Multiple Mechanisms of Apparent Motion Perception

Thesis by

Ching Elizabeth Ho

Computation and Neural Systems

California Institute of Technology

Pasadena, CA 91125

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

2000

Acknowledgement

Life is a dance. We do not ask why we are put there. We are just there to put in ourselves and perform our best. I am grateful for everything life has given me during these five years at Caltech, for all the things I learned, all the knowledge I gained, for all the happiness, as well as all the pains I experienced, and for all the wonderful people I came to know.

I would like to give my special thanks to my advisors, Professor George Sperling and Professor Shinsuke Shimojo. George has been a great teacher. He has encouraged me to develope my various skills and potentials, and his coaching has helped shape this thesis. Shin, with his wide range of knowledge and expertise in psychophyiscs, broadened my perspective and deepened my insight. It is truly fortunate for me to interact with and learn from them. I thank Professor Christof Koch for his advice and hospitality, and the writing of the paper on the first part of this thesis, which was conducted in his lab. I thank Prof. John Allman and Prof. Richard Andersen. John has been a great inspiration from the very beginning. Richard's kindness and generosity has helped me in various different situations. I benefitted a lot as a fellow of Sloan Center for Theoretical Neurobiology at Caltech, which was wonderfully organized and managed by Richard.

I owe many of my happy days in Caltech to the wonderful collegeaus and friends, Martin Stemmler, Wyeth Bair, Gary Holt, Laurent Itti, Amit Manwani, Yuki Kamitani, Katz Watanabe, Jennifer Sun, Christine Chee, Raymond Chan, Joseph Umbro, and Robert

Peters. I thank our effecient and helpful secretary, Johanna Weber.

It is my great pleasure to come to know Professors Francis Crick, Kam Lau, Partha Mitra, Colin Blakemore, Tai Sing Lee, Hoi-Fung Chau and Tai-Kai Ng. They have given me much inspirations over the years. Their friendship and warmth have sustained my interest in science and given me much happiness in life. Collaborations with Partha and with Hoi-Fung were wonderful learning experiences for me.

Thanks to my music teachers, Margaret Lambert and Dr. Joseph Huszti. They not only teach me music, but also open my eyes to see many beautiful things in life.

Most importantly, I'd like to thank my parents. Without their understanding and encouragement throughout these years, this thesis would not be possible.

Abstract

Motion perception is a complex phenomenon. Recently, multiple categories of motion perception have been defined through properties of the stimulus: e.g., short-range and long-range motion(Braddick, 1980), or first-order (luminance-defined), second-order (texture-defined) (Chubb & Sperling, 1988; Chubb & Sperling, 1989; Cavanagh & Mather, 1989), and third-order (pattern-tracking) (Lu & Sperling, 1995a; Lu & Sperling, 1995b) motion. This thesis elucidates the mechanisms of motion perception for a class of ambiguous motion stimuli. In particular, two competing motion systems were found to be involved in the perception of nominal second-order motion stimuli. These systems were hypothesized to explain the inter-observer differences in the perceived direction of motion, and to explain the differences in performance under a dual task paradigm. The interference effects seen in the dual task performances implied the involvement of multiple forms of/distributed attention. Choice of systems could be influenced by attentional instructions and training in addition to stimulus conditions.

In viewing an ambiguous texture-defined motion stimulus first devised by Werkhoven *et al* (1993), the observers fell into two distinct groups based on the direction of perceived motion. The differences were interpreted in terms of the algorithms used to extract motion: one group using a second-order motion process, the other, a third-order motion process.

This was investigated further using a dual-task paradigm in which the interference between two tasks indicated the nature of processing involved. Observers who used third-order motion processing experienced interference with a letter-recognition task, and a more severe interference in dual third-order motion tasks. Observers who used second-order motion processing experienced interference with another second-order motion detection task, but not with the letter-recognition task.

These observations suggest that two systems, a second-order system and a third-order system, are involved in the perception of the nominal second-order motion stimuli. The performance of observers can be interpreted in terms of a simple architecture of motion processing and attentional resource. Insofar as task interference implies competition for attentional resource, the complex and apparently paradoxical interference effects of second-order motion perception suggest that there are multiple attentional resources, or in another word, attention is distributed. Whether two tasks interfere or not depends on whether they require the same attentional resource. A quantitative method, Structure-Attention-Mapping (SAM) was formulated, and a model architecture for the motion pathways was proposed using this method. It was found to be able to explain the data in the experiments.

While different observers have different innate tendencies to use one pathway over the other, the selection of pathway is not fixed but can be induced by attentional instructions, training, temporal frequency, and viewing conditions. High temporal frequencies and monocular displays favored the second-order system, low temporal frequencies and interocular displays favor the third-order system. In addition, stimuli composed of patches with

orthogonal slant patterns (one slant could be selectively attended) favored the third-order direction; same-slant stimuli favored the second-order direction. A study of luminance-defined motion showed that it can also be perceived by a third-order motion system as well as a first-order motion system. The competition between first- and third-order systems qualitatively resembles that between second- and third-order motion systems.

Contents

List of Figures

Chapter 1 Introduction

1.1 Visual Perception of Apparent Motion

Visual motion perception has long been a central theme of study in psychophysics. It has been evident, from very early, that the complex phenomena of motion perception must involve more than one kind of computations (Braddick, 1974; Braddick, 1980; Van Santen & Sperling, 1984; Adelson & Bergen, 1985; Cavanagh, 1992; Wilson *et al.*, 1992; Boulton & Baker, 1993). One of the better known recent dichotomies is Braddick's long range and short range motion systems proposed in 1974. The defining characteristic of these two motion processes is their spatio-temporal range, the distance and the duration. Braddick (1980) proposed a set of criteria for the stimuli that activate short range and long range processes. The short-range process is intensity-based and supposed to correspond to a low-level mechanism, while the long-range process is token-based, and corresponds to a high-level, cognitive mechanism. These categories provide certain insight into the fact that more than one mechanism is necessary to explain our motion perception. However, subsequent research (Mather *et al.*, 1985; Cavanagh & Mather, 1989, and others) showed that short-range and long-range were only consequence of the stimuli used in the two paradigms. They were only incidental properties of those motion mechanisms behind them. The short-range, long-range distinction does not reveal the mechanisms themselves. Later in this thesis, it will be shown that different mechanisms can be used to perceive motion from the same stimuli.

1.1.1 First-Order Motion

The mechanism of the first-order (or Fourier) motion system was established in the mid eighties. The first-order motion system detects the spatio-temporal variation of luminance. What we now call first-order motion was for a time called Fourier motion because the class of equivalent detectors that have been proposed for first-order motion detect the components corresponding to motion energy in an Fourier analysis of stimulus luminance. Reichardt proposed what is now called the Reichardt model as a theory for beetle vision(Reichardt, 1957). It was then adapted to human vision (Van Santen & Sperling, 1984; Van Santen & Sperling, 1985) and a mathematically equivalent models, the motion energy model (Adelson & Bergen, 1985), was proposed. Various other theories of first-order motion, such as a gradient model, have been shown to be equivalent or nearly equivalent to the Reichardt model, and this class is sometimes called Standard Motion Analysis. (Note: These models are only equivalent in their overall system functions; the neural implementation of first-order motion is not yet fully understood.) Fig.1.1 shows a sinewave grating and a half-Reichardt model. The full model has a mirror symmetric component that detects motion in the opposite direction and the full model outputs the difference between the outputs of the two halves.

Van Santen & Sperling (1984) and Wilson (Wilson, 1994) gave the following example for the response (R) of a Reichardt detector to a cosine grating moving at velocity v may be shown to be:

$$R = M^2(\omega)\sin(\omega v \Delta t)\sin(\omega \Delta x)$$

where ω is the spatial frequency of the grating, and Δx and Δt are the spatial offset and time lag associated with the the Reichardt detector, respectively. $M(\omega)$ is the response of a receptive field or filter, such as that of a simple cell in primary visual cortex, to the grating as a function of its spatial frequency.

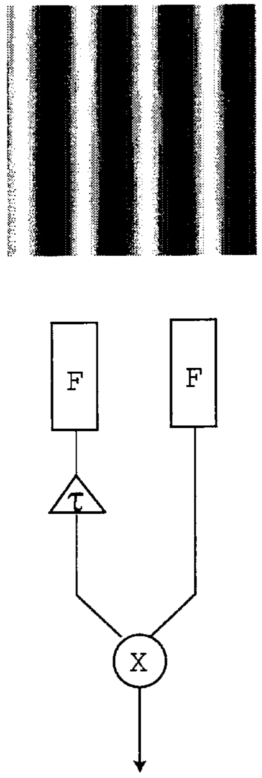


Figure 1.1: first-order example and half-Reichardt detector. Reichardt model is one of the standard model for first-order (luminance-defined) motion. It consists of two spatial filters separated at a distance $\Delta\vec{x}$ and a time delay (temporal filter), Δt, feeding into a cross-correlator. A luminance bar moving at velocity $\frac{\Delta\vec{x}}{\Delta t}$ will produce a strong signal. The motion energy model (Adelson & Bergen, 1985) is another model for first-order motion which is mathematically equivalent to the full-Reichardt model (not illustrated here).

1.1.2 Second-Order Motion

Once a theory of first-order motion had been established and supported by a number of strongly corroborative experimental results, it didn't take psychophysicists long to invent stimuli whose motion could not be detected by Standard Motion Analysis. Indeed, some of these stimuli had been around since the 1970s.

The example in fig.1.2 is a moving sine wave modulator multiplied to a static black-and-white noise carrier. Chubb and Sperling (1988; 1989) proved that this kind of stimulus is drift balanced, that is, the expectation of the amplitude of any spatial and temporal frequency component that represents leftward motion is equal to the expectation of the amplitude of the same frequency that represents rightward motion.

Apparent motion in these drift balanced stimuli is just as vivid and powerful as motion in ordinary moving stimuli. Observers do no know whether apparent motion is being produced by first- or second-order computations. Nevertheless, Standard Motion Detectors cannot compute the motion direction of drift balanced stimuli.

Chubb and Sperling (1988; 1989) proposed that motion of drift-balanced stimuli is revealed by preprocessing of the visual stimuli by a "texture grabber" which is then followed by Standard Motion Analysis. The texture grabber consists of a spatio-temporal filter followed by a fullwave rectifier. Texture grabbers report the quantity of texture at each neighborhood of the visual field. Second-order motion is computed on the quantity of texture (e.g. contrast, spatial frequency, flicker, and etc) in the same way that first-order motion is computed on the quantity of light.

Using a very sensitive test, the phase test, Lu and Sperling (1995b) showed that first- and second-order motion are analyzed in different systems, whose outputs are then combined to produce apparent motion.

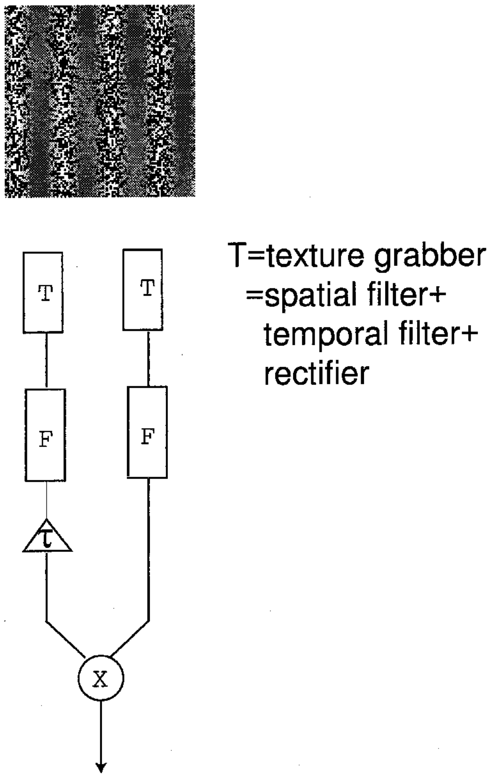


Figure 1.2: Example of a second-order motion stimuli and a model of second-order motion detector by Werkhoven *et al* (1993).

1.1.3 Third-Order motion

Again, once explicit mechanisms for first- and second-order motion had been defined, psychophysicists quickly invented stimuli that produced good apparent motion and which were invisible to the first- and second-order mechanisms.

An example of such a stimulus is a motion-defined motion grating as shown in fig.1.3(a). The stimulus consists of the random dots. The arrows indicate the direction of motion of random dots. The probability waveform governs whether a dot at a particular horizontal location moves up or down. When this probability waveform moves horizontally, one can perceive the motion. However, this motion direction cannot be computed by either first- or second-order motion detectors. Another example is the dynamic random-dot stereo-depth grating as shown in fig.1.3(b). The depth amplitude is defined by the disparity of the corresponding dots in the two eyes. The grating moves in one direction. The depth grating does not have any consistent spatio-temporal variation in luminance or contrast so it can neither be detected by first-order nor second-order motion detectors.

For revealing mechanisms, the most significant of these new displays is an alternating feature display in which successive frames are composed of completely unrelated patterns, but have the property that each frame defines an area of figure, and the figure-area moves in a consistent direction from frame to frame. A map that records the location of areas of the visual field is called a salience map, following the terminology suggested by Koch and Ullman (Koch & Ullman, 1985). Lu and Sperling (1995a) proposed that a third-order motion system computes motion of the areas defined as figure in a salience field. They noted that all the third-order motion stimuli had the same temporal tuning function, a low pass filter with a corner frequency of 3-6 Hz. For comparison, the corner frequencies of first- and second-order motion are 10-12 Hz. Furthermore, the third-order stimuli are indifferent to the eye of origin; thresholds are the same whether all stimuli are presented to the same eye or when they alternate between eyes. In contrast, first- and second-order systems seem to be almost exclusively monocular, they fail with interocular stimuli. There are other differences: The first- and second-order motion systems exhibit pedestal immunity (observers perceive unimpaired motion in stimuli superimposed on a stationary pedestal) whereas the third-order system fails to detect linear motion and reports only the wobble component. Brain lesions affect first- and second-order systems differently; the effects of attention are different.

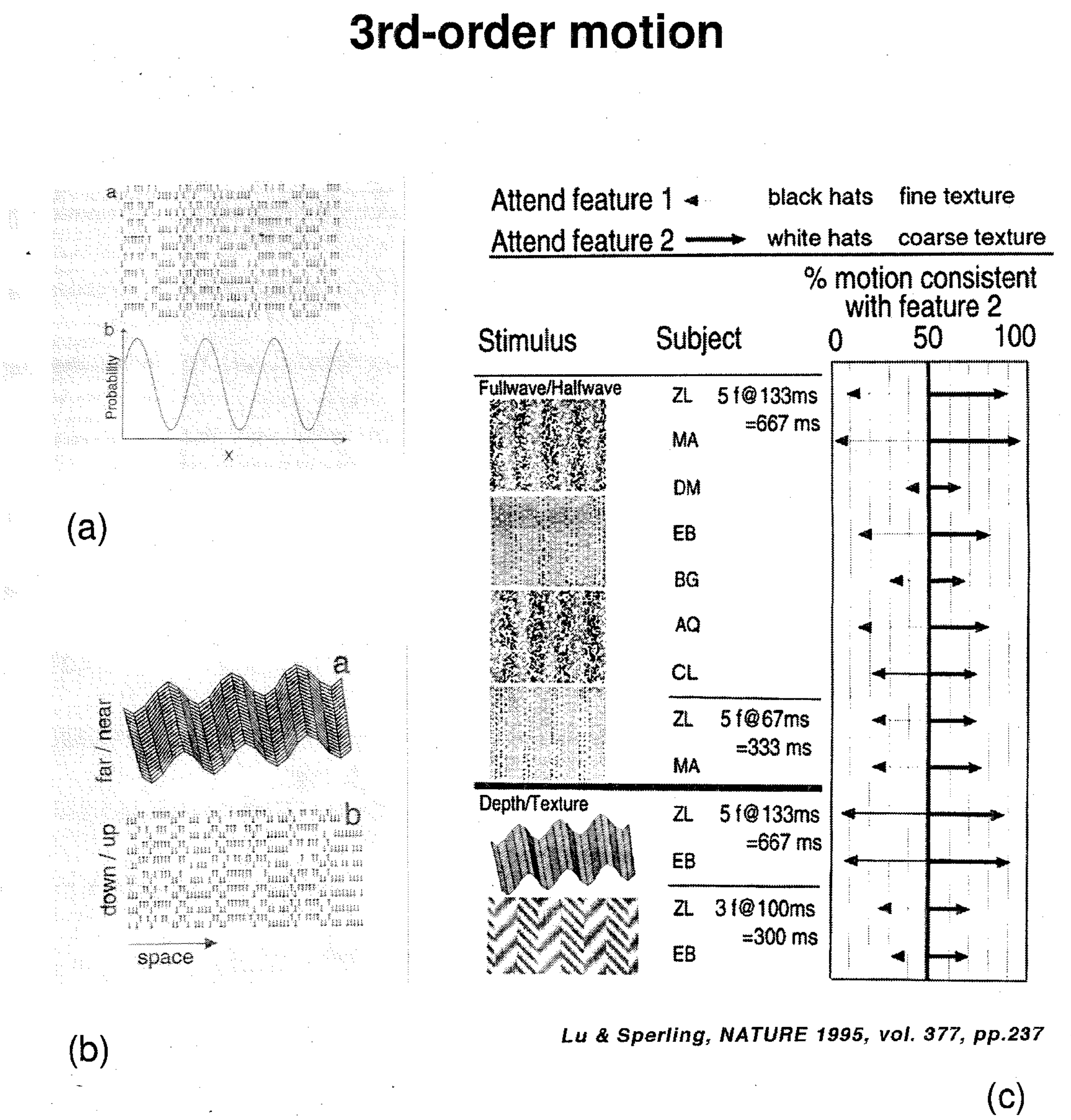


Figure 1.3: (a) Motion-defined motion (reproduced from Lu & Sperling (1995b)). (b)(Upper row) Dynamic random-dot stereo-depth grating (reproduced from Lu & Sperling (1995b)). (c)Alternating feature display.

