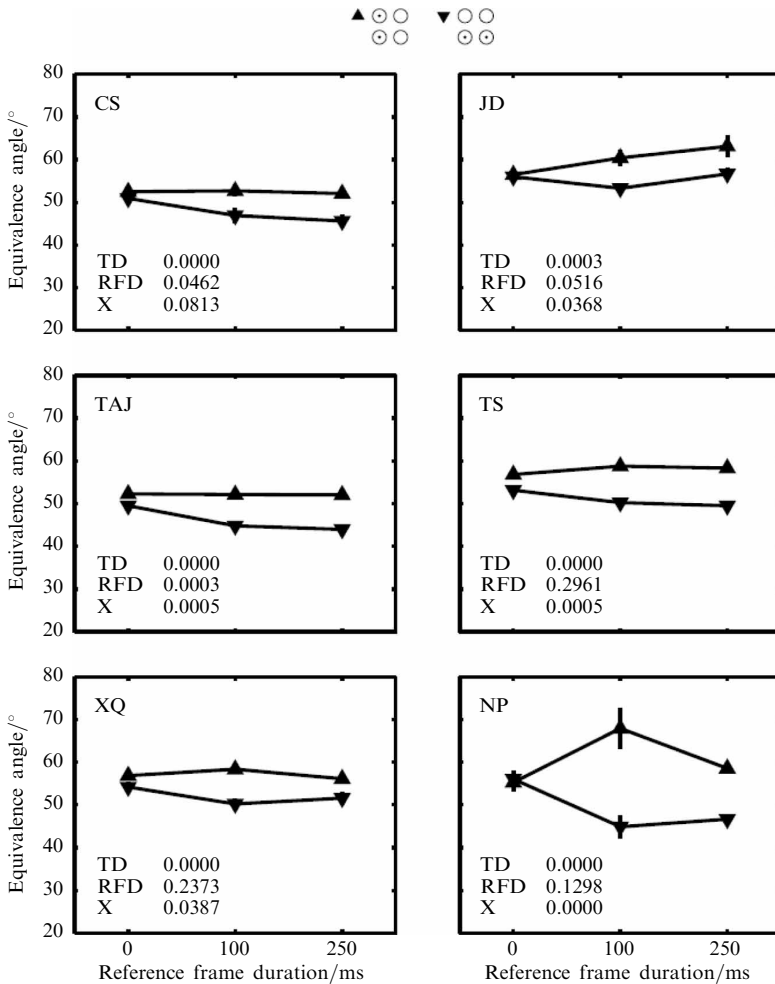
Angle  $\rho$ , determining the position of the motion tokens, was constrained to have a value between  $10^\circ$  and  $80^\circ$  and was varied with a resolution of  $0.25^\circ$ . When tokens overlapped spatially in the reference frame (ie at values of  $\rho$  close to  $10^\circ$  and  $80^\circ$ ), their luminance modulations were added. This occurred only in a few trials early in each block before the staircase procedure started to converge on a threshold. At none of the observed threshold values for  $\rho$  did the tokens overlap.

#### 3.2 Results

Equivalence angles with their standard errors are presented in figure 5, for each of the observers individually (where standard error bars are not visible they do not exceed the graphing symbol). Also presented in the figures are  $p$ -values assessing the statistical



**Figure 4.** A reference frame used in experiment 1. Tokens are a high-pass and a low-pass filtered version of a bright disk (see text for details). Angle  $\rho = 60^\circ$  in this figure.



**Figure 5.** Results of experiment 1, in which tokens differed with regard to their spatial-frequency content. Thresholds and standard errors for obtained equivalence angles are shown for each observer as a function of token distribution (TD; spatial frequencies match horizontally or vertically, see legend) and reference frame duration (RFD; 0, 100, or 250 ms). Where standard error bars are not visible, they do not exceed symbols. Also shown are  $p$ -values, based on bootstrap simulations (see appendix) for the main effects of TD and RFD and for the interaction between TD and RFD (X).

reliability of the main effects of TD, RFD, and their interaction (X). These  $p$ -values are based on bootstrap simulations, and should be interpreted as standard statistical  $p$ -values. The details of the bootstrap simulations are presented in the appendix. Based on the nearest-neighbor principle alone, one would expect equivalence angles to have a value of  $45^\circ$  (where motion tokens lie on the corners of an imaginary square). Instead, almost all equivalence angles have a value greater than  $45^\circ$ , indicating an overall preference for perceiving vertical motion over horizontal motion. This anisotropy replicates that found by Gengerelli (1948), who argued that this preference can be related to a stronger affinity between stimuli represented in the same cerebral hemisphere—a finding later confirmed by Chaudhuri and Glaser (1991).