1.2 Summary of this thesis

In chapter 2, I will describe how the observers fell into two distinct groups based on the direction of perceived motion in the context of an ambiguous motion task involving a nominal second-order stimulus. The differences were interpreted in terms of the algorithms used to extract motion: one group using a second-order motion process, the other, a third-order motion process. This will be further investigated, in chapter 3, by using a dual-task paradigm in which the interference between the two tasks indicated the nature of processing involved. Observers who used third-order motion processing experienced interference with letter-recognition, and a more severe interference in dual third-order motion tasks. Observers who used second-order motion processing experienced interference with another second-order motion detection, but not with letter-recognition. To explain these results, a quantitative approach, Structure-Attention Mapping, is introduced in chapter 4, to relate psychological Attention Operating Characteristic (AOC) graphs to structures in the brain. A model architecture for the attentional bottlenecks of motion processing was proposed to explain all the observed behaviors. These experimental data and a model suggested two pathways, a second- and a third-order motion pathways are involved involved in the motion perception of the texture-defined stimulus. In chapter 5, motivated by these results and those of Lu & Sperling (1995b), factors affecting observers' selection of motion pathways for motion perception were investigated. Attentional instructions, training, and stimulus parameters, such as temporal frequency, grating slant, and monocular vs interocular viewing conditions were all factors affecting observers' selection of motion pathways. Chapter 6 demonstrates that qualitatively similar behavior due to the involvement of the third-order motion can also be found in luminance-defined motion stimuli. Chapter 7 summarizes the factors for selection of pathways and discusses the results. The Appendix describes the black-white calibration procedure and some results obtained before calibration were presented, and discusses some issues in attention.

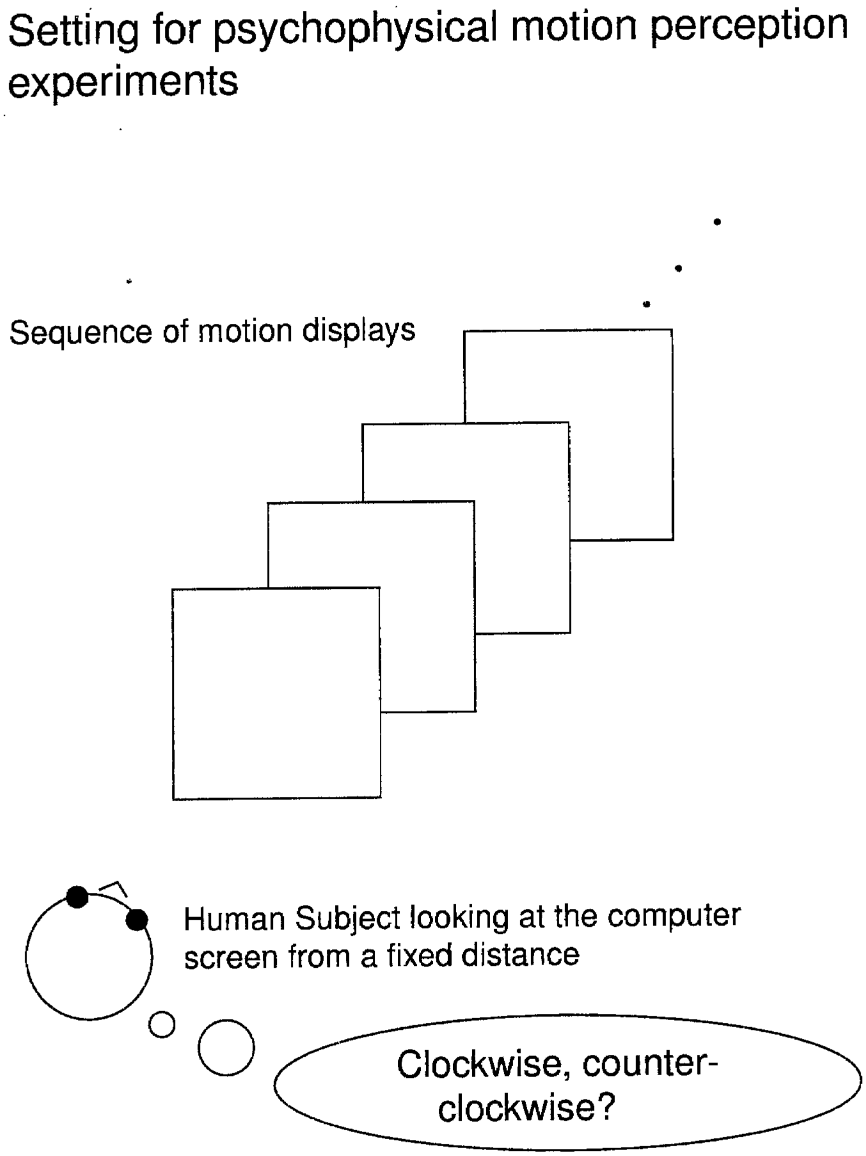


Figure 1.4: Experimental setting for the psychophysics experiments of motion perception and letter recognition described in this chapter. Subjects were seated in front of a computer screen with their eyes at a fixed distance from the screen. A series of frames were presented in sequence. At the end of one sequence, observers had to report their perceptual judgements.

Chapter 2 Divergence in Motion Perception

2.1 Introduction.

Imagine the car of the future on the six-lane freeways of Los Angeles: As we drive, a head-up-display of the road-map hovers in the near field of our vision. While we are paying attention to figure out the next exit shown on the road-map, will our brains still be able to do motion processing to avoid the car swerving into our lane in front of us?

Motion detection and shape recognition in primates are known to be processed in the dorsal and ventral streams, respectively (Ramachandran & Gregory, 1978; Ungerleider & Mishkin, 1982; Livingstone & Hubel, 1987; DeYoe & Van Essen, 1988; Zeki & Shipp, 1988; Zeki, 1990). Yet, these two streams are not entirely separate entities. Cross-talk is found between these two streams (Blasdel *et al.*, 1985; Rockland, 1985; Shipp & Zeki, 1989a; Shipp & Zeki, 1989b).

The study in this and the next two chapters began with the goal of investigating the role of attention in second-order motion perception by comparing motion performance in the presence of an attention-demanding recognition task with that in the absence of a recognition task. One such second-order motion task involves an ambiguous texture-defined motion stimulus (Fig.2.1 (a)), devised by Werkhoven, Sperling and Chubb (Werkhoven *et al.*, 1993).

The first experiment was to reproduce their earlier results. Surprisingly, observers fell into two distinct groups based on the direction of perceived motion: one group behaved similarly as Werkhoven *et al*'s observers, the other group behaved completely differently. The differences were interpreted in terms of the algorithms used to extract motion: one group using a second-order motion process which operates according to Werkhoven *et al*'s motion energy model; the other, a third-order motion process which tracks motion of a certain pattern.

2.2 Motion Stimulus

The stimulus consisted of ambiguous second-order motion displays in which apparent motion is carried by the textural properties. In these stimuli, identical displays can be perceived to rotate in either clockwise or counter-clockwise directions. A *heterogeneous* motion direction, defined by alternating sectors of texture s (standard) and texture v (variable), competes with a *homogeneous* motion direction, defined solely by sectors of texture s (Werkhoven *et al.*, 1993). Fig.2.1 shows a schematic representation of the motion stimuli. It was first devised by Werkhoven *et al* (1993). A series of eight frames $(f_1, f_2, \cdots)$ is shown in succession, each frame lasting 125 msec. Frame f_{n+2} is similar to frame f_n, but rotated clockwise by a 45° angle. Unfolding the annular stimulus, the horizontal axis represents the angular positions ϕ, while the vertical axis represents time. Two different textures (spatial frequencies) s and v are present in the odd frames, while the even frames only contain sectors with texture s. Phase is randomized in each frame, each sector. When frame f_n and frame f_{n+1} are presented in succession, two opposing interpretations of the motion direction are possible, as indicated by the arrows:

- Homogeneous motion pathway: matches between sectors of identical texture s. In this case, motion will be perceived in the counter-clockwise direction.
- Heterogeneous motion pathway (counterclockwise): matches between sectors of texture s and texture v. In this case, motion will be perceived in the clockwise direction.

Spatial frequencies for texture s is 1.26c/deg, and for texture v is 0.83c/deg. Outer radius is 3.00 deg. Inner radius is 1.50 deg. The motion path chosen is determined by the motion strength (Werkhoven *et al.*, 1993) as shown in fig.2.1 (b). Their result is shown in fig.2.1(c).

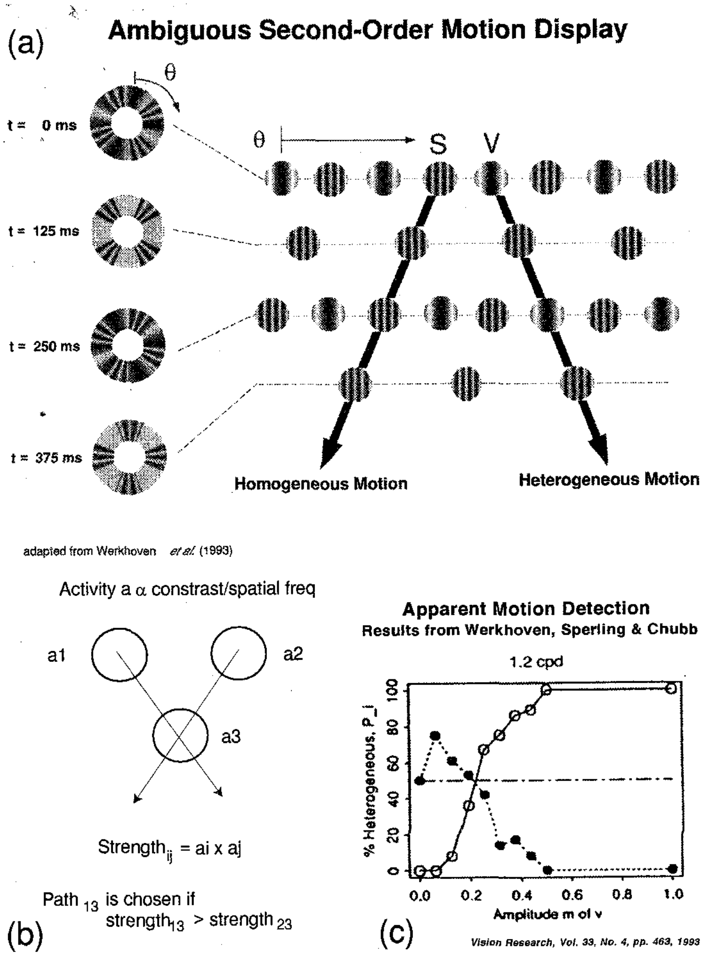


Figure 2.1: (a) Schematic explanation of ambiguous second-order motion displays, reproduced from Werkhoven *et al.* (1993). (Left) A series of eight frames $(f_1, f_2, \cdots)$ is shown in succession, each frame lasting 125 msec. Frame f_{n+2} is similar to frame f_n, but rotated clockwise by a 45° angle. (Right) Unfolding the annular stimulus, the horizontal axis represents the angular positions θ, while the vertical axis represents time. Two different textures (spatial frequencies) s and v are present in the odd frames, while the even frames only contain sectors with texture s. Phase is randomized in each frame, each sector. When frame f_n and frame f_{n+1} are presented in succession, two opposing interpretations of the motion direction are possible, as indicated by the arrows. Spatial frequencies for texture s is 1.26c/deg, and for texture v is 0.83c/deg. Outer radius is 3.00 deg. Inner radius is 1.50 deg. (b) The energy model by above authors for the ambiguous second-order motion stimulus shown in the upper row. Activity of each patch is directly proportional to the contrast and inversely proportional to the spatial frequency. Subjects' motion perception depends on the strength of the motion path, which is the product of the activities of two patches from consecutive frames. (c) The open dots are their results of on the experiment described in the text. When patches v's contrast is low, motion perception is along homogeneous path. When v's contrast is high, motion perception is along heterogeneous path.

2.3 Experiments and Results

Four sets of experiments were performed. Experiment I involved a single task identical to that in Werkhoven *et al*(1993). Experiments II-IV involved dual tasks. In Experiment II, observers simultaneously performed a Rapid Serial Visual Presentation (RSVP) letter identification task(Sperling *et al.*, 1971; Sperling & Melchner, 1978) and the apparent motion task. In Experiment III, observers performed dual motion tasks. Experiment IV was a control in which two letter-recognition tasks were performed.

The procedure of Experiment I is the same as in Werkhoven *et al* (1993). Fig.2.1 schematically illustrates how competing(homogeneous vs heterogeneous) motion directions are constructed. Fig.2.2 shows the actual stimuli and results. Observers had to fixate at the central fixation point. From trial to trial, the direction of rotation and the contrast of texture v were chosen randomly. Observers had to perform a forced-choice task, reporting whether the direction of motion was clockwise or counterclockwise. The observers in the current study fell into two categories. For six of ten observers, the percentage of trials in which the heterogeneous direction of motion is perceived increases sigmoidally as the contrast modulation m_v of texture v increases (Fig.2.2, Group A). This is in accordance with the results of Werkhoven *et al* (1993). However, four of ten observers perceived mostly homogeneous motion for all values of m_v, as shown in Fig.2.2 (Group B). They perceived the direction of motion of the sectors with texture s without explicit instructions to do so. Observers were specifically asked to fixate at the center. When being asked afterwards about their subjective perception, they reported having perceived four sectors with texture s. This ruled out the possibility of tracking one particular sector. Instead, these observers seem to perceive motion carried by the attended pattern. The performances of most observers lie at one of the two extreme ends of the distribution of heterogeneous motion. This led me to suspect that the perception of the ambiguous second-order motion stimuli may also be processed through another pathway which computes motion of particular patterns.

Because of the bimodal grouping of observers, in Experiments II and III, observers were split into two groups; each group was given different instructions. The observers in group A, whose responses in experiment I were similar to that on fig.2.1(Lower right), were asked to perceive global motion while fixating at the center. The observers in group B were asked to perceive motion of sectors with texture s, i.e., to attend selectively to s, while fixating at the center. Since group B observers attended to the higher spatial frequency texture s, they were assumed to have activated the third-order (attention-generated) motion mechanism(Lu & Sperling, 1995a; Lu & Sperling, 1995b).

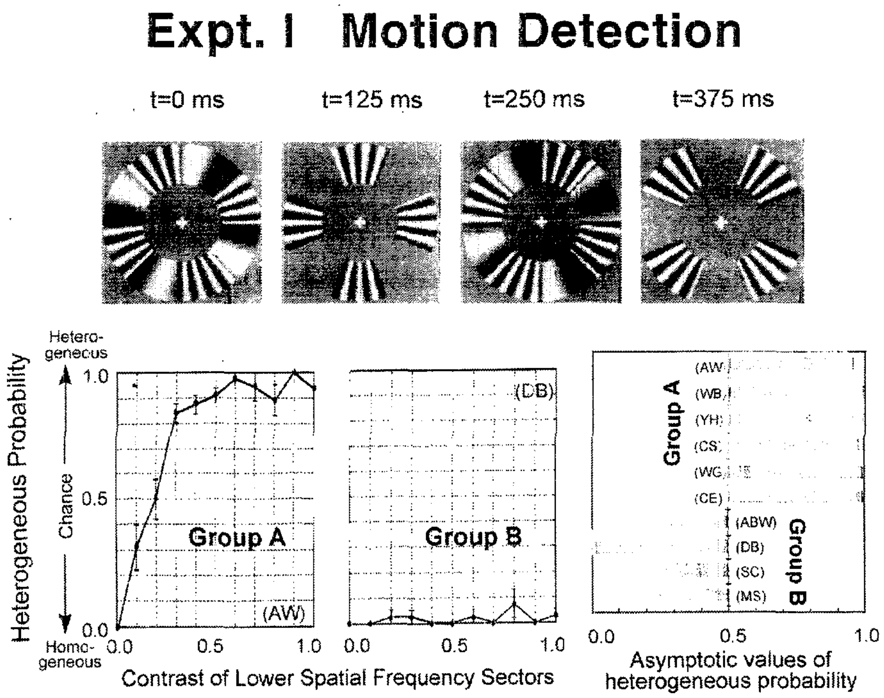


Figure 2.2: **(Top)** Experiments on the perception of second-order motion, using the stimuli described in Fig. 2.1 and illustrated in the top row (expt. I). Observers indicated the perceived motion direction by pressing one of two buttons. For group A **(left panel)**, the heterogeneous motion direction dominates perception as the contrast modulation of the texture type v, which are the low spatial frequency sectors, increases. Group B observers show the opposite behavior **(middle panel)**: they perceive only the homogeneous direction of motion, independent of the contrast modulation of the texture type v. The dominance of the heterogeneous direction (for high contrast stimuli) is displayed for all observers in a bar chart **(right panel)**. Observers: (A) AW, WB, YH, CS, WG, CE (B) DB, SC, ABW, MS.

Chapter 3 Divided Attention Experiments

3.1 Introduction.

The dual tasks paradigm has been used for studying divided attention for centuries. Sperling and Melchner (1978) devised a useful tool, the Attention Operating Characteristic (AOC) graph (see fig.3.1), in which performance of one task is plotted against the performance of another, to measure whether two tasks interfere with each other or if they can operate concurrently. When the dual task performance lies on the upper right corner, the two tasks do not interfere. If it lies between the single task performances, the two tasks interfere. Experiments using dual tasks paradigm to investigate the attentional role in motion perception will be investigated in this chapter.

3.2 Experiments

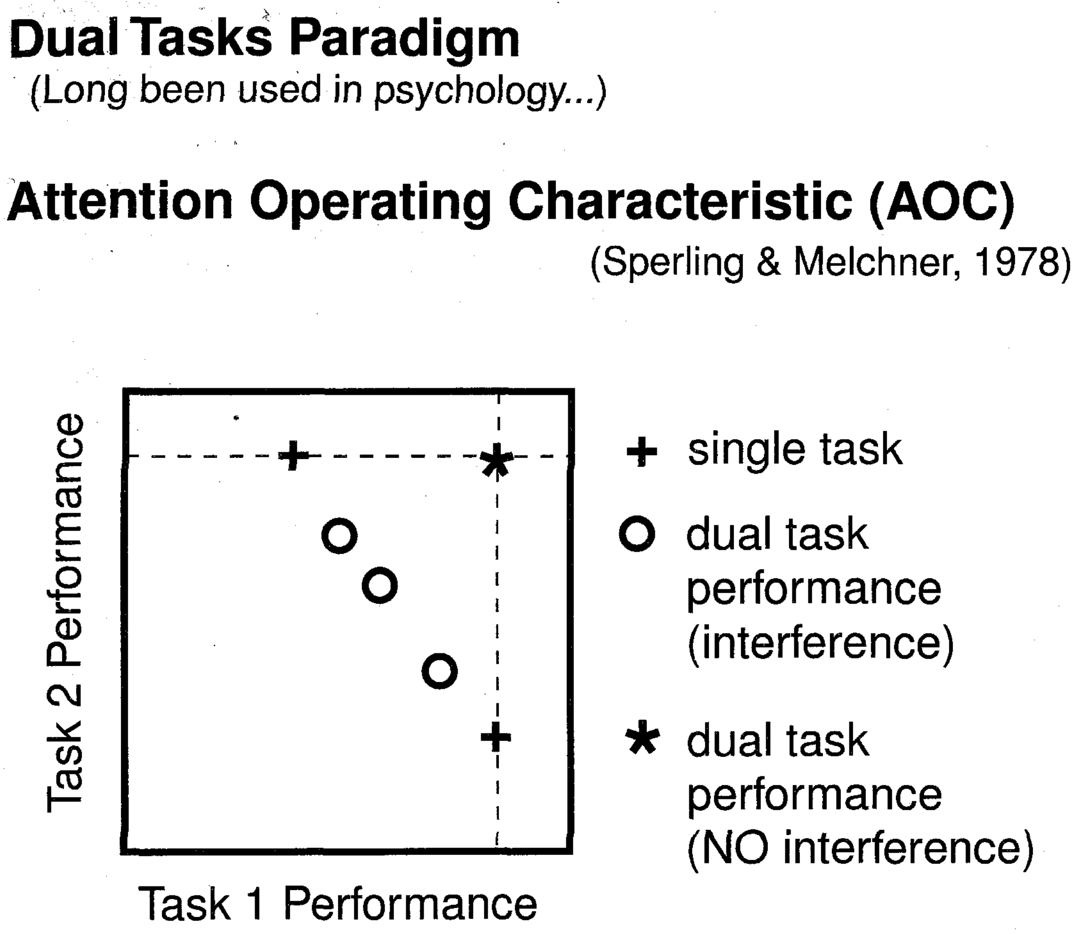
The two motion detection instructions were intended to amplify the natural tendency of the observers. In preliminary experiments in which no explicit instructions were given, the results were qualitatively similar. On the other hand, when observers were asked to do what the opposite group did, their performance, even on single tasks, often fell to chance, making it impossible to investigate the dual tasks effect.

For each combination of dual tasks, five attentional instructions were given: two single task instructions, and three dual task instructions, namely dual tasks with attention given to task 1, dual tasks with attention given to task 2, dual tasks with attention equally split between tasks 1 and 2. Blocks of trials gathered under different instruction were treated as separate data points on the AOC graph. The three dual-task instructions served to spread out the data in the AOC graphs.

In Experiment II, observers were required to concurrently perform a Rapid Serial Visual Presentation (RSVP) letter identification task(Sperling *et al.*, 1971; Sperling & Melchner, 1978) as well as the apparent motion task. **RSVP letter task :** A target letter (a vowel) is presented before the annular motion stimulus is shown. Eight letters, subtending 0.5° of visual angle, are flashed in sequence together with the motion stimulus. Each frame lasts for 115msec-130msec, depending on each observer's performance on letter-recognition task with respect to frame time. The target letter appears between 0 and 3 times in this sequence; consonants serve as distractors. The task of the observer is to count the number of times (0-3) the target letter appeared in the sequence. Two precautions are used to avoid repetition blindness (Kanwisher *et al.*, 1997): no target letters are allowed to appear in two successive frames, and a time delay of 533 msec (very long) separates the appearance of the cue that indicates the target letter and the appearance of the first of the eight letters in sequence. Fig.3.2 shows the performance of the observers. First consider Group A.

The dual task results lie on the upper-right corner of the AOC curves (Fig.3.2 Group A). This indicates no interference between the two tasks. For group B, interference occurred between the RSVP letter-recognition and third-order motion detection. Fig.3.2 (group B) shows the data and linear fit for the dual task performance.

Concurrent tasks of higher-order motion perception and rapid serial visual presentation (RSVP) letter counting (Expt II). Observers are divided into two groups on the motion task, groups A and B, as explained in the text. Second and third rows are, respectively, the results of group A observers (instructions: perceive global motion) and group B observers (instructions: perceive motion of sectors with texture s, treating the sectors with texture v as distractors.). **(left column)** Graphs of heterogeneous dominance vs. contrast modulation of lower spatial frequency sectors, m_v, for one observer from group A, AW, and one from group B, DB, are shown. Observers were alternately asked to report only the direction of motion, or to do both tasks while paying full attention to the RSVP letter task. Trials on which the observer gave an incorrect response on the RSVP task were excluded from further analysis. For group A observers, the motion detection results in both conditions were statistically indistinguishable. However, there is a clear difference between the two curves for group B observers ($P < 0.01$). Observers from group B perceived the homogeneous direction of motion when paying attention to motion detection, but gave chance responses or indicated heterogeneous directions of motion when paying attention to the RSVP letter task. Observers : (A) AW, CS, CE, WB (B)DB, SC, MS, ABW. **(right column)** AOC graphs of RSVP performance vs. heterogeneous motion pathway dominance at contrast $m_v = 0.4$ for two observers each from group A and B. Performance within the white area is above chance for both tasks and those within the grey area are below chance for one or both tasks. Some of the single task performances of heterogeneous dominance are shifted slightly to show the density of data when two or more overlap. Performance of two observers from each group is shown: (A) AW, CE (B) DB, MS; the other two observers in each group also have similar AOC graphs.



Dual Tasks Paradigm

(Long been used in psychology...)

Attention Operating Characteristic (AOC)

(Sperling & Melchner, 1978)

Figure 3.1: Attention Operating Characteristic (AOC) as a quantitative analysis for studying divided attention in a dual task paradigm.

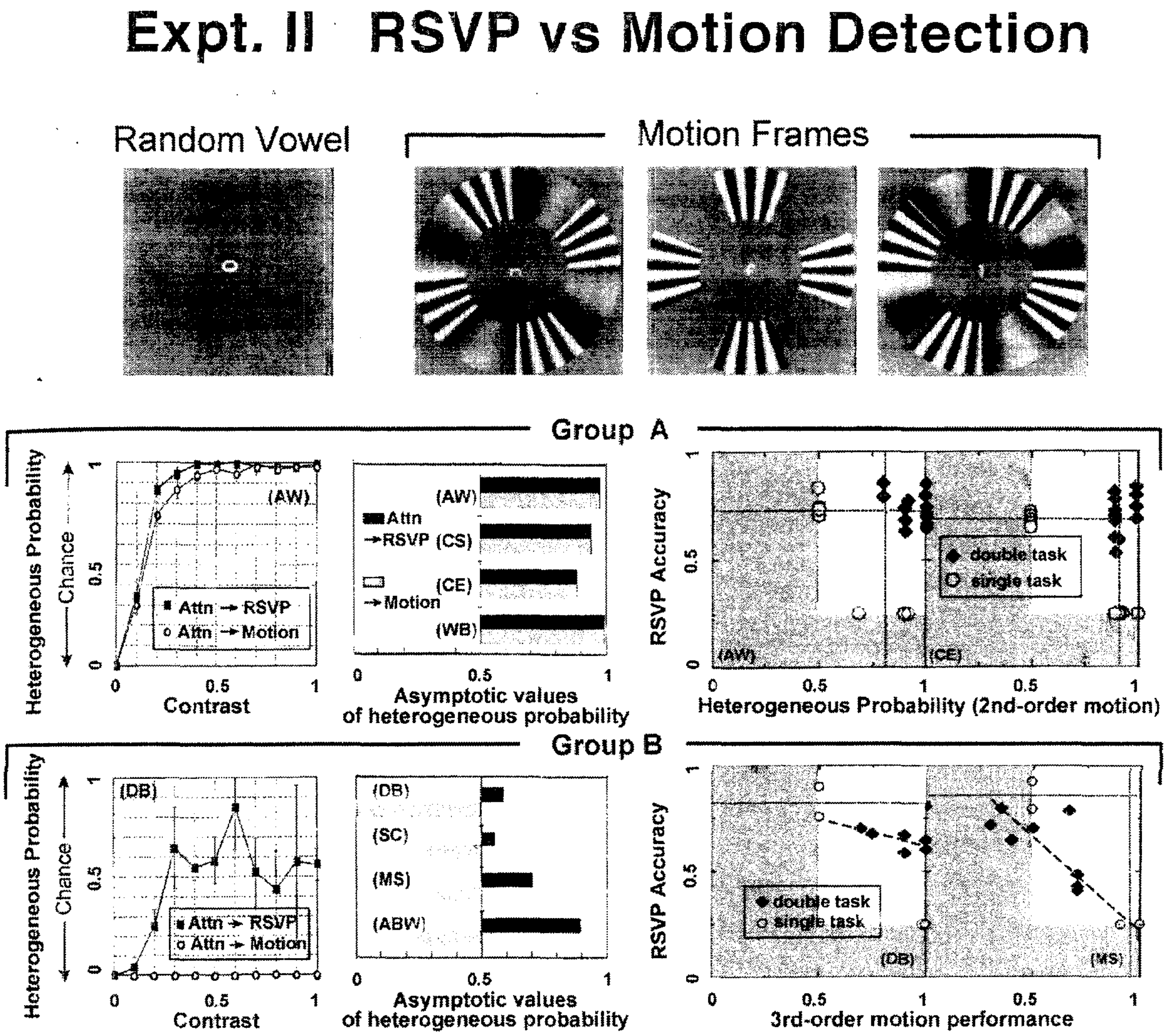


Figure 3.2: Concurrent tasks of higher-order motion perception and rapid serial visual presentation (RSVP) letter counting (Expt II).

Experiment III was a control experiment of dual letter-recognition tasks(Duncan, 1993). Interference was found between two letter-recognition tasks. The result confirms Duncan (1993)'s finding that two letter recognition tasks interfere with each other.

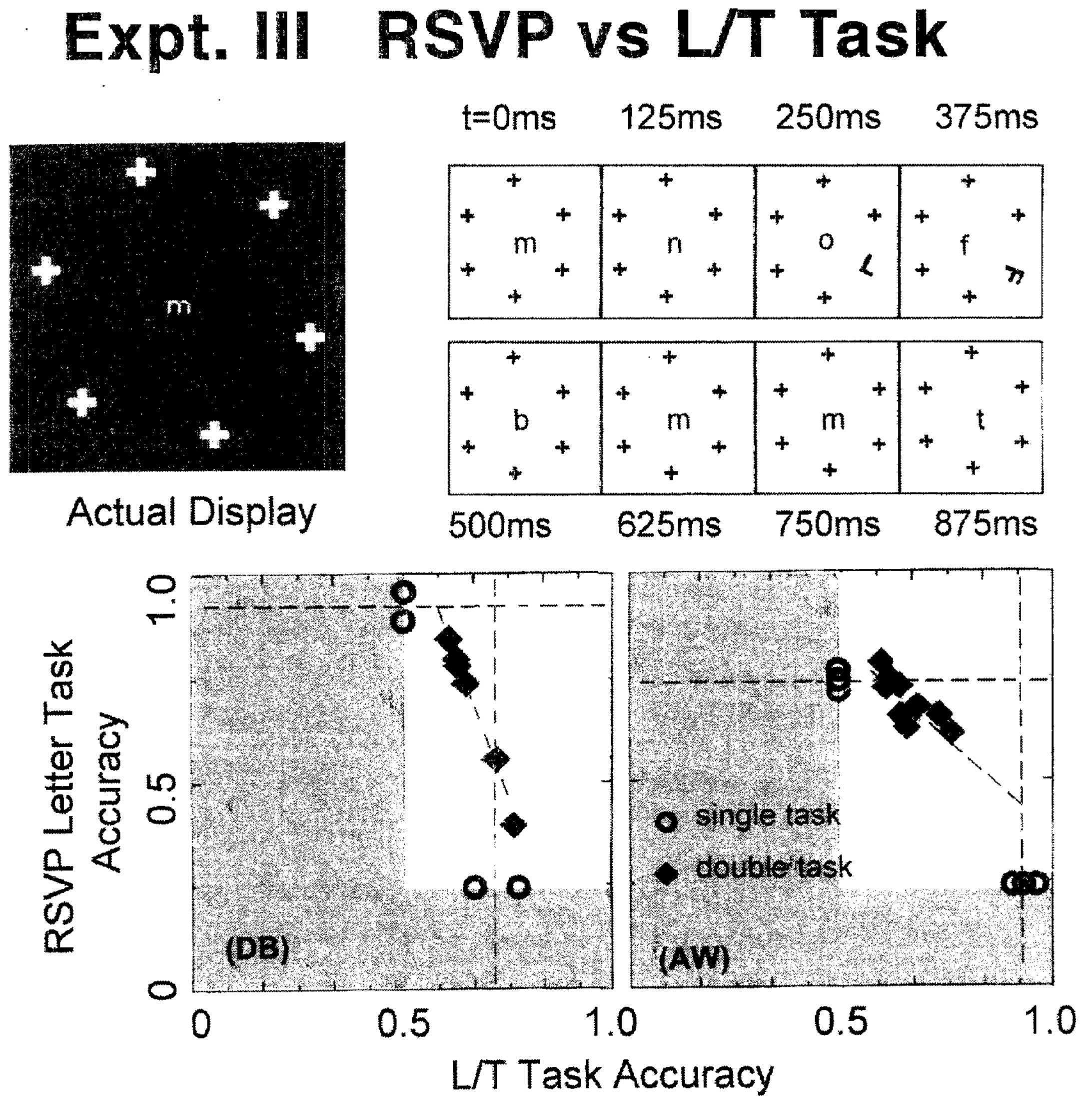


Figure 3.3: Concurrent tasks of rapid serial visual presentation (RSVP) letter counting and L/T discrimination (Expt III). The RSVP letter counting task is the same as that in Expt II. The L/T task is a 2AFC task in which at one of the six locations(random location), during one of the eight frames (random time), an 'L' or a 'T' is presented, and masked by an 'F'. Subjects has to report whether it is an 'L' or a 'T'.

In Experiment IV, two ambiguous motion stimuli are put together in the form of inner and outer annuli for dual motion perception tasks. Stimuli are shown in fig.3.4. A total of four frames are presented. Two distractor annuli are placed in between the inner and outer annuli to avoid the perception that the two annuli always seem to rotate in the same directions. The inner and outer radii of the inner annulus are 0.8° and 1.6° respectively, and that of the outer annulus are 4.0° and 5.0° respectively. Group A was asked to perceive global motion while group B was asked to perceive motion of the sectors with texture s. Results are shown in Fig.3.4. Interference occurs for both group A and group B observers, but the patterns of interference are different. For group A, experimental dual task performance is straight or slightly curved line segments above the diagonal line joining the single task performance. For group B, the dual third-order motion task, the dual task performance data scatter in a concave region under the diagonal line joining the single tasks performance in the AOC graph, as shown in Fig.3.4 group B.

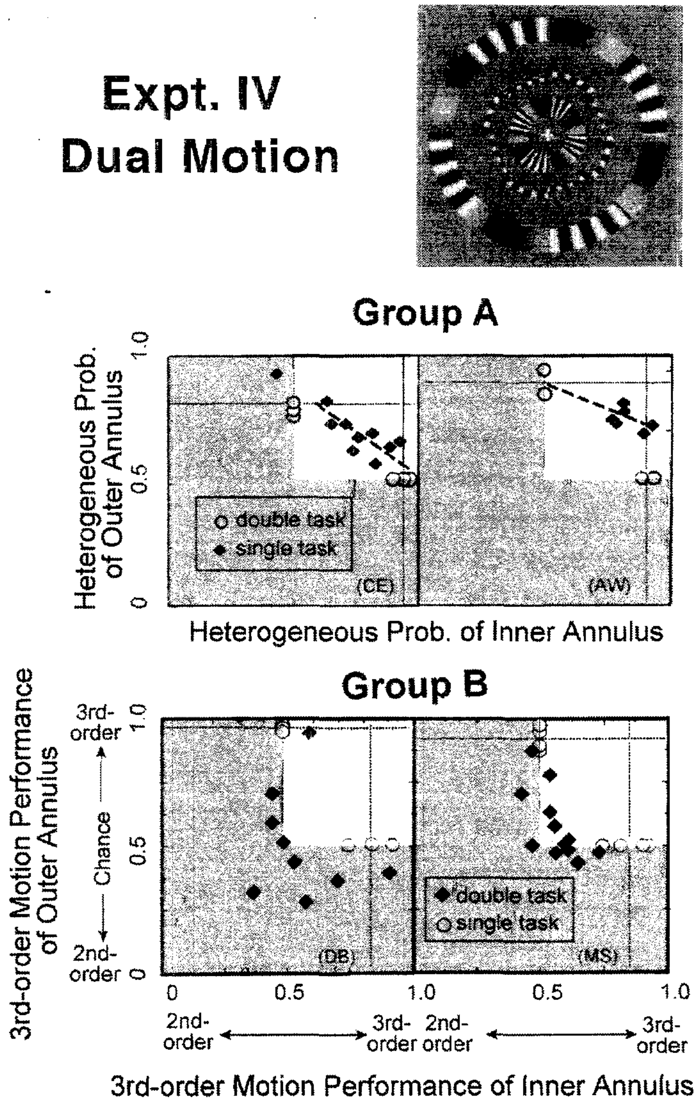


Figure 3.4: **(Top row)** Stimuli used in Expt. IV dual-motion detection tasks. **(Second row)**AOC graphs for group A observers. The abscissa and the ordinate represent heterogeneous direction of motion perception for inner and outer annuli, respectively. **(Bottom row)**AOC graphs for group B observers. Axes represent perception of the homogeneous motion direction of outer and inner annuli. Data for dual-task performance lie in a concave area under the diagonal line. AOC graphs of two observers from each group is shown : (A) AW, CE; (B) DB, MS. Two other observers for each group show similar behavior.