From figure 5, it is apparent that TD (featural properties of tokens match horizontally or vertically) and RFD (0, 100, or 250 ms) interact: for all observers the effect of TD increases with RFD. This interaction is statistically significant for five of the six observers as indicated by the  $p$ -value associated with the interaction (X). It is important to note that for three of the observers (CS, JD, and NP) no effect of token distribution on equivalence angles was observed at all when the reference frame was not presented before the onset of the motion sequence. In other words, when the tokens were positioned such that  $\rho$  was equal to the equivalence angle, horizontal motion would be perceived on 50% of the trials, regardless of whether the figural integrity of the motion tokens was preserved along the motion path or not. Where it was observed, the direction of the effect of TD was consistent with the idea that matches between tokens containing like spatial frequencies are preferred over matches between tokens containing different spatial frequencies.

## 4 Experiment 2: Color

### 4.1 Stimuli

In experiment 2, motion tokens were defined by color. For each individual observer, a red–green subjective equiluminance point was established through a flicker-minimization technique. Observers minimized the appearance of flicker while the display presented red–green flicker (isolating the CRT’s red and green phosphors, respectively). The flicker was temporally square-wave at 20 Hz. The luminance of the red half-cycle remained constant at  $11.6 \text{ cd m}^{-2}$ , while the observer, through the method of limits, varied the luminance of the green half-cycle to minimize the appearance of flicker.

Motion tokens were then created by modulating the red and green phosphors simultaneously in antiphase as follows:

$$L_r = L_{rm} [1 + p_c G(x, y)],$$

$$L_g = L_{gm} [1 - p_c G(x, y)],$$

where  $L_r$  and  $L_g$  are the red and green phosphor luminances, respectively;  $L_{rm}$  is the mean red phosphor luminance ( $11.6 \text{ cd m}^{-2}$ ), and  $L_{gm}$  is the green phosphor luminance subjectively matched to  $L_{rm}$ ;  $p_c$  indicates color polarity (setting the value of  $p_c$  to 1 or  $-1$  creates red and green tokens, respectively) and  $G(x, y)$  is a Gaussian modulation:

$$G(x, y) = \exp\left(-0.5 \frac{x^2 + y^2}{\sigma^2}\right),$$

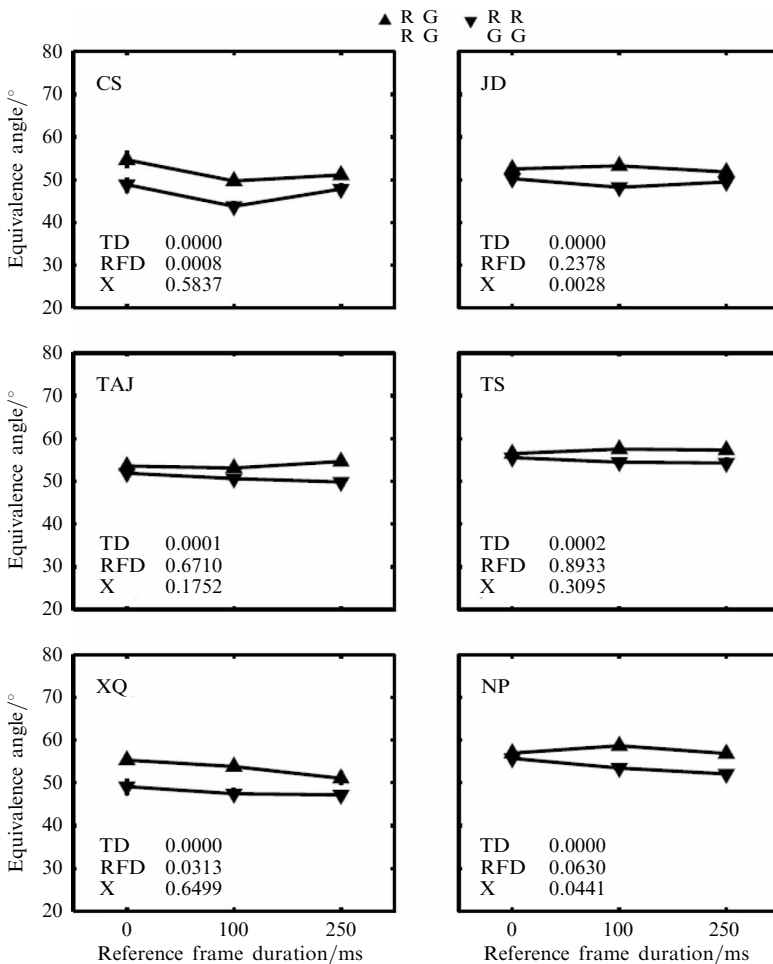
where  $x$  and  $y$  are spatial coordinates relative to the center of the motion token and  $\sigma = 0.4 \text{ deg}$ . In the background,  $L_r = L_{rm}$  and  $L_g = L_{gm}$ , such that the motion tokens were subjectively isoluminant with the background.

As in experiment 1, angle  $\rho$  determining the position of the motion tokens was constrained to have a value between  $10^\circ$  and  $80^\circ$  and was varied with a resolution of  $0.25^\circ$ . When tokens overlapped spatially in the reference frame (ie at values of  $\rho$  close

to  $10^\circ$  and  $80^\circ$ ), their red and green luminance modulations were added individually. This occurred only at a few trials early in each block before the staircase started to converge on a threshold. At none of the observed threshold values for  $\rho$  did the tokens overlap.

#### 4.2 Results

Equivalence angles with their standard errors are presented in figure 6, for each of the observers individually. Also included in the figure are  $p$ -values associated with the reliability of the effects (see appendix). As in experiment 1, equivalence angles are almost exclusively greater than  $45^\circ$ . For four of the six observers (JD, TAJ, TS, and NP), the pattern of results is similar to that obtained in experiment 1. The equivalence angles for these observers reveal a preference to match tokens based on their color but only when the motion sequence was preceded by the reference frame. The bias, when present, was smaller than that obtained in experiment 1, however. These four observers displayed no effect of TD when the motion sequence was not preceded by the reference frame. The results of observers CS and XQ, on the other hand, indicate no



**Figure 6.** Results of experiment, in which tokens differed with regard to their color. Thresholds and standard errors for obtained equivalence angles are shown for each observer as a function of token distribution (TD; colors match horizontally or vertically, see legend) and reference frame duration (RFD; 0, 100, or 250 ms). Where standard error bars are not visible, they do not exceed symbols. Also shown are  $p$ -values, based on bootstrap simulations (see appendix) for the main effects of TD and RFD and for the interaction between TD and RFD (X).

interaction of TD and RFD. Rather, these observers displayed a preference to match tokens of the same color at all levels of RFD.

Notwithstanding the effort made here to make the motion tokens in the present experiment subjectively isoluminant relative to the background, it is possible that this process was successful only for observers CS and XQ but failed for observers JD, TAJ, TS, and NP. It is also possible that, owing to physiological factors, internal luminance artifacts were introduced, making the tokens accessible to an achromatic motion mechanism as suggested by Mullen et al (2003). If this is indeed the case, the results suggest that early assignment of correspondence requires the motion tokens to be accessible to a luminance mechanism. The results of CS and XQ suggest the possibility that, for these observers, the motion tokens were not accessible to a luminance mechanism, in which case correspondence could only be assigned by a higher-level, controlled process acting on representations of the tokens derived from a chromatic mechanism. In order to test this possibility, observers CS and XQ participated in a control experiment in which the motion tokens, while still being subjectively isoluminant relative to each other, were made to differ in luminance from the background, thereby making them accessible to a luminance mechanism. If early assignment of correspondence indeed requires motion tokens to be accessible to a luminance mechanism, and chromaticity of tokens exerts effect only through a slow, controlled process, observers CS and XQ should now display the same patterns of results that they displayed in experiment 1. In the control experiment, the red motion tokens were defined by:

$$L_r = L_{rm} [1 + G(x, y)],$$

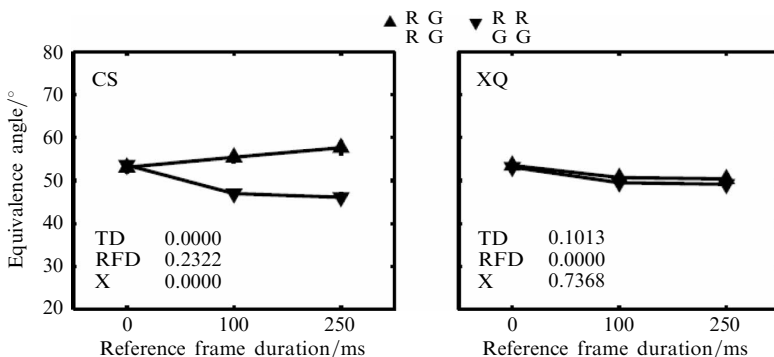
$$L_g = L_{gm},$$

while the green motion tokens were defined by:

$$L_r = L_{rm},$$

$$L_g = L_{gm} [1 + G(x, y)],$$

where the symbols are defined as above. In words, tokens were defined by a luminance increase in one of the red and green phosphors while the luminance in the other phosphor remained constant. In all other respects, the control experiment was identical to that described above. Results are presented in figure 7. Observer CS now displays the same pattern of results as observers JD, TAJ, TS, and NP (and the same pattern of results this observer displayed in experiment 1), while observer XQ now displays a lack of effect of TD for any level of RFD. Contrasted with the results in the main experiment for these two observers, this suggests that early assignment of correspondence, which disregards token similarity, is mediated by an achromatic (luminance) mechanism.