The results of the experiments reveal fundamental differences in the way the ambiguous motion stimulus of Figs. 2.1(Upper Row) and 2.2 was perceived by group A and group B observers. If we assume that there is only one pathway and interpret the data in Experiment II as that when observers attended to the central RSVP task, the activities of higher spatial frequency sectors decreased, the dual motion task performance should had similar patterns in the AOC graph for groups A and B such that both had straight lines or both had curves. This contradicts the results in Experiment III which show different patterns in group A's vs group B's AOC graphs. A more possible reason is that textures to RSVP the ambiguous motion stimulus is processed by two motion pathways in the brains, a second-order motion pathway and a third-order(attention-generated) motion (Lu & Sperling, 1995a; Lu & Sperling, 1995b) pathway. When observers attend to a particular pattern, such as texture s, motion composed only of this pattern (the homogeneous direction) is seen by the third-order system. The second-order pathway computes the motion to be in either the homogeneous or the heterogeneous direction, depending on the activities of the sectors (Werkhoven *et al.*, 1993). These two pathways are different in their architectures (as being discussed in more detail in the Discussion Section), giving the different patterns of AOC graphs in the experiments in this study.

In summary, two pathways were found for the ambiguous motion stimuli, a second-order motion pathway and a third-order motion pathway. Interference occurred for dual letter-recognition tasks. A third-order motion task interfered with an RSVP letter task, and it interfered with another third-order motion task in a more severe manner. A second-order motion task interfered with another second-order motion task, but a second-order motion task did not interfere with an RSVP letter task.

Tasks	Interference Results
Letter vs Letter	Yes
2nd-order motion vs Letter	No
2nd-order motion vs 2nd-order motion	Yes
3rd-order motion vs Letter	Yes
3rd-rder motion vs 3rd-order motion	Yes Yes

3.3 Discussion

Insofar as task interference implies the need for attention, the complex interference effects and the apparently paradoxical interference effects of second-order motion perception imply that there are multiple forms of attention or in another word, attention is distributed. Whether two tasks interfere or not depends on whether they require the same resource.

Insofar as spatio-temporal processing is assumed to be carried out in the dorsal stream, and pattern recognition in the ventral stream, the interference patterns suggest that second-order motion may be computed entirely in the dorsal stream, while third-order motion may involve two computational processes, one of which shares computational resources with the letter-recognition task in the ventral stream.

These results were reported in Ho & Koch (1996) and Ho (1996; 1998a; 1998b).

Chapter 4 Bringing Everything Together: Structure-Attention Mapping

In order to explain the complicated interference effects from the previous chapters, a method, Structure-Attention Mapping (fig4.1) was devised, based on the assumption that attention is related to the fraction of limited resource for computation of the given task in the brain.

4.1 Structure-Attention-Mapping (SAM)

Attention has been generally assumed to be related to the fraction of limited resources one uses. Psychologists consider attention problems as classroom attendence problems(Sperling & Dosher, 1986). Physiologist look at the neuronal firing rates or firing rate contrasts to see the effect of attention, fMRI scientists look at correlates of blood flow. In the brain, there are certain areas that provide the resources for the computations necessary to perform each task. Here, the computational resources will be termed *processors.* Using a model of limited resources, I show that one can derive different AOC graphs from the assumed architecture of the processors. This method, Structure-Attention Mapping (SAM), help to understand the different forms of interferences quantitatively from the physiological point of view. I will use the result later to explain the experimental data. Depending on the architecture of how different attention processes are constructed, several cases will be discussed in the following sections.

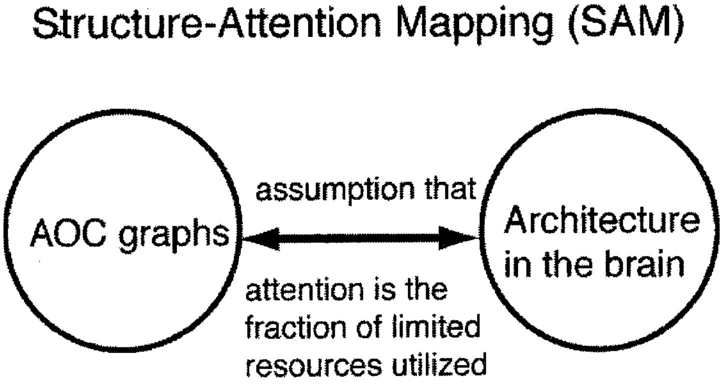


Figure 4.1: Structure Attention Mapping. A method to relate AOC graphs to the architecture in the brain, based on the assumption that attention is related to the fraction of limited resource in the brain used for processing a task.

4.1.1 Two tasks sharing single attention-processor

Let us consider an attention-processor which is the computational resources at the bottleneck of two or more tasks. The attention-processor between two different pairs of tasks, say tasks (X,Y) and tasks (X,Z), may be different. The fraction of resources used is a continuous value from 0 to 1. The probability of the neurons of successfully computation of the task as a function of attention-processor utilization, x, without the consideration of noise is very likely a threshold function of 0 and 1.

$$p(x) = \begin{cases} 0 & x < x_{t'} \\ 1 & x \geq x_{t'} \end{cases} \tag{4.1}$$

When a observer performs the alternative forced choice tasks, he makes guesses when the computations fail, giving a nonzero performance at chance and there may also be some other errors that limit the maximum performance from 1 to a smaller value. Hence the performance as a function of attention-processor utilization, x, without the consideration of noise is very likely a threshold function :

$$O(x) = \begin{cases} a & x < x_{t'} \\ b & x \geq x_{t'} \end{cases} \tag{4.2}$$

where a is the chance level and b is the maximum performance. When we obtain the probability over a large number of trials, the attention of each single trials will not be the same, but a distributed around a certain mean value, $\langle x \rangle$. For simplicity, I write $\langle x \rangle$ as x. Then the performance will become a sigmoid. Including the factor of noise, the sigmoid will be further smoothed out. Making a piecewise linear approximation, and an assumption that the significant part is the slope and the saturation, I found that this model fits all of our data. So I postulate the performances as a function of attention-processor utilizations for all the tasks of my experiments to be:

$$P(x) = \begin{cases} a + (b - a)\frac{x}{x_t} & x < x_t \\ b & x \geq x_t \end{cases} \tag{4.3}$$

One can think of it intuitively as that when no attention is paid, performance is at chance, and when attention or the utilization of attention-processor is increased, the performance increases, until it comes to certain point that maximum performance is reached, then further increase of attention or attention-processor utilization does not help anymore.

For given two concurrent tasks, X and Y, parametrically plotting P_y and P_x, performances of task Y and task X respectively, one can obtain the AOC graphs based on the given architecture, as shown in fig.4.2. Some examples of the AOC graphs are shown in fig.4.6 and fig.4.7.

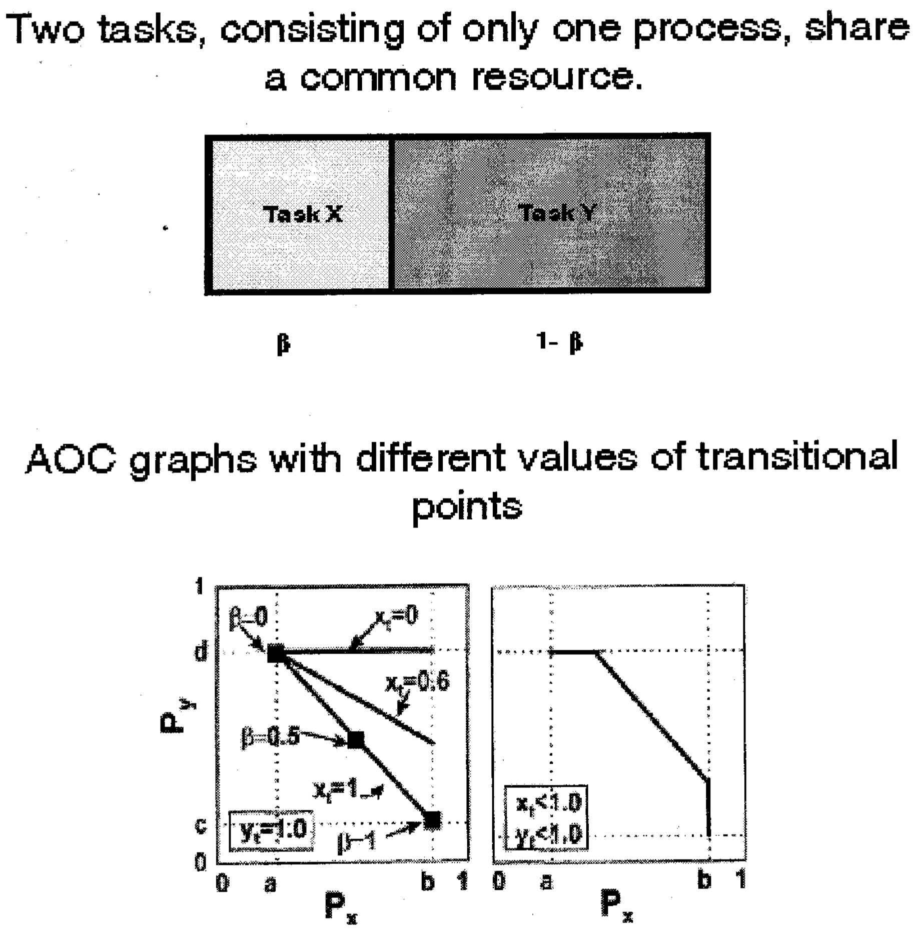


Figure 4.2: One possible set of architecture: two tasks sharing one attentional bottleneck. Simulated AOC graphs are shown below the resource architecture. For different values of transitional points, x_t and y_t, for tasks X and Y, the AOC graphs are different. The larger the transitional points are, i.e. the more attentional resource required for each task alone, the stronger the interference is. The strongest interference pattern is a diagonal line joining the single tasks performances For weaker interferences, the dual task data will lie on a line above the diagonal line.

4.1.2 Two tasks sharing compound attention-processors

If there exist two such attention-processors working independently (that is, the processor-utilizations are uncorrelated in each trial), and the final result must combine the outcome of both attention-processors, the performance will be a product of the two separate performances. Let us call the two processors, processor 1 and processor 2, with processor utilization, α and β, and processor 1 precedes processor 2 such that information is first feed into processor 1 and its outcome is feed into processor 2 for further processing. If processor 1 does not get the right result, processor 2 will not be able to compute at all. If processor 1 get the right answer, the performance will depend on processor 2's performance. The probability of successful computations by both processors, $p(\alpha, \beta)$ is the probability of successful computation of the first processor, $p_1(\alpha)$, multiplies that of second processor given it has been successfully computed in the first processor, $p_2(\beta)$, where α and β are the processor utilization of processor 1 and 2 respectively.

$$p_1(\alpha) \sim \alpha / x_{t1} \tag{4.4}$$

$$p_2(\beta) \sim \beta / x_{t2} \tag{4.5}$$

$$p(\alpha, \beta) \sim p_1(\alpha) p_2(\beta) \tag{4.6}$$

Hence the performance on the task as a function of utilizations, α and β can be postulated as:

$$P(\alpha, \beta) = \begin{cases} a + (b-a)\frac{\alpha\beta}{x_{t1}x_{t2}} & \alpha < x_{t1} \text{ and } \beta < x_{t2} \\ a + (b-a)\frac{\alpha}{x_{t1}} & \alpha < x_{t1} \text{ and } \beta \geq x_{t2} \\ a + (b-a)\frac{\beta}{x_{t2}} & \alpha \geq x_{t1} \text{ and } \beta < x_{t2} \\ b & \alpha \geq x_{t1} \text{ and } \beta \geq x_{t2} \end{cases} \tag{4.7}$$

where a is the chance level of performance and b is the maximum performance, and P_1 and P_2 are the probabilities of successful computations of processor 1 and 2 respectively. If one of the two tasks is very easy such that either $P_1 \simeq 1$ or $P_2 \simeq 1$, or if α and β on each trial are strongly coupled, P behaves like a single processor.

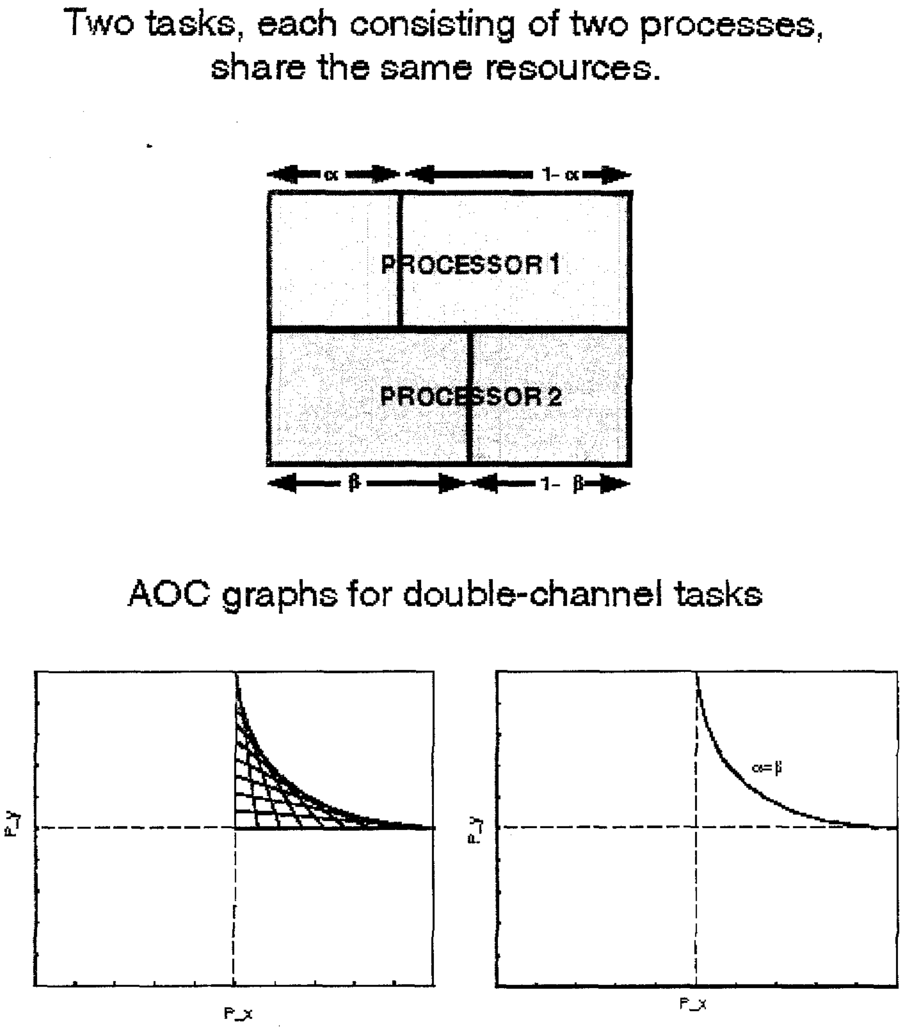


Figure 4.3: Another possible architecture: two tasks compete for two consecutive attentional bottlenecks.

4.1.3 Two tasks utilizing completely separate attention-processors

If cells are functionally distinct with respect to the two given tasks, according to SAM, they are completely separate processors. Since each task can use all the resource in respective processor,

$$P_x = b \tag{4.8}$$

$$P_y = d \tag{4.9}$$

with notations as in single processor case above. The double tasks performance are on the upper right corner. An example of AOC plot can be found in fig.4.6 Expt.II.

4.2 Two motion pathways for one motion stimulus

4.2.1 Two computation-pathways architecture:

The results of the experiments led me to propose the architecture for the second-order motion stimuli and letter recognition tasks as shown in fig.4.5. Two separate computation pathways compute motion: PRP and MPP for third-order motion and SMC for second-order motion. Inputs are sent to both pathways. Each pathway computes the given tasks independently and the results compete according to their outputs and their weights. Assuming the decision process is a summation of inputs from the two pathways, we have:

$$\mathbf{P}_{\mathbf{x}} = qP_x(\alpha_1) + (1-q)Q_x(\alpha_2) \tag{4.10}$$

$$\mathbf{P}_{\mathbf{y}} = qP_y(1-\alpha_1) + (1-q)Q_y(1-\alpha_2) \tag{4.11}$$

where $\mathbf{P}_{\mathbf{x}}$ and $\mathbf{P}_{\mathbf{y}}$ are the final performances of tasks X and Y respectively, α_1 and α_2 are the utilizations for PRP and MPP, and SMC respectively (assuming equal average utilization along same computation-pathway), P_x and P_y are performances for the computation-pathway along PRP and MPP for tasks X and Y respectively, Q_x and Q_y are performances for the computation-pathway along SMC for tasks X and Y respectively, the P's and Q's correspond to the compound attention-processor equation 4.7, q is the normalized weight for the output of computation-pathway along PRP and MPP. This weight is close to 0 group A observers, and close to 1 for group B observers. It can be changed voluntarily when one wants to pay attention to certain feature, forcing that particular feature to become "salient" with respect to this particular competition. Hence, q is small when perceiving global motion, and q is close to 1 when perceiving the motion of certain features.