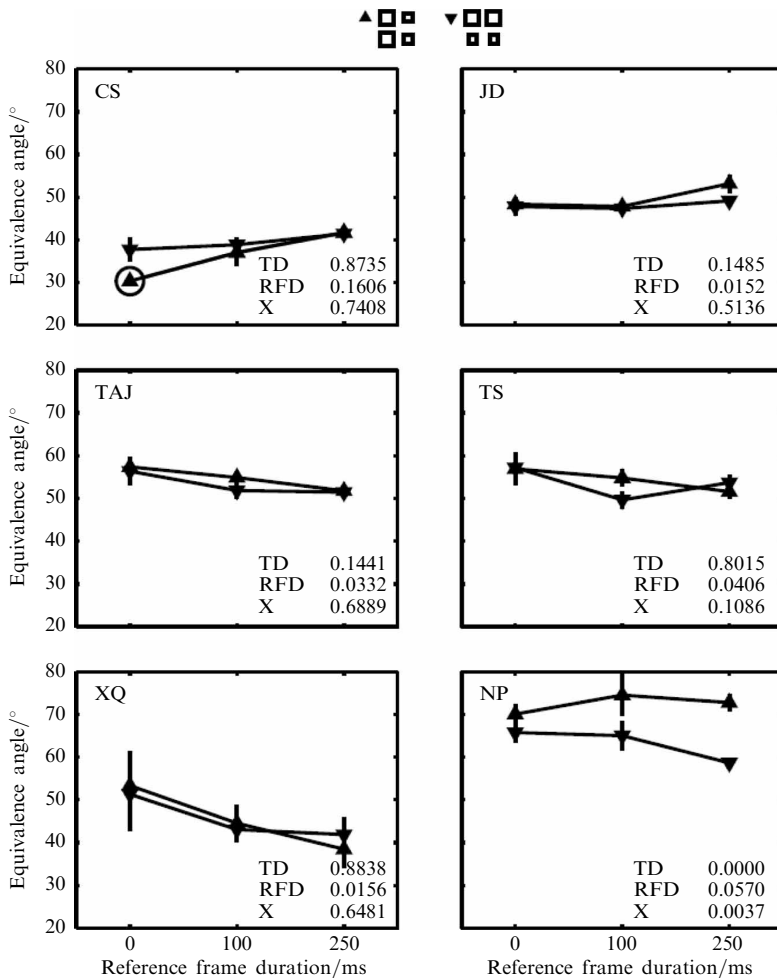


**Figure 7.** Results for observers CS and XQ in a control experiment, in which tokens were made to differ in luminance from the background.

## 5 Experiment 3: Depth

### 5.1 Stimuli

Motion tokens consisted of square random-dot patches ( $0.9 \text{ deg} \times 0.9 \text{ deg}$ , dot size  $1.7 \text{ min of arc} \times 1.7 \text{ min of arc}$ ) presented against a random-dot background ( $5.6 \text{ deg} \times 5.6 \text{ deg}$ ). Motion tokens were defined by (crossed) binocular disparity (6.6 or 13.3 min of arc relative to the background) such that all tokens were perceived to be floating in front of the background. The random-dot patterns (both background and tokens) were refreshed every frame, such that the display was truly cyclopean. Left-eye and right-eye views were presented in alternate frames. Participants wore ferro-electric liquid crystal goggles (Cambridge Research Systems) synchronized with the display to obtain fusion. In order to avoid spatial overlap between tokens in the reference frame, angle  $\rho$  (figure 2) was constrained to have a value between  $17.5^\circ$  and  $72.5^\circ$  with a resolution of  $2.5^\circ$ .



**Figure 8.** Results of experiment 3, in which tokens differed with regard to binocular disparity. Thresholds and standard errors for obtained equivalence angles are shown for each observer as a function of token distribution (TD; disparities match horizontally or vertically, see legend) and reference frame duration (RFD; 0, 100, or 250 ms). Where standard error bars are not visible, they do not exceed symbols, except for the condition indicated by the circled symbols, for which the standard error could not be determined. This was due to a very shallow slope of the psychometric function in this condition, leading to a failure of convergence of the fitting procedure for many of the bootstrap simulations. Also shown are  $p$ -values, based on bootstrap simulations (see appendix) for the main effects of TD and RFD and for the interaction between TD and RFD (X).

---

Since no subpixel resolution was used in experiment 3, the actual angles differed somewhat from those given above. For each value of  $\rho$ , constrained as indicated above, the ideal token positions were determined and tokens were placed at the closest possible locations given the limited (1 pixel, 1.7 min of arc) spatial resolution. The resulting angles  $\rho$ , based on the actual positions of the tokens, were used in the data-fitting procedure.

## 5.2 Results

Equivalence angles with their standard errors are presented in figure 8, for each of the observers individually. Also included in the figure are  $p$ -values associated with the reliability of the effects (see appendix). TD had little, if any, effect on equivalence angles for any of the naive observers. Any of the small differences in equivalence angles that were observed between the two different token distributions were inconsistent between observers and between experiments. Only the pattern of results of author NP mimics that found in experiments 1 and 2.

It is not clear why no consistent effect of relative depth was obtained for the naive observers. A reliable effect of relative token depth has been demonstrated previously by Green and Odom (1986), He and Nakayama (1994), and Prins and Juola (2001). However, without exception, the reported effect of relative depth was small. When the effects of retinal position of tokens and relative depth of tokens are expressed in the same metric, namely the object-centered Euclidian distances between tokens, the 3-D to 2-D distance trade-off is roughly an order of magnitude. In any case, it is not clear why (except for observer NP) no effect of relative token depth was obtained here.

## 6 General discussion

Two main findings were obtained in the experiments described here. First, the familiar biasing effect of token featural similarity on the correspondence process in long-range apparent motion was replicated, at least for the spatial-frequency content of the tokens and color of tokens. That is, all else equal, tokens are preferentially matched with like tokens. More importantly, however, it is shown that this biasing effect only develops over time and lags behind the effect of token proximity in the retinal projection. When the featural properties of the potential token matches are revealed only at the time the second stimulus frame appears (ie RFD = 0 ms), correspondence is assigned purely on the basis of the spatial coordinates of the motion tokens in the retinal projection, without any regard to the featural properties of the tokens. From inspection of the results, the biasing effect showed no consistent differences between RFDs of 100 and 250 ms and apparently is completed after at most 400 ms (the actual time between the appearance of the reference frame and the appearance of the second stimulus frame).

The present results confirm that the effect of featural properties of tokens is small relative to the effects of the relative distances in the retinal projection between tokens. Studies that have demonstrated effects of featural properties of tokens have typically, if not exclusively, done so with displays in which the position of motion tokens was such that either solution to the correspondence problem was generated with approximately equal probabilities. When the retinal coordinates are not equated in this way, no measurable effect may be obtained (Prins and Juola 2001; experiment 1 and experiment 2 have demonstrated this for the effect of the relative depth of motion tokens), presumably because the effect of token features is overshadowed by the effect of the relative proximity between tokens.

Overall, these results are consistent with the idea that long-range motion is mediated by a mechanism which acts initially on the spatiotemporal distribution of motion tokens only. The results of experiment 2, and in particular the control experiment performed there, tentatively suggest that this mechanism is luminance-based. That is, motion tokens apparently need to be accessible to a luminance mechanism in order for

---

this mechanism to assign correspondence without reference to the featural properties of the tokens. Any featural similarities among motion tokens are disregarded by this mechanism, even when these differences lie in the luminance domain as in experiment 1 (in which motion tokens differed with regard to their luminance spatial frequency content). Since the effects of featural differences lag behind the effect of the spatial distribution of motion tokens in the retinal projection, these effects appear to be mediated by a (slow) top-down process.

The nature of the top-down process remains largely a matter of speculation. However, it is quite likely that the top-down process is mediated by attention. It has been shown previously that attention affects the correspondence process in apparent motion. For example, Sperling and Lu (1995) instructed observers to attend to one of two featural dimensions along which tokens varied in their experiments and found that the perceived direction of motion was affected by which of the two features was attended. Suzuki and Peterson (2000) found that the direction of perceived motion can be influenced by sheer intention of the observer. These researchers merely instructed observers to 'try' to see motion one way or another and found that this instruction influenced which motion was perceived. It is conceivable that in the present experiment an expectation is generated by the observer based on a 'perceptually intelligent' (eg Sigman and Rock 1974) system regarding which solution to the correspondence problem should be perceived given the relative featural properties of the motion tokens. Through attentional processes the participant may then be able to bias the correspondence process accordingly. Attentional processes are, of course, slower than pre-attentive processes (eg Treisman and Gelade 1980), which explains why here the effect of featural properties lags behind the effect of the retinal coordinates of the motion tokens.