4.2.2 Fitting the data

Experiment II : Second-order motion vs RSVP letter recognition

No interference for second-order motion vs RSVP letter recognition task, as shown in fig.3.2(A). Interference occurs for pattern-tracking third-order motion vs RSVP letter recognition task, as shown in fig.3.2(B).

For part (A), the RSVP letter task vs second-order motion, the double tasks results lie on the upper-right corner of the AOC curves. The result can only be that there are two separate processors for the two tasks.

For part (B), fig.4.4 shows the least-square fit for the data of the double tasks performance of Experiment RSVP vs pattern-tracking third-order motion. If we assume there is a single processor for both RSVP and third-order, pattern-tracking motion, the result implies that $x_t < 1$, that is, the performance of a third-order, pattern-tracking motion task saturates before full attention is used. With some linear algebra, we can obtain the ratio of y_t to x_t, calculate the exact value of x_t, and hence the ratio of the difficulties of the two tasks. Together with the result from Experiment RSVP vs L/T that $y_t = 1$, we can exactly calculate the value of x_t.

On the other hand, if we assume the two tasks are computed in partially overlapped, but otherwise, distinct pathways, we will get the overlapping region, $k = 1 + r - ry_t$. Together with the result that $y_t = 1$, we get $k = 1$, meaning the processor of third-order, pattern-tracking motion is completely included in the attentional processor of RSVP letter task.

With some linear algebra, we can also compute what the maximum value for r, the ratio of attentional size of RSVP to that of third-order, pattern-tracking motion, is.

$$P_x = q_1 P_y + q_2 \tag{4.12}$$

$$P_x = b \to P_y = S_y \tag{4.13}$$

$$P_y = d \to P_x = S_x \tag{4.14}$$

$$S_x = a \tag{4.15}$$

$$a = q_1 d + q_2 \tag{4.16}$$

With single processor model

$$q_1 = -\frac{b-a}{d-c}\frac{y_t}{x_t} \tag{4.17}$$

$$q_2 = a + \frac{d(b-a)}{d-c}\frac{y_t}{x_t} \tag{4.18}$$

$$\frac{y_t}{x_t} = -q_1\frac{d-c}{b-a} \tag{4.19}$$

if $y_t = 1$

$$x_t = -\frac{b-a}{q_1(d-c)} \tag{4.20}$$

With multiple processor model

$$q_1 = -\frac{b-a}{d-c}\frac{ry_t}{x_t} \tag{4.21}$$

$$q_2 = a + \frac{b-a}{x_t}[\frac{c}{d-c}ry_t + 1 + r - k] \tag{4.22}$$

for $q_2 = a$

$$k = 1 + r - ry_t \tag{4.23}$$

if $y_t = 1$

$$k = 1 \tag{4.24}$$

from eqn(23)

$$\frac{r}{x_t} = -q_1\frac{b-a}{d-c} \tag{4.25}$$

$$r \leq -q1\frac{b-a}{d-c} \tag{4.26}$$

$$\tag{4.27}$$

This gives a maximum ratio for task Y to task X. The data and fittings of DB and SC are shown in table 4.1.

For observer (DB) in fig.4.4,

q_1	q_2	a	b	c	d
-2.14	2.27	0.5	1.0	0.25	0.82

Single processor model (assuming $y_t = 1$) gives :

$$x_t = 0.40$$

Single processor model (assuming $y_t = 1$) gives :

$$\frac{r}{x_t} = 2.49$$

$$x_t \leq 1 \Rightarrow r \leq 2.49$$

We got the ratio of size of attention processor of RSVP to that of third-order, pattern-tracking motion has to be less than 2.4866.

For observer (SC) in fig.4.4,

q_1	q_2	a	b	c	d
-1.38	1.52	0.5	1.0	0.25	0.75

Single processor model (assuming $y_t = 1$) gives :

$$x_t = 0.72 \tag{4.32}$$

Single processor model (assuming $y_t = 1$) gives :

$$\frac{r}{x_t} = 1.38 \tag{4.33}$$

$$x_t \leq 1 \Rightarrow r \leq 1.38 \tag{4.34}$$

4.2.3 Model

Attention has been generally assumed to be related to the fraction of limited resources one uses(Sperling & Dosher, 1986). In the brain, there are certain areas that provide the bottleneck resources for the computations necessary to perform each task. Here, the computational bottleneck resources will be termed *processors*, and the pathways of these processors, *pathways*. Using a model of limited resources, one can derive quantitatively the different shapes of AOC graphs from the assumed architecture of the processors (see Appendix).

Under this model, all tasks require attention. It is only a matter of how much attention the task requires. The level of interference can be simulated with the equations in Appendix. The only time two tasks do not interfere is when their computational pathways are non-overlapping. Different pathways require different forms of attentions.

Two letter-recognition tasks required the same form of attention, and similarly for two second-order motion detection task and two third-order motion detection tasks. Therefore interference occurred in all these dual tasks combination. However, the second-order motion detection task required a different form of attention from the RSVP letter-recognition task. Hence no interference occurred in this combination. On the other hand, part of the third-order motion computation required the same form of attention as the RSVP letter-recognition task did, so interference occurred(Ho & Koch, n.d.).

The results of all the four experiments in this study can be captured by the simple model architecture in Fig.4.5 with two computational pathways. The first computational pathway performs third-order motion detection and consists of a pattern-recognition processor (PRP) that performs pattern recognition functions, followed by a motion-from-pattern processor (MPP) that performs motion computations on attended patterns of the stimulus. Second-order motion detection occurs along a separate computational pathway—the second-order-motion processor (SMP). PRP is separated from SMP. MPP may or may not overlap with SMP which is outside the scope of this study.

The different perceptions of motion by group A and group B observers in Experiment I can be interpreted as that, group A observers primarily used the second-order motion pathway while group B observers primarily used the third-order motion pathway. In Experiment II, RSVP letter tasks and third-order motion detection both require resources from the PRP leading to interference in group B observers. Second-order motion detection occurs in SMP which is separate from PRP, explaining the lack of interference between the RSVP letter-recognition task and second-order motion detection for Group A in Experiment II. In fact, for some group A observers, the RSVP letter tasks even improved their second-order motion perception since they may had used some third-order motion mechanism when they did motion detection only. Note that in the AOC graphs of fig.3.2 (Group B), the motion direction judgement for the dual task falls to chance before the letter-recognition performance reaches single task performance. The reason is that when observers were paying attention to the RSVP letter task, their performance on the third-order motion detection was poor, hence the motion perception was given by the second-order mechanism. In Experiment III, the concave nature of the AOC graphs for dual third-order motion tasks further supports the hypothesis that the two processors PRP and MPP are separate, consecutive stages in third-order motion detection.

4.2.4 The Model in relation to physiological findings

In the primate visual system, it is believed that shape recognition is processed through the ventral, or "what" stream, which passes through the inferior temporal area (IT), while motion is processed through the dorsal, or "where" stream, passing through the medial temporal lobe (MT)(Maunsell & Van Essen, 1983a; Maunsell & Van Essen, 1983b; Albright, 1982; Albright, 1984) and medial superior temporal area (MST)(Graziano *et al.*, 1994). According to these neurophysiological findings, the RSVP letter task is probably processed in the ventral stream, somewhere near human equivalent of IT. Since third-order motion shares the pattern-recognition processor (PRP) with the RSVP letter task, PRP is also located in the ventral stream. The motion-from-pattern processor (MPP) of the third-order motion pathway, being completely distinct from the PRP, is possibly in the dorsal stream, though this is not directly inferred from the results in this study. Second-order motion is proposed to be processed through V1-V2-MT (Wilson *et al.*, 1992; Vaina & Cowey, 1996). Our experiments show that it does not interfere with the RSVP letter-recognition task. This supports that second-order motion processing, and hence the SMP of the model in Fig.4.5, may be entirely in the dorsal stream. This is consistent with the models of Wilson *et al*(Wilson *et al.*, 1992) and the findings of Vaina and Cowey(Vaina & Cowey, 1996).

Recent findings (Dobbins *et al.*, 1998) showed that distance, a known dorsal stream function, was also found in ventral stream. Hence, while motion is generally known to be a dorsal stream function, It is possible that some of the early processes of third-order motion is computed in ventral stream.

parameters and results	DB	SC
a	0.5	0.5
b	1.0	1.0
c	0.25	0.25
d	0.82	0.75
q_1	-2.14	-1.38
q_2	2.27	1.52
Single Channel : x_t(assuming $y_t = 1$	0.40	0.72
Multiple Channel : $\frac{r}{x_t}$ (assuming $y_t = 1$)	0.92	1.38
$r \leq$	2.49	1.38

Table 4.1: Parameters, fittings and results for single and multiple processor models for RSVP vs third-order, pattern-tracking motion. Subjects: DB, SC.

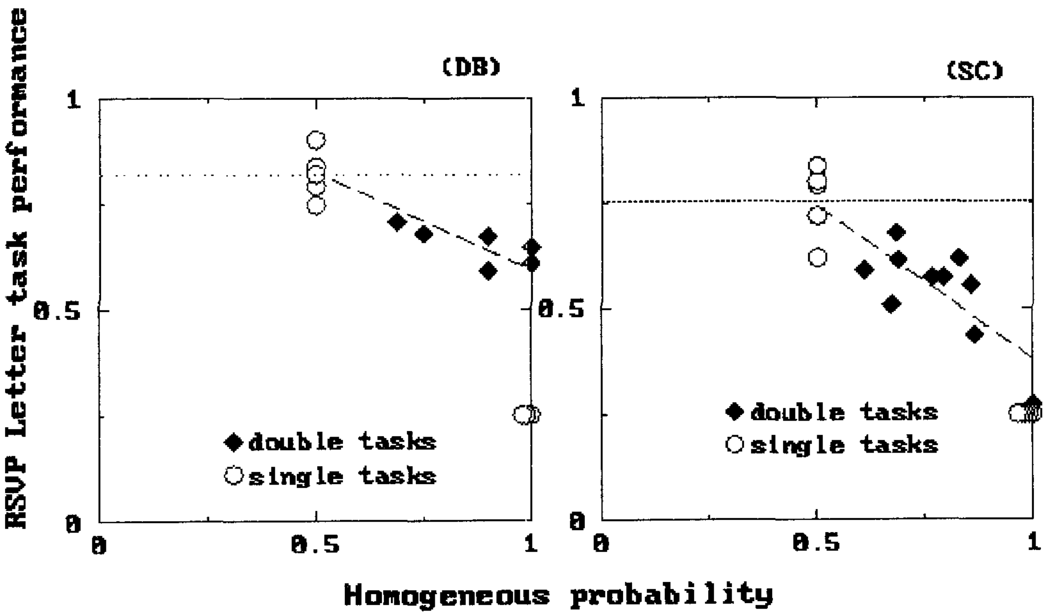


Figure 4.4: Least Square fits for third-order, pattern-tracking motion. The single performances of motion task are plotted with offsets in order to distinguish points of same values. Their actual values are all equal to 1. Subjects DB, SC

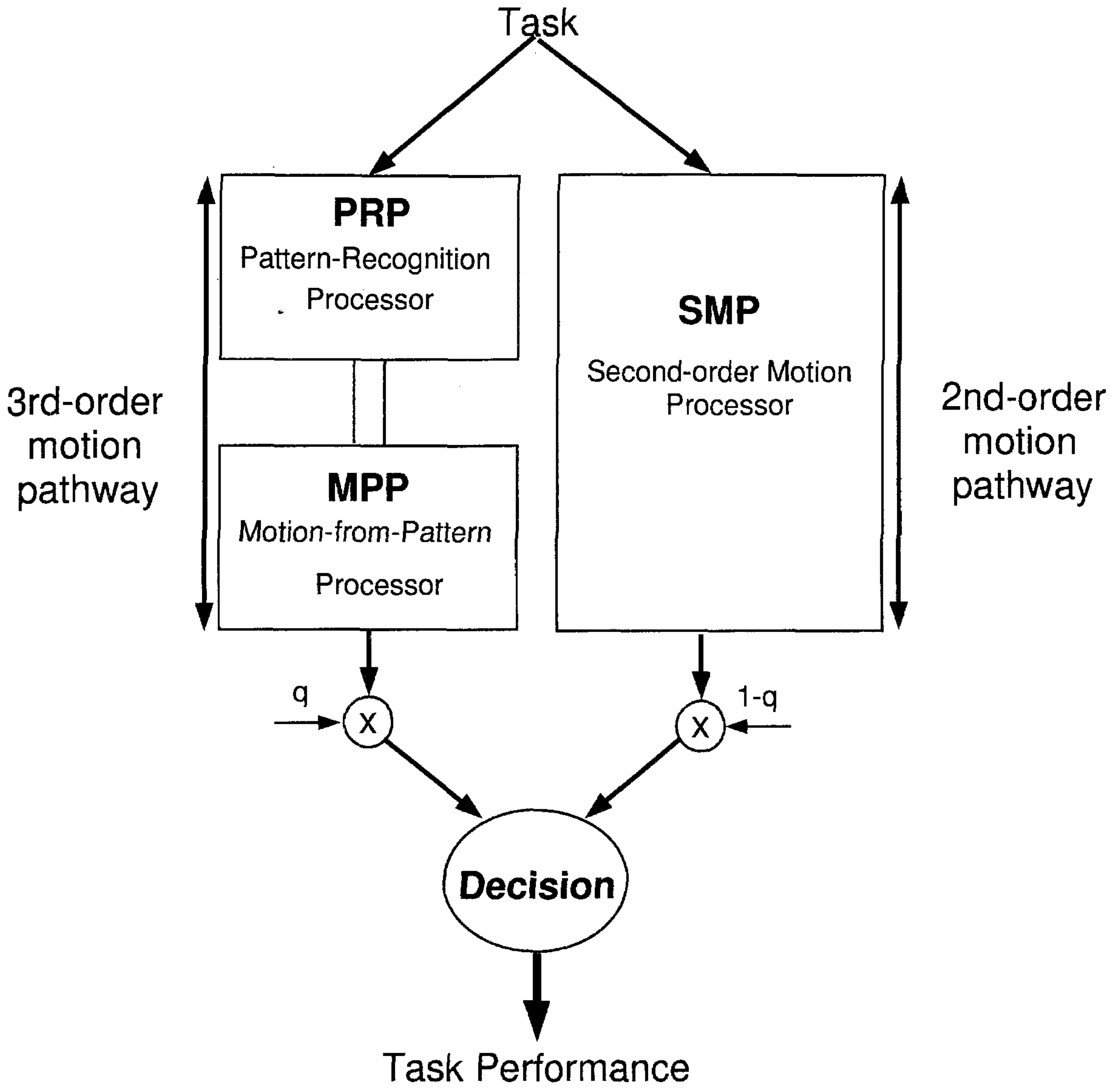


Figure 4.5: A model architecture fitting all of the data. The model contains three elements : (1)pattern-recognition processor (PRP), (2)motion-from pattern processor(MPP) (3)second-order motion processor(SMP). PRP and MPP form the third-order motion computation-pathway, SMP is the second-order motion computation-pathway. Sensory information is sent to both computation-pathways and processed independently. The output of each pathway is multiplied by normalized weight, q and (1-q), which bias the direction of perceived motion. The weights q and (1-q) are different for different observers, and are assumed to be a function of selective attention.

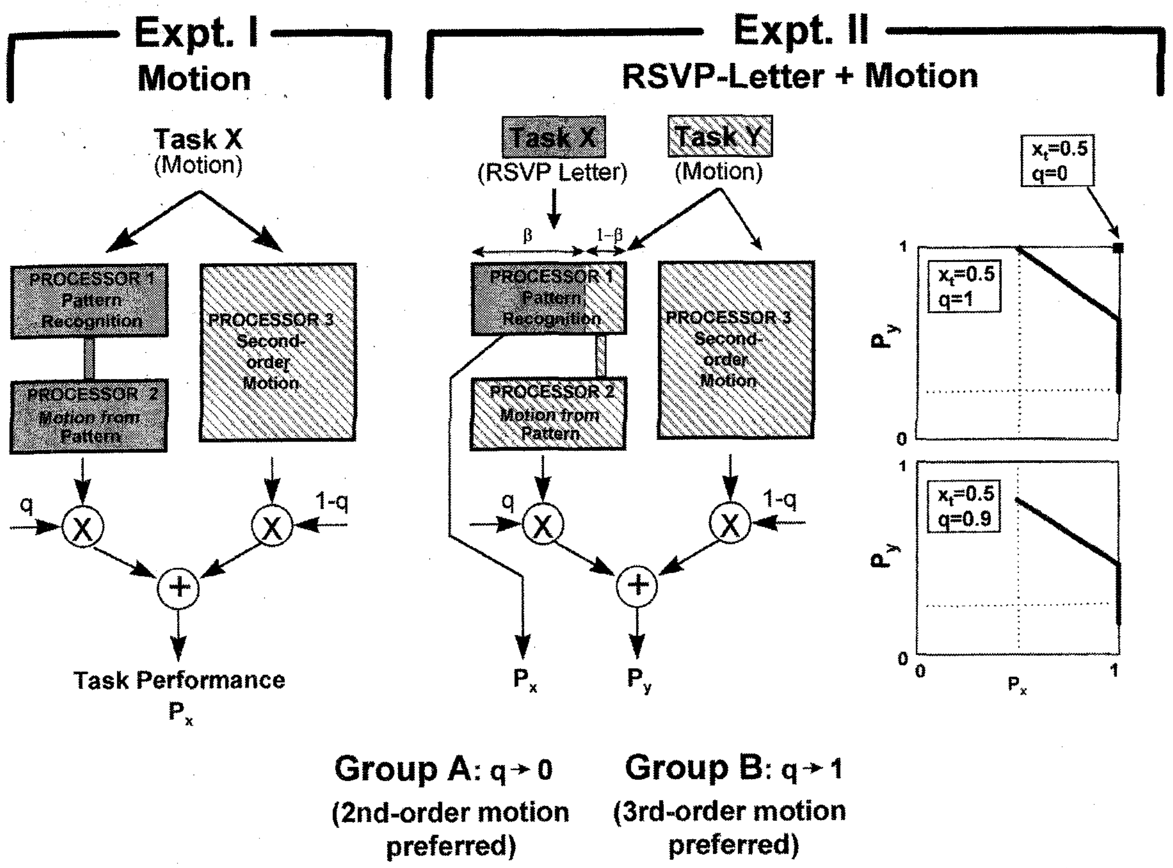


Figure 4.6: The model explaining data of Expts I and II.