Presumably, given the high ecological value of motion in the visual image, it would be beneficial for an animal's visual system to detect motion (ie assign correspondence) as early as possible. Evolutionary pressure must have strongly favored those animals that were able to detect their prey's or predator's motion before being detected themselves. The present results support this notion: initial correspondence is assigned based solely on information that is available early, namely the retinal coordinates of the motion tokens. Speculatively, the slow top-down effects of properties other than the retinal coordinates of the tokens may serve several functions. First, once motion is detected, it would be beneficial to generate that solution to the correspondence problem which corresponds to the most ecologically valid interpretation of the motion. Second, it is possible that, in the absence of luminance-defined motion, motion detection relies entirely on a higher-level interpretative mechanism. This idea is tentatively supported by the results of observers CS and XQ in experiment 2, where tokens were defined by color. The results of observers CS and XQ in this experiment were the only results reported in this paper to show a reliable effect of token distribution when the reference frame was not presented. This effect disappeared for both observers when the tokens were made to be luminance-defined in a control experiment. Finally, it is conceivable that under certain circumstances top-down control might speed up the detection of motion. Williams et al (1986) proposed a model of apparent-motion perception in which the activity of directionally sensitive neurons is updated iteratively on the basis of the excitatory and inhibitory interactions among them. Neurons selective for like directions will mutually excite each other, while the interactions among neurons selective for different directions are mutually inhibitory. Such a mechanism will, over time, converge on a solution to the motion correspondence problem and do so more quickly when the activity of neurons selective for one direction of motion is much stronger from the outset compared to the activity of neurons selective for another direction. However, when the activities of neurons selective for the two possible directions of motion are approximately equal at the outset, a stable state will be reached more slowly.

Hock et al (2003) devised a model similar to that of Williams et al and used it to simulate the dynamical phenomena of bistability, spontaneous switching, and hysteresis. Convergence towards a stable solution in the model of Hock et al will again only be developed over time, especially when, based on the retinal coordinates of the tokens alone, no strong preference for a particular solution to the correspondence problem exists. An attentional top-down process may serve to tip the balance such as to lead to a faster convergence of the motion mechanism towards a stable state.

## References

- Adelson E H, Bergen J R, 1985 "Spatiotemporal energy models for the perception of motion" *Journal of the Optical Society of America A* **2** 284–298
- Baddick O, 1973 "The masking of apparent motion in random-dot patterns" *Vision Research* **13** 355–369
- Braddick O, 1974 "A short-range process in apparent motion" *Vision Research* **14** 519–527
- Burt P, Sperling G, 1981 "Time, distance, and feature-trade-offs in visual apparent motion" *Psychological Review* **88** 171–195
- Caelli T, Manning M, Finlay D, 1993 "A general correspondence approach to apparent motion" *Perception* **22** 185–192
- Cavanagh P, Mather G, 1989 "Motion: The long and short of it" *Spatial Vision* **4** 103–129
- Chaudhuri A, Glaser D A, 1991 "Metastable motion anisotropy" *Visual Neuroscience* **7** 397–407
- Dawson M, 1991 "The how and why of what went where in apparent motion: modeling solutions to the motion correspondence problem" *Psychological Review* **98** 569–603
- Efron B, Tibshirani R J, 1993 *An Introduction to the Bootstrap* (Boca Raton, FL: Chapman & Hall/CRC)
- Gengerelli J A, 1948 "Apparent movement in relation to homonymous and heteronymous stimulation of the cerebral hemispheres" *Journal of Experimental Psychology* **38** 592–599
- Green M, 1986 "What determines correspondence strength in apparent motion?" *Vision Research* **26** 599–607
- Green M, Odom J V, 1986 "Correspondence matching in apparent motion: Evidence for three-dimensional spatial representation" *Science* **233** 1427–1429
- He Z J, Nakayama K, 1994 "Apparent motion determined by surface layout not by disparity or three-dimensional distance" *Nature* **367** 173–175
- Hock H S, Schöner G, Giese M, 2003 "The dynamical foundations of motion pattern formation: stability, selective adaptation and perceptual continuity" *Perception & Psychophysics* **65** 429–457
- Judd C M, McClelland G H, 1989 *Data Analysis: A Model Comparison Approach* (San Diego, CA: Harcourt Brace Jovanovich)
- Kolers P A, Green M, 1984 "Color logic of apparent motion" *Perception* **13** 249–254
- Kolers P A, Pomerantz J R, 1971 "Figural change in apparent motion" *Journal of Experimental Psychology* **87** 99–108
- Kolers P A, Grünau M von, 1976 "Shape and color in apparent motion" *Vision Research* **16** 329–335
- Koriat A, 1994 "Object-based apparent motion" *Perception & Psychophysics* **56** 392–404
- Lu Z-L, Sperling G, 1995 "The functional architecture of human visual motion perception" *Vision Research* **35** 2697–2722
- Mack A, Klein L, Hill J, Palumbo D, 1989 "Apparent motion: evidence of the influence of shape, slant, and size on the correspondence process" *Perception & Psychophysics* **46** 201–206
- Mather G, 1988 "Temporal properties of apparent motion in subjective figures" *Perception* **17** 729–736
- Mullen K T, Yoshizawa T, Baker C L, 2003 "Luminance mechanisms mediate the motion of red-green isoluminant gratings: the role of temporal chromatic aberration" *Vision Research* **43** 1235–1247
- Navon D, 1976 "Irrelevance of figural identity for resolving ambiguities in apparent motion" *Journal of Experimental Psychology: Human Perception and Performance* **2** 130–138
- Oyama T, Simizu M, Tozawa J, 1999 "Effects of similarity on apparent motion and perceptual grouping" *Perception* **28** 739–748
- Pentland A, 1980 "Maximum likelihood estimation: the best PEST" *Perception & Psychophysics* **28** 377–379
- Prins N, Juola J F, 2001 "Relative roles of 3-D and 2-D coordinate systems in solving the correspondence problem in apparent motion" *Vision Research* **41** 759–769