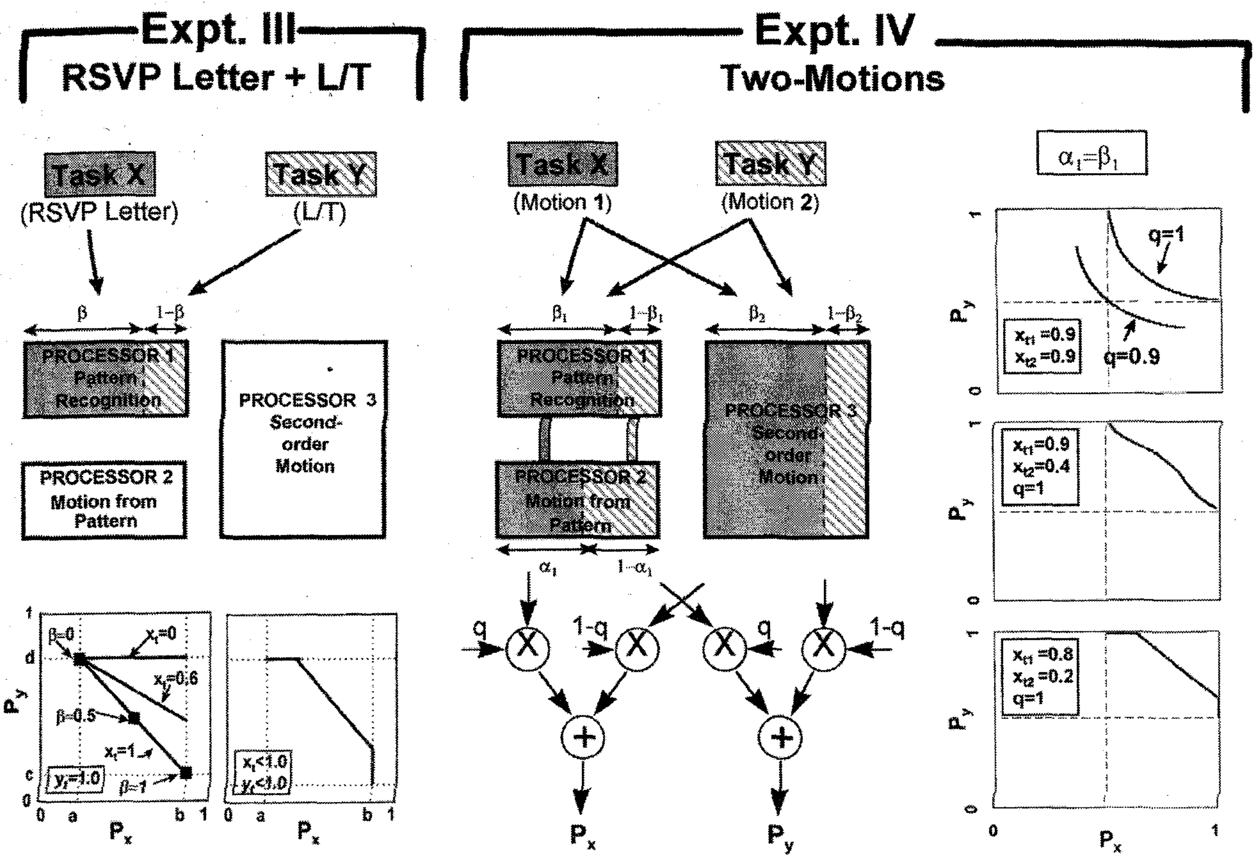


Figure 4.7: The model explaining data of Expts III and IV.

Chapter 5 Selecting Second- and Third-order Motion Systems

5.1 Introduction:

The three pathways for perceiving the first-, second- and third-order motion probably exist in every normal person. However, not everyone activates their pathways to the same extent. This may be due to genetic background or experience. Individual differences in the activation of our various functional pathways are often found in ordinary life. Some people are talented singers, actors or athletes, some are not. Training and practising can compensate for a lack of talent. Though everybody's ability to access these pathways may be different, everybody possesses these pathways. However, correct instructions must be provided in order to achieve the ideal results. In classical singing, systematic procedures for training vocal techniques are emphasized. Similarly, in theater acting, techniques like Stanislavsky's "The Method", are used for training actors in vocal, relaxation, movements, and affective memory. Athletes and dancers go through similarly or even more rigorous trainings. In the following chapters, factors concerning the selection of the three functional pathways for motion perception will be investigated, and procedures for activating observers' pathways and for inducing observers to select one over the others will be developed.

5.1.1 Change of terminology: Pathway → System

The term "pathway" which means the collection of attentional bottlenecks for perceptual processing is more meaningful for dual task experiments, where there is competition with another task. In the following experiments, only motion perception is considered. The bottleneck will be the decision process for the three motion mechanisms. Hence, the term "pathway" loses its meaning, and can be confusing. Therefore, "system" will be used to replace "pathway" from now on.

5.1.2 Backgrounds and Objectives

In the previous chapters, it was found that some observers perceived ambiguous texture-defined stimuli as moving in one direction, corresponding to motion computation by a second-order motion system, whereas other observers perceived the same stimuli as moving in the opposite direction, corresponding to a third-order motion system. In this chapter, the selection of higher-order motion systems will be examined more carefully by manipulating observers' perceptions of an ambiguous texture-defined motion stimulus. Lu and Sperling (1995b) showed that (1) first-, second- and third-order motions have corner frequencies at 12Hz, 12Hz and 3Hz respectively, and (2) third-order motion is intrinsically binocular whereas first- and second-order motion are primarily monocular. The goal here is to show that depending on stimulus parameters, training, and attentional instructions, every observer can be induced to perceive a texture-defined *stimulus in either a second-* or a third-order motion direction. These manipulations bias the competition between the second-order and the third-order motion perception mechanisms. In chapter 6, competition of the third-order motion system with a first-order motion system in a luminance-defined stimulus will be investigated.

Werkhoven, Sperling, and Chubb (1993) showed that second-order motion followed a motion-energy model as illustrated in Fig. 2.1 (b) and is basically not affected by grating slant. Later, Werkhoven, Sperling and Chubb (1994) found a small effect of grating slant that disappeared at high temporal frequencies, an apparent violation of their energy model. The results in this chapter will show that this slant effect is actually due to the third-order motion perception, not to second-order motion perception. Hence the new result supports Werkhoven *et al*'s (1993) earlier energy model for second-order motion that grating slant is irrelevant.

5.2 Stimuli for Testing Second- vs Third-Order Motion Perception

For most texture-defined motion stimuli, the directions perceived by the second-order and the third-order systems are the same. The stimulus here has the property that directions perceived by the second-order and third-order systems are in the opposite directions. The stimulus is an ambiguous motion display which appears to move in either of two directions (see Fig. 5.1). It consists of a sequence of eight or more frames (total display duration no shorter than 0.25 second) with alternating grating patches, s (standard), and v (variable). The s and v patches have the same spatial frequencies. A final frame, similar to the first frame, and with the same duration, was also displayed at the end. In the orthogonal-slant conditions, they are orthogonal to each other, blocks of target slant (same as slant in patches of s) at 45° interleft with blocks or target slant at $-45°$. Patches of type s appears in all frames while patches of type v only appears in alternate frames. Two motion directions, here defined as the second-order direction and the third-order direction can be perceived from each stimulus. The contrasts of s and v are adjusted ($m_s < m_v$) such that according to the motion energy model of Werkhoven *et al* (1993), as shown in Fig. 2.1(b), second-order direction is along $v \rightarrow s \rightarrow v \rightarrow s$, as shown in Fig. 5.1. If observers attend to target slant, which has the same direction as s, in successive frames, the other direction, $\rightarrow s \rightarrow s \rightarrow s$, This direction, computed by a pattern(feature)-tracking, third-order system, is called the third-order direction.

Spatial phases of the sinusoids within each patch are randomized in each frame to eliminate any first-order component. The starting frame is randomized between s, v, s, v, ..., v,s,v,s..., and s, blank, s, blank,... and blank,s,blank,s,.... The direction of the second-order motion is randomized between left and right. These randomizations produced thousands of stimuli of the same type. This makes it impossible for the observers to learn anything about specific stimuli. Blocks of trials, typically 80, with target slant of 45° and −45° are alternated. Subjects have to report the direction motion perception with three confidence levels for each direction (6AFC). Fig. 5.1 shows four samples out of a longer sequence of motion frames used in Condition I.

A technical complication was revealed by some recent studies (Smith & Ledgeway, 1997; Nam & Chubb, 1998; Lu & Sperling, 1999) that there is an asymmetric distortion in human perception of black and white, such that sensitivity to negative contrast ("black") is greater than sensitivity to positive contrast ("white"). According to Lu & Sperling (1999), distortion is proportional to contrast. Hence a first-order component is often present in a texture-defined stimulus. Lu & Sperling (1999) devised a black-white calibration method to eliminate the first-order component due to perceptual distortion. The texture-defined stimuli used in this study all went through such black-white calibration (see Appendix); the contrast of all patches was kept below 0.4.

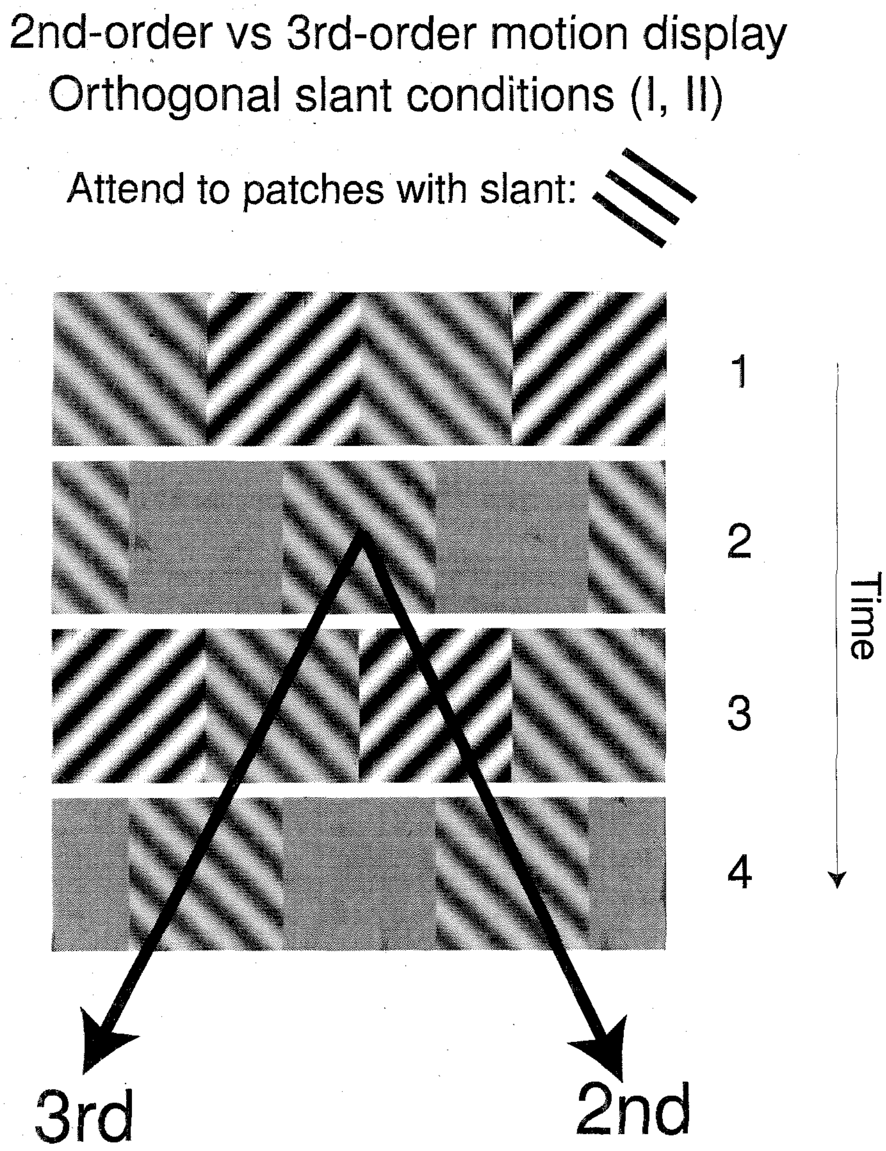


Figure 5.1: Second- vs third-order motion stimulus with orthogonal-slant patches. Four frames of a longer motion sequence of texture-defined motion display are shown, each consists of a strip with eight alternating patches, *s* (standard) and *v* (variable). Patches *s* contain grating slants at one of 45° in this example. Patches *v* contains grating slants at −45° here. The grating slants are consistent throughout a motion sequence. *s* patches appear in every frame while *v* patches appear either only in odd-numbered or only in even-numbered frames. The contrast of *s*, and *v*, are 0.2 and 0.4 respectively. Two motion directions, the second-order and the third-order, can be perceived in each motion sequence.

5.3 Initial Tendencies

Subjects were asked to judge motion directions. Before being given specific attentional instructions, three out of seven observers perceived motion in the third-order direction. and four perceived motion in the second-order motion direction (Fig. 5.2. In the following sections, various manipulations were tested for their effectiveness in altering observers' perceived direction of motion.

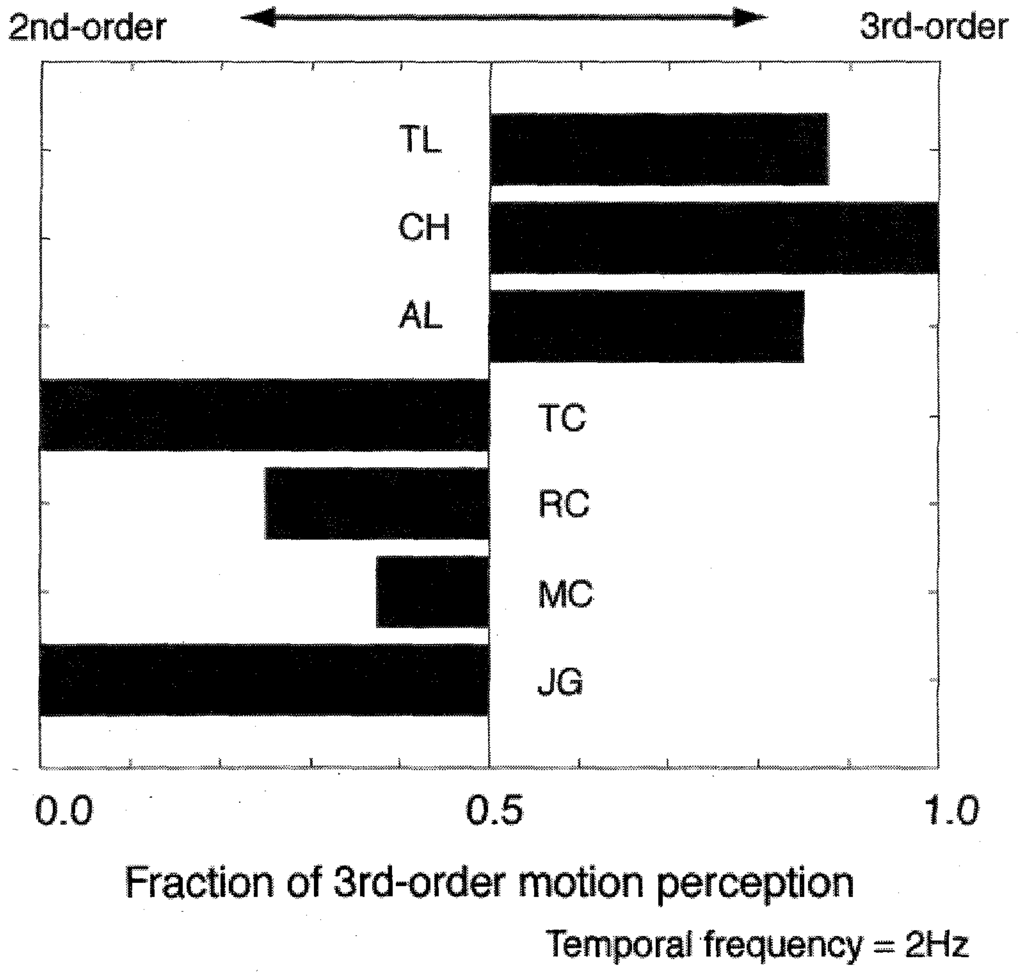


Figure 5.2: Direction of perceived motion at temporal frequency of 2Hz, before attentional instructions on whether to attend to a particular slant direction. Subjects JG, MC, RC, TC, AL, CH and TL.

5.4 Changing the Direction of Apparent Motion from Second- to Third-Order : Attentional Instructions

In the experiments reported in chapter 3 (Ho, 1998b), report of the third-order direction was enhanced when observers who perceived the third-order direction initially were told to attend to patches of s; and report of the second-order direction was enhanced when observers who perceived the second-order direction initially were told to attend to the global motion of the annulus. It had been shown previously that observers' consistency in reporting motion in either the second- or the third-order direction was enhanced by attentional instructions. It had also been hypothesized in the model of Fig. 4.5 (Ho, 1998b) that the weight of each system could be changed by top-down attention. Therefore, the first factor investigated here was attentional instructions. How effective is an attentional instruction to attend to a particular direction of slant in switching observers' motion perception from the second-order direction to third-order direction?

The four observers who perceived motion in the second-order direction were told to:

"Fixate at the fixation point. Attend to the patches with target grating slant 45° or −45° (demonstrate example), and try to perceive the motion of the attended patches. Report whatever the strongest motion direction perceived, regardless whether it is by tracking target patches or not."

Two observers were able to perceive the third-order direction. As shown in Fig. 5.3, these two observers' perception of motion direction at low temporal frequencies switched completely from the second-order direction to the third-order direction. At higher temporal frequencies, all observers' perception of motion direction was close to chance.

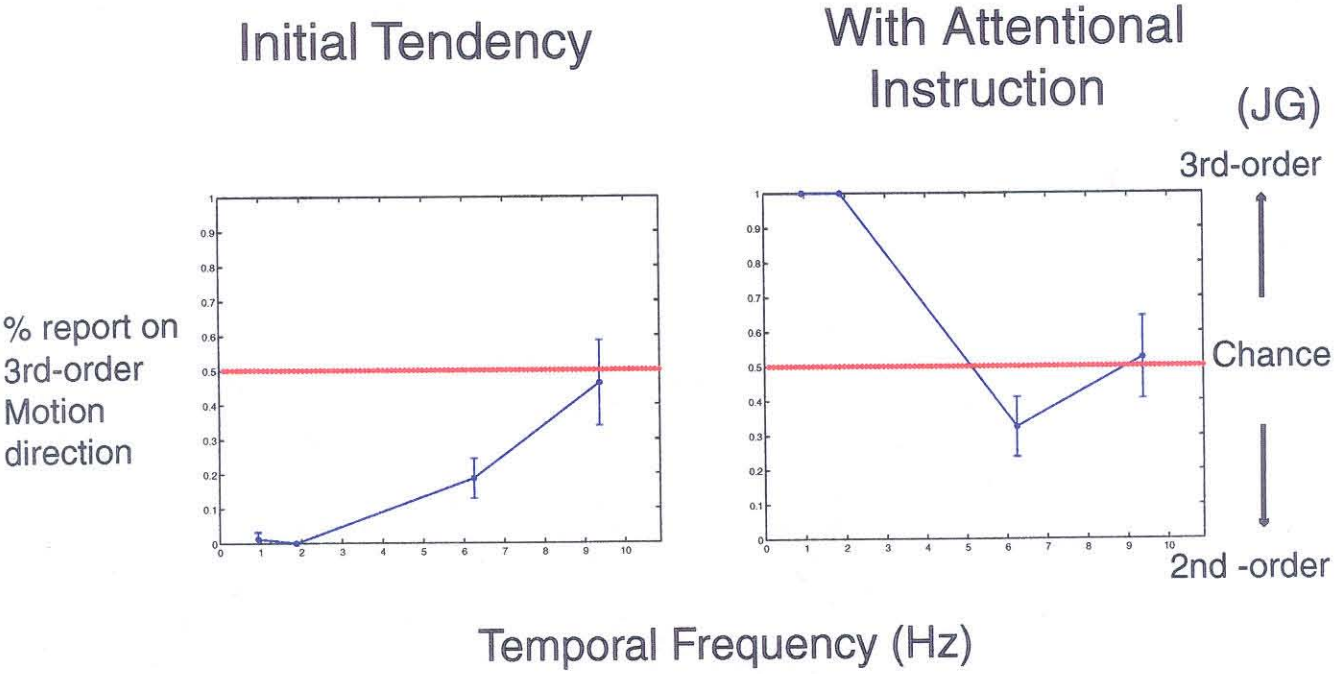


Figure 5.3: **Perceived direction of motion before and after instruction to attend to a particular slant direction.** Before the attentional instruction, this observer's initial tendency was to perceive the second-order direction for all frequencies until it was too high that the performance fell to chance. After the attentional instructions, this observer perceived the third-order direction at low temporal frequencies. Subject: JG. One other observer behaved similarly.

these two observers' perception of motion direction at low temporal frequencies switched completely from the second-order direction to the third-order direction. At higher temporal frequencies, all observers' perception of motion direction was close to chance.

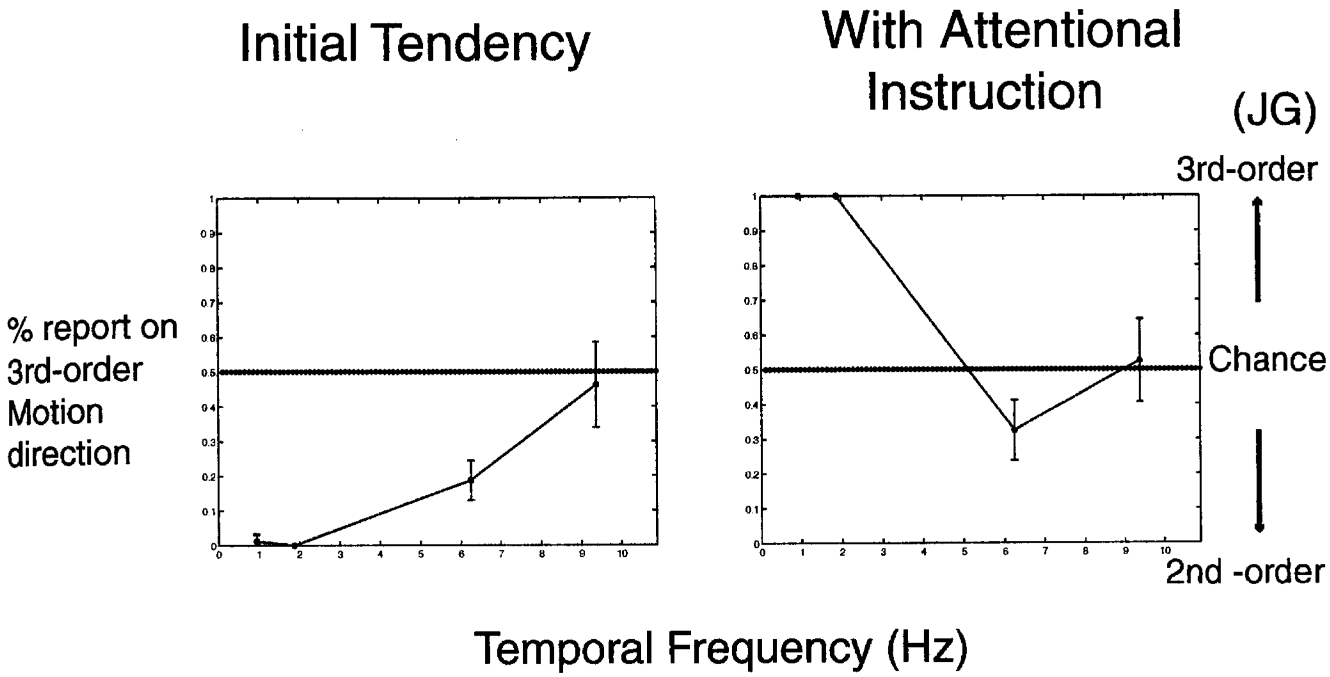


Figure 5.3: **Perceived direction of motion before and after instruction to attend to a particular slant direction.** Before the attentional instruction, this observer's initial tendency was to perceive the second-order direction for all frequencies until it was too high that the performance fell to chance. After the attentional instructions, this observer perceived the third-order direction at low temporal frequencies. Subject: JG. One other observer behaved similarly.

5.5 Training

After the instruction to attend to a particular direction of slant, two of the four observers whose initial tendency was to perceive second-order direction, perceived motion in the third-order direction. The other two did not. A method of fading was used to train the remaining two observers to attend to slant. First the observers were presented with a slow motion stimulus, less than 1 Hz, to help the observers to recognize the target grating slant. This produced an increased tendency to perceive motion in the third-order direction, but it was not completely successful. Subsequently, the observers was presented with a stimulus with patches s (the to-be-attended slant) colored red, as shown in Fig. 5.6. The helping color was subsequently faded so that the training stimuli turned back to the original stimuli. Data of one observer are shown in Fig. 5.5. Before training was given, when this subject was presented with motion stimuli as shown in Fig. 5.1, the second-order direction was mainly perceived. After going through the training procedure for about ten minutes, when this subject was presented with the stimuli as shown in Fig. 5.1 again, the third-order motion direction was mainly perceived. The other observer behaved similarly.

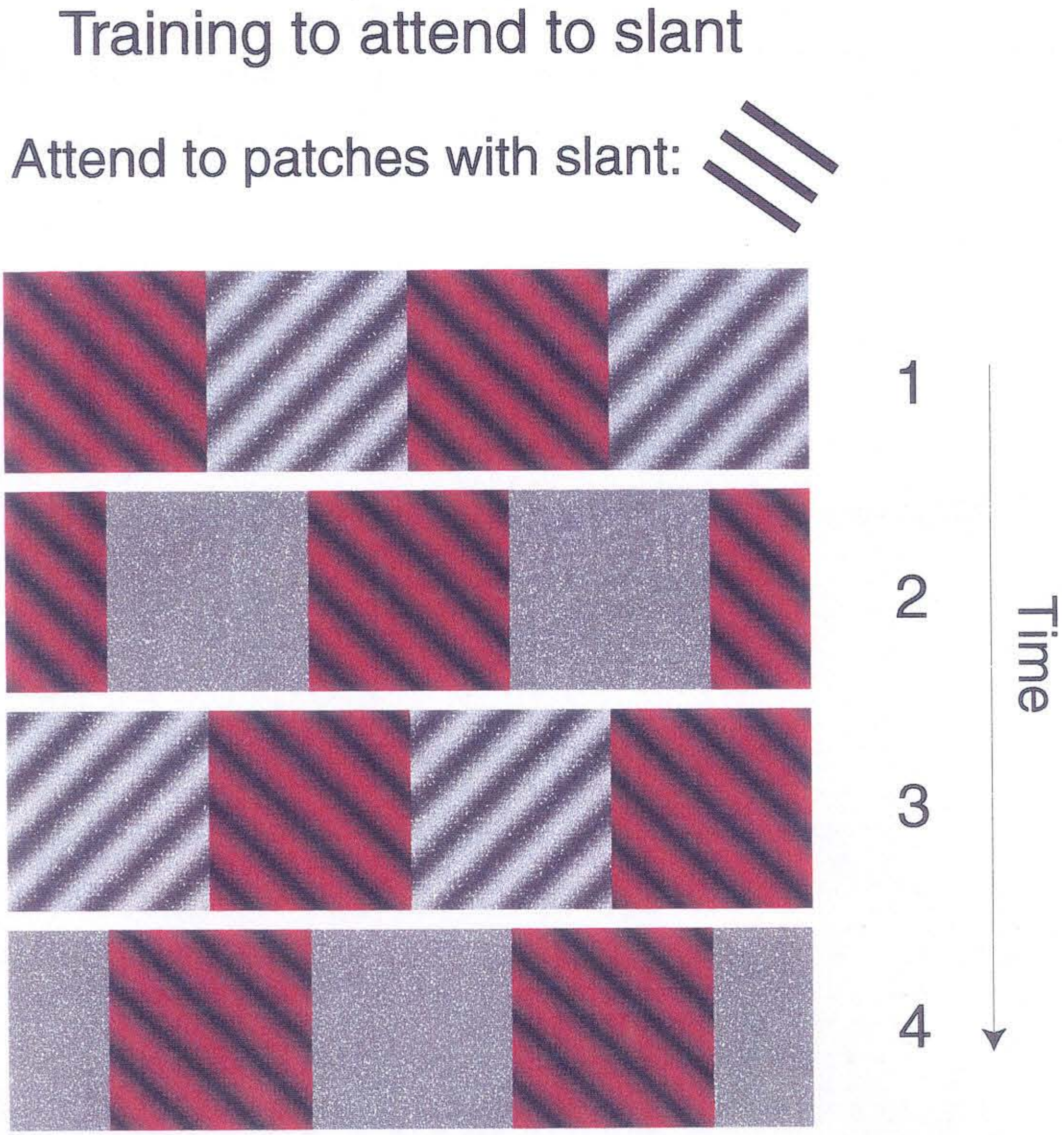


Figure 5.4: **Training-stimulus for perceiving slant.** Training with color and with the fading method was used for two observers who still could not consistently perceive third-order motion after being given attentional instruction to attend to a particular slant direction, and practising with motion frames at a very low temporal frequency.

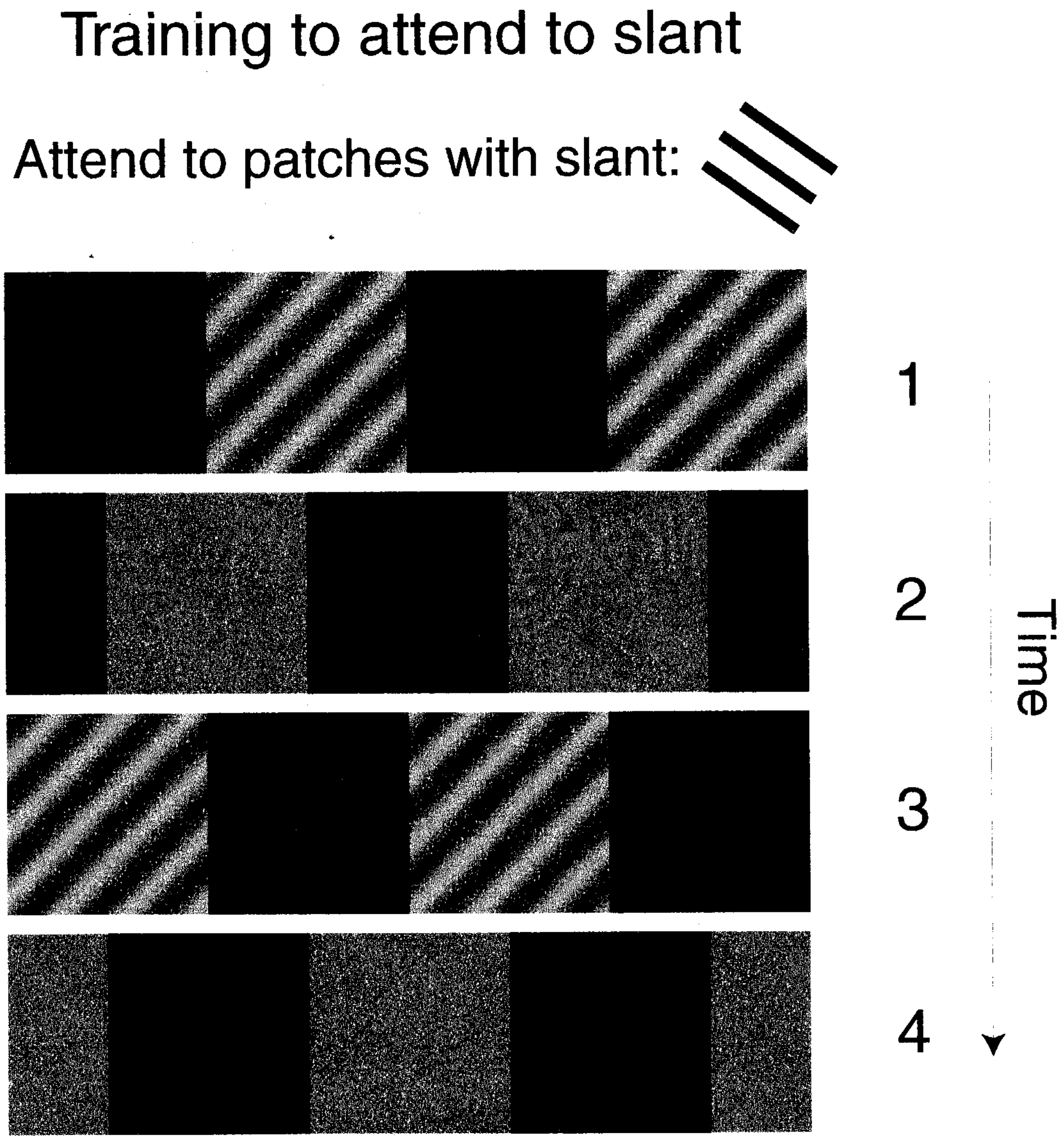


Figure 5.4: **Training-stimulus for perceiving slant.** Training with color and with the fading method was used for two observers who still could not consistently perceive third-order motion after being given attentional instruction to attend to a particular slant direction, and practising with motion frames at a very low temporal frequency.