- Reichardt W, 1969 "Movement perception in insects", in *Processing of Optical Data by Organisms and Machines* Ed. W Reichardt (New York: Academic Press) pp 465–493
- Schneider W, Shiffrin R M, 1977 "Controlled and automatic human information processing: 1. Detection, search, and attention" *Psychological Review* **84** 1–66
- Scott-Samuel N E, Georgeson M A, 1999 "Feature matching and segmentation in motion perception" *Proceedings of the Royal Society of London, Series B* **266** 2289–2294
- Shechter S, Hochstein S, 1989 "Size, flux, and luminance effects in the apparent motion correspondence process" *Vision Research* **29** 579–591
- Shechter S, Hochstein S, 1990 "On and off pathway contributions to apparent motion perception" *Vision Research* **30** 1189–1204
- Shechter S, Hochstein S, Hillman P, 1988 "Shape similarity and distance disparity as apparent motion correspondence cues" *Vision Research* **28** 1013–1021
- Sigman E, Rock I, 1974 "Stroboscopic movement based on perceptual intelligence" *Perception* **3** 9–28
- Sperling G, Lu Z-L, 1995 "Attention affects the perceived direction of visual motion" *Investigative Ophthalmology & Visual Science, Supplement* **36** 854
- Suzuki S, Peterson M A, 2000 "Multiplicative effects of intention on the perception of bistable apparent motion" *Psychological Science* **11** 202–209
- Treisman A M, Gelade G, 1980 "A feature-integration theory of attention" *Cognitive Psychology* **12** 97–136
- Ullman S, 1979 *The Interpretation of Visual Motion* (Cambridge, MA: MIT Press)
- Ullman S, 1980 "The effect of similarity between line segments on the correspondence strength in apparent motion" *Perception* **9** 617–626
- Wertheimer M, 1912 "Experimentelle Studien über das Sehen von Bewegung" *Zeitschrift für Psychologie und Physiologie der Sinnesorgane* **61** 161–265
- Williams D, Phillips G, Sekuler R, 1986 "Hysteresis in the perception of motion direction as evidence for neural cooperativity" *Nature* **324** 253–255
- Zhuo Y, Zhou T G, Rao H Y, Wang J J, Meng M, Chen M, Zhou C, Chen L, 2003 "Contributions of the visual ventral pathway to long-range apparent motion" *Science* **299** 417–420

## Appendix

### Model comparisons

In order to test for the effects of token distribution (TD), reference frame duration (RFD), and their interaction (X), a variant of the likelihood-ratio test was used. The likelihood-ratio test can be used to compare the fit of two models when the parameter space of one model (referred to here as the constrained model) is a subset of the parameter space of the full model. In order to test for the statistical significance of the effects, the six parameters corresponding to the thresholds were reparametrized. Before reparametrization, the full (augmented) model contains seven parameters: the six individual thresholds ( $\alpha_{\text{TD,RFD}}$ , where  $\text{TD} = \{\text{H}, \text{V}\}$  and  $\text{RFD} = \{0, 100, 250\}$ ) and a common slope parameter ( $\beta$ ). The threshold parameters were reparametrized by using contrasts. Contrasts are routinely used in the context of the general linear model to test for main effects and interactions of the independent variables (eg Judd and McClelland 1989). The following matrix consists of six rows, each containing a contrast. The six contrasts are orthogonal. Let us symbolize the reparametrized parameters determining thresholds by  $\theta$ .  $\theta$  and  $\alpha_{\text{TD,RFD}}$  are related thus:

$$\begin{aligned}
 & [\theta_{\text{M}} \quad \theta_{\text{TD}} \quad \theta_{\text{RFD1}} \quad \theta_{\text{RFD2}} \quad \theta_{\text{X1}} \quad \theta_{\text{X2}}] \times \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & -1 & -1 & -1 \\ 2 & -1 & -1 & 2 & -1 & -1 \\ 0 & 1 & -1 & 0 & 1 & -1 \\ 2 & -1 & -1 & -2 & 1 & 1 \\ 0 & 1 & -1 & 0 & -1 & 1 \end{bmatrix} \\
 & = [\alpha_{\text{H},0} \quad \alpha_{\text{H},100} \quad \alpha_{\text{H},250} \quad \alpha_{\text{V},0} \quad \alpha_{\text{V},100} \quad \alpha_{\text{V},250}].
 \end{aligned}$$

After reparametrization, the full model thus again contains seven parameters. Six of these ( $\theta$ ) determine the thresholds for the six conditions after matrix multiplication with the contrast matrix.  $\theta_M$  corresponds to the mean threshold across all conditions,  $\theta_{TD}$  codes for a main effect of TD,  $\theta_{RFD1}$  and  $\theta_{RFD2}$  together code (exhaustively) for a main effect of RFD,  $\theta_{X1}$  and  $\theta_{X2}$  code (exhaustively) for the TD  $\times$  RFD interaction. The remaining parameter directly corresponds to the common slope of the fitted logistic functions. Having reparametrized the parameters determining threshold values in this manner, we can construct appropriate constrained models to test for the main effects and the interaction of the independent variables. The parameters for all models were estimated by using a maximum likelihood criterion. The full model for all these comparisons is that in which all  $\theta$  (and thus all thresholds,  $\alpha$ ) and the common slope ( $\beta$ ) are free to vary. In the test for a main effect of TD, the null hypothesis states that parameter  $\theta_{TD}$  equals 0. Thus, the constrained model will be identical to the full model except that parameter  $\theta_{TD}$  is fixed at 0. In terms of threshold values, this amounts to constraining  $\alpha_{TD,RFD}$  such that the average of  $\alpha_{TD=H}$  equals the average of  $\alpha_{TD=V}$ . Similarly, in order to test for a main effect of RFD,  $\theta_{RFD1}$  and  $\theta_{RFD2}$  should be fixed at 0 in the constrained model. In order to test for an interaction of TD and RFD,  $\theta_{X1}$  and  $\theta_{X2}$  should be fixed at 0 in the constrained model.

Having defined the full and constrained models as above we can generate maximum-likelihood estimates of our parameters under these models and determine the maximum likelihoods under each of the models. If the maximum likelihoods of the data are assessed under both the full and the constrained model [let us call these  $\max(L_{full})$  and  $\max(L_{constr})$ , respectively], we can compute the likelihood ratio  $\lambda$ :

$$\lambda = \frac{\max(L_{constr})}{\max(L_{full})}.$$

When the null hypothesis is true (ie the parameters that are fixed in the constrained model are, in fact, equal to 0),  $-2 \log \lambda$  is asymptotically distributed as  $\chi^2$ , with degrees of freedom equal to the difference in the number of free parameters between the two models.

However, since the data were gathered by means of a staircase procedure, the number of observations per stimulus level was typically very low (often only one observation per stimulus level), leading to an unacceptably poor correspondence between the actual sampling distribution and the theoretical  $\chi^2$  distribution. For that reason, empirical sampling distributions of  $-2 \log \lambda$  were created by simulating the experiment. For each combination of observer, experiment, and hypothesis to be tested, the best fitting values of the  $\theta_i$  and  $\beta$  for the constrained model were determined. These values were then used in simulations of the experiment. In the simulations the exact same values of the independent variables (ie values of  $\rho$ ) that the human observer actually tested at were used (these values differ between experiments and observers since they were determined by the staircase procedures at the time of testing). For each simulated experiment, the value of  $-2 \log \lambda$  was determined. The collection of these values served as the empirical sampling distribution of  $-2 \log \lambda$ . All empirical sampling distributions were based on  $B = 10\,000$  simulations. Reported in the result figures are, for each of the effects, the proportions of simulations resulting in a value of  $-2 \log \lambda$  greater than that observed from the human observer. These values are estimates of the probability of obtaining a value of  $-2 \log \lambda$  as high as, or higher than, that obtained from the human observer in case the effect parameter has a value of 0.

**Conditions of use.** This article may be downloaded from the E&P website for personal research by members of subscribing organisations. This PDF may not be placed on any website (or other online distribution system) without permission of the publisher.