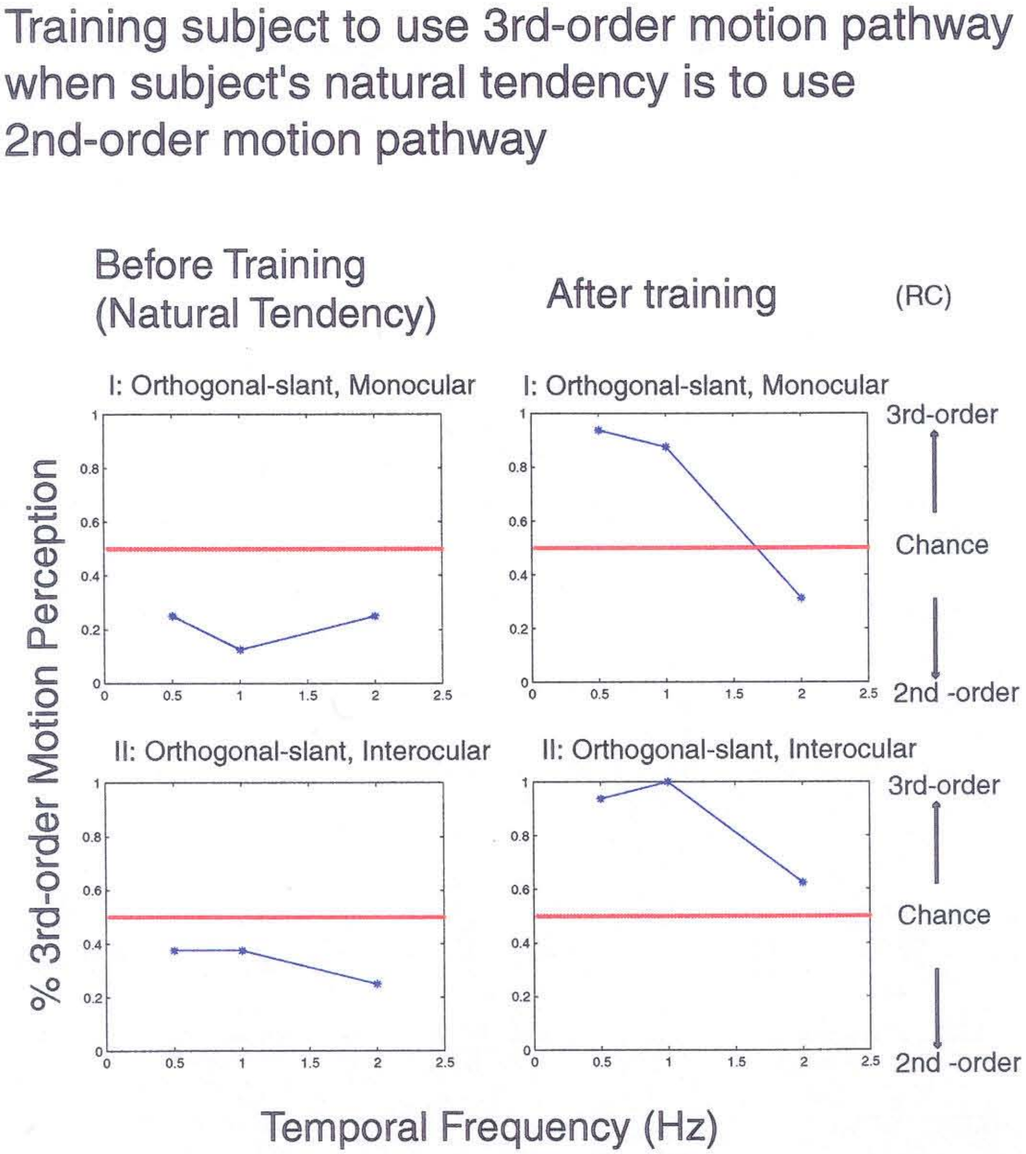


Figure 5.5: Effect of training. Perceived direction of motion before and after training with fading. Figure shows observers' motion perception of orthogonal-slant stimuli under monocular and interocular viewing conditions, before and after training. Subject: RC. The other observer behaved similarly.

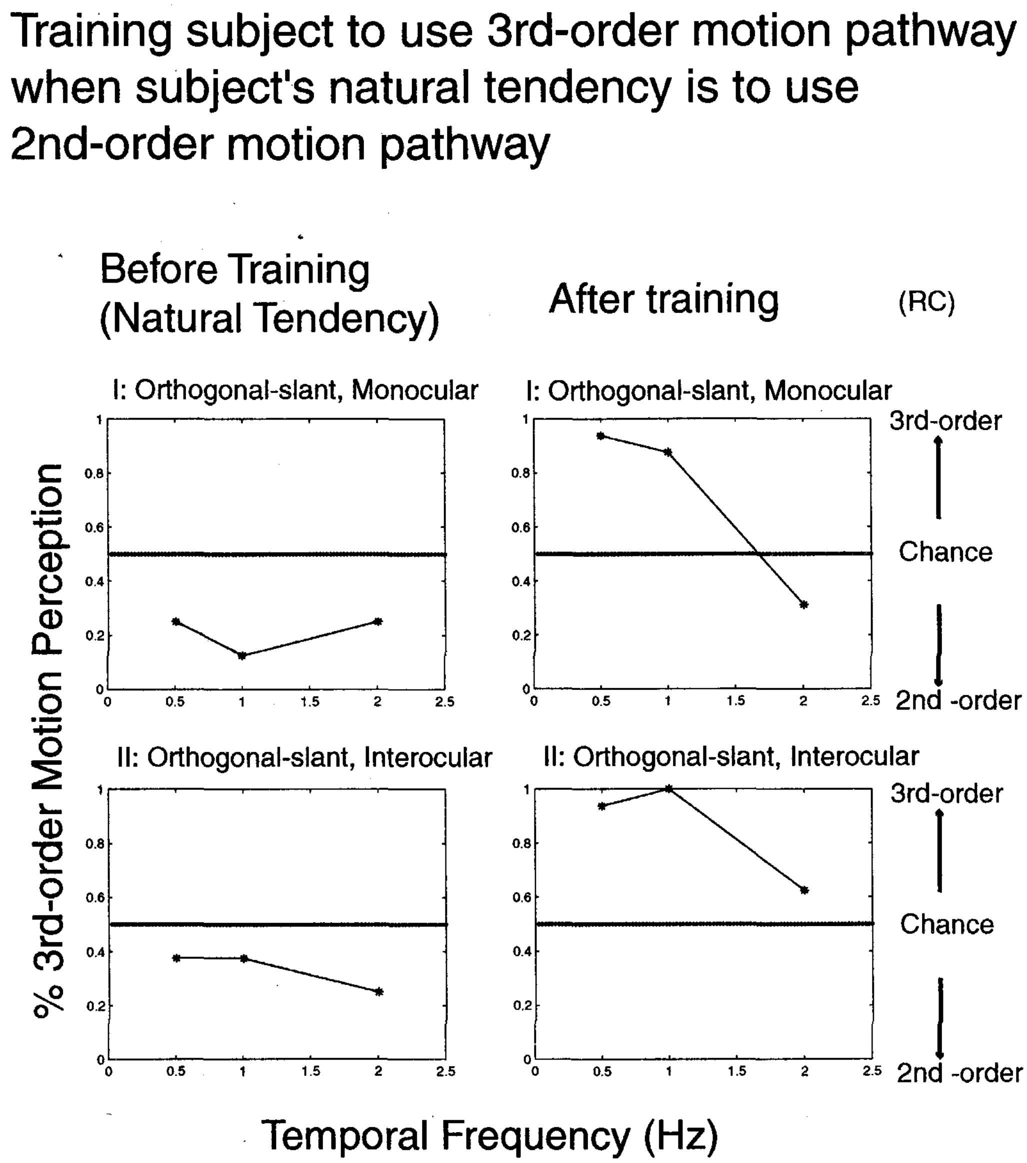


Figure 5.5: Effect of training. Perceived direction of motion before and after training with fading. Figure shows observers' motion perception of orthogonal-slant stimuli under monocular and interocular viewing conditions, before and after training. Subject: RC. The other observer behaved similarly.

5.6 Stimulus Parameters

After every observer was able to perceive the third-order direction in the orthogonal-slant conditions, and second-order direction in the same-slant conditions, other stimulus parameters were manipulated. Lu & Sperling (1995b) showed that first-, second- and third-order motion had corner frequencies, 12 Hz, 12 Hz and 3 Hz, respectively. They also showed that third-order motion was intrinsically binocular whereas first- and second-order motions were primarily monocular. Therefore, we predict that for temporal frequencies lower than about 3 Hz, both the second-order and third-order directions can be perceived, while for temporal frequencies higher than 3 Hz, only second-order direction can be perceived. We also predict that with monocular viewing, both the second-order and the third-order directions can be perceived, whereas under interocular condition, only the third-order direction can be perceived. It is reasonable to assume slant is irrelevant to the second-order motion system, whereas slant is a cue for figure (as opposed to ground) under the third-order computation. Hence, the third-order direction can be perceived in the orthogonal-slant stimuli but not in the same-slant stimuli, whereas the second-order direction can be perceived equally well for both the orthogonal-slant and the same-slant stimuli. Therefore, we studied a full range of temporal frequencies, with both monocular and interocular viewing conditions, and with both the orthogonal-slant and same-slant stimuli.

5.6.1 Procedures

Viewing Conditions

Two kinds of stimuli were used: orthogonal-slant (adjacent s and v patches have orthogonal slant, Fig. 5.1, and same-slant (adjacent s and v patches have same slant, Fig. 5.6). There were two viewing conditions. In monocular viewing, one of the two eyes, selected randomly on each trial, received the stimulus. Consecutive frames were shifted by 90°. In interocular viewing, motion displays were presented alternately to the two eyes. On a random half of the trials, the left eye received the first frame, otherwise, the right eye. Consecutive frames in each eye, individually, were shifted by 180°. This is counterphase flicker. Perceiving consistent apparent motion requires observers to combine information from the two eyes.

Experiments with four conditions were conducted: Two types of stimuli (same, orthogonal slant) times two viewing conditions (monocular, interocular)(Fig. 5.8). Blocks of same slant stimuli were conducted separately from blocks of orthogonal slant stimuli. Within a block, monocular and interocular conditions were randomly interleaved, with equal numbers of monocular and interocular trials.

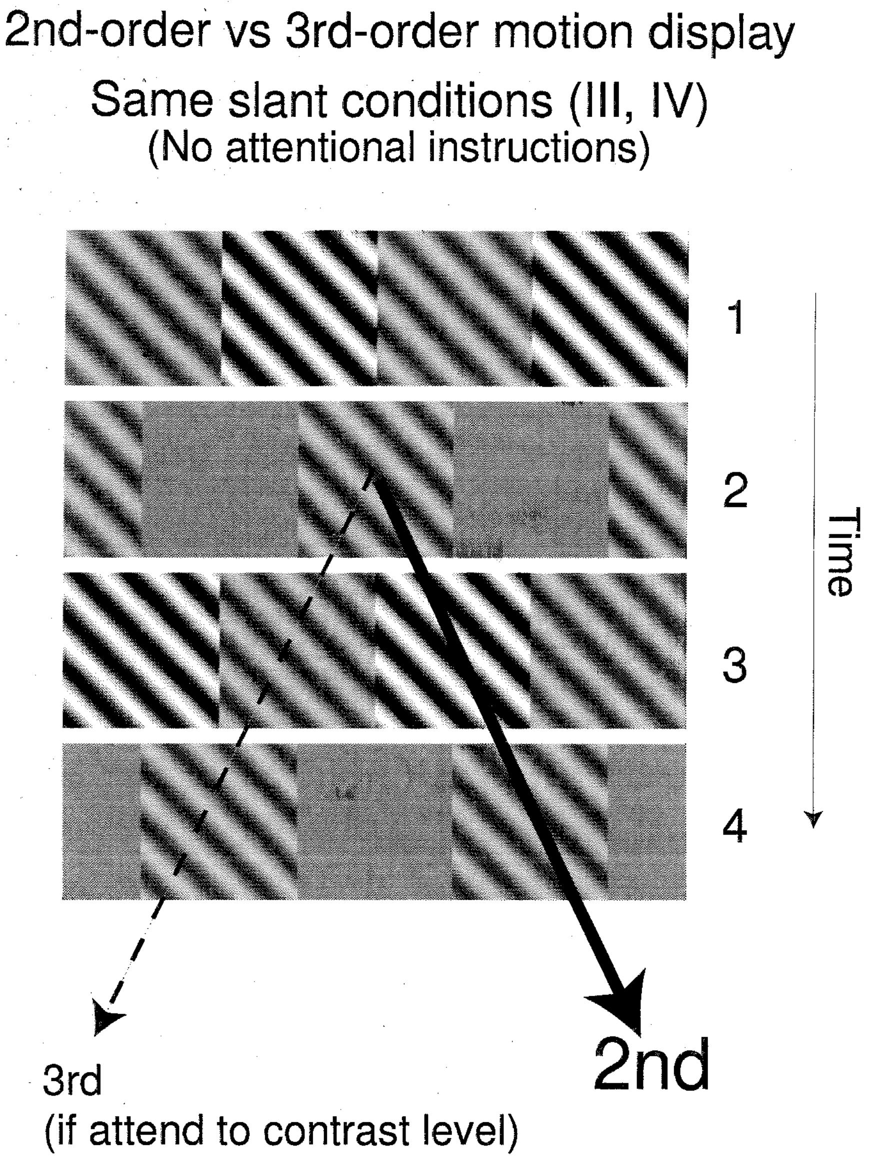


Figure 5.6: Motion stimulus for the same-slant conditions (conditions III and IV). Four frames of a longer motion sequence of texture-defined motion stimulus are shown, each consists of a strip with eight alternating patches, s (standard) and v(variable). Patches s and v contain grating slants at the same direction, either 45° or −45°. The slant directions are consistent throughout a motion sequence. s appears in every frame while v appears only in either odd or even numbered frames. Two motion directions, the second-order and the third-order can be perceived in each stimulus. The contrast of s, and v, are 0.2 and 0.4 respectively. Normally, the second-order direction is perceived. However, if observers were able to attend to a specific contrast level, the third-order direction might be perceived.

5.7 Changing the Direction of Apparent Motion from Third- to Second- Order: Attentional Instructions

In same-slant stimuli, two observers said they attended to and tracked low contrast patches. They perceived motion in the third-order direction. The left panels of Fig. 5.7 show data from one such observer. This viewing mode was changed by telling the observers to attend not to contrast but to slant, to give equal attention to all patches with the same slant, and to report global motion. The right panels of Fig. 5.7 show the results after these additional attentional instructions. The data of Fig. 5.7 show that instructions to attend to slant in same-slant stimuli changed the perceived direction of motion from the third- to second-order direction. It is interesting to note that for orthogonal slant stimuli, the instruction to attend to slant had the opposite effect, changing the perceived direction of motion from the second- to the third-order direction.

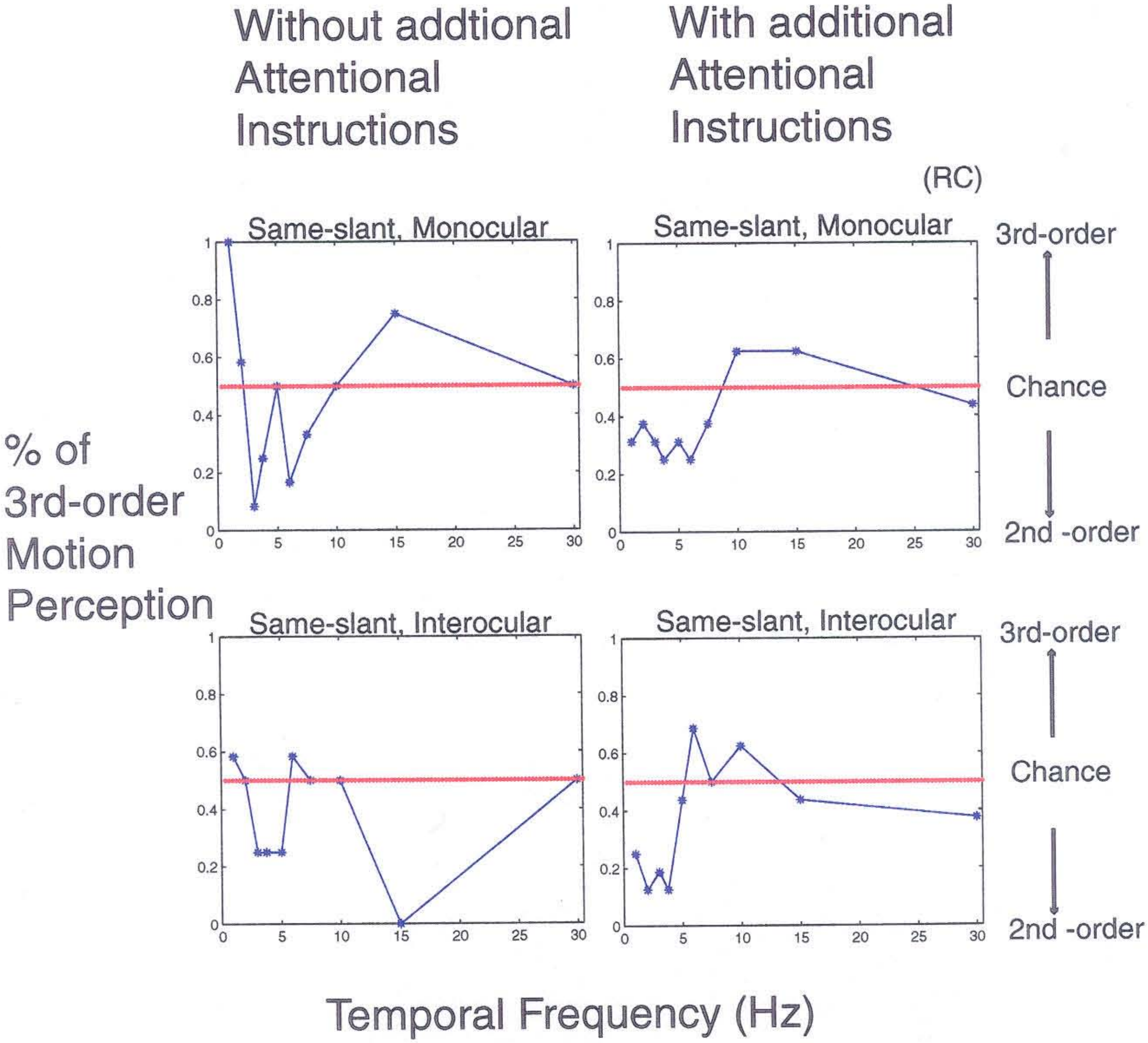


Figure 5.7: Perceived direction of motion before and after instructions to attend not to constrast but to slant. Subject RC.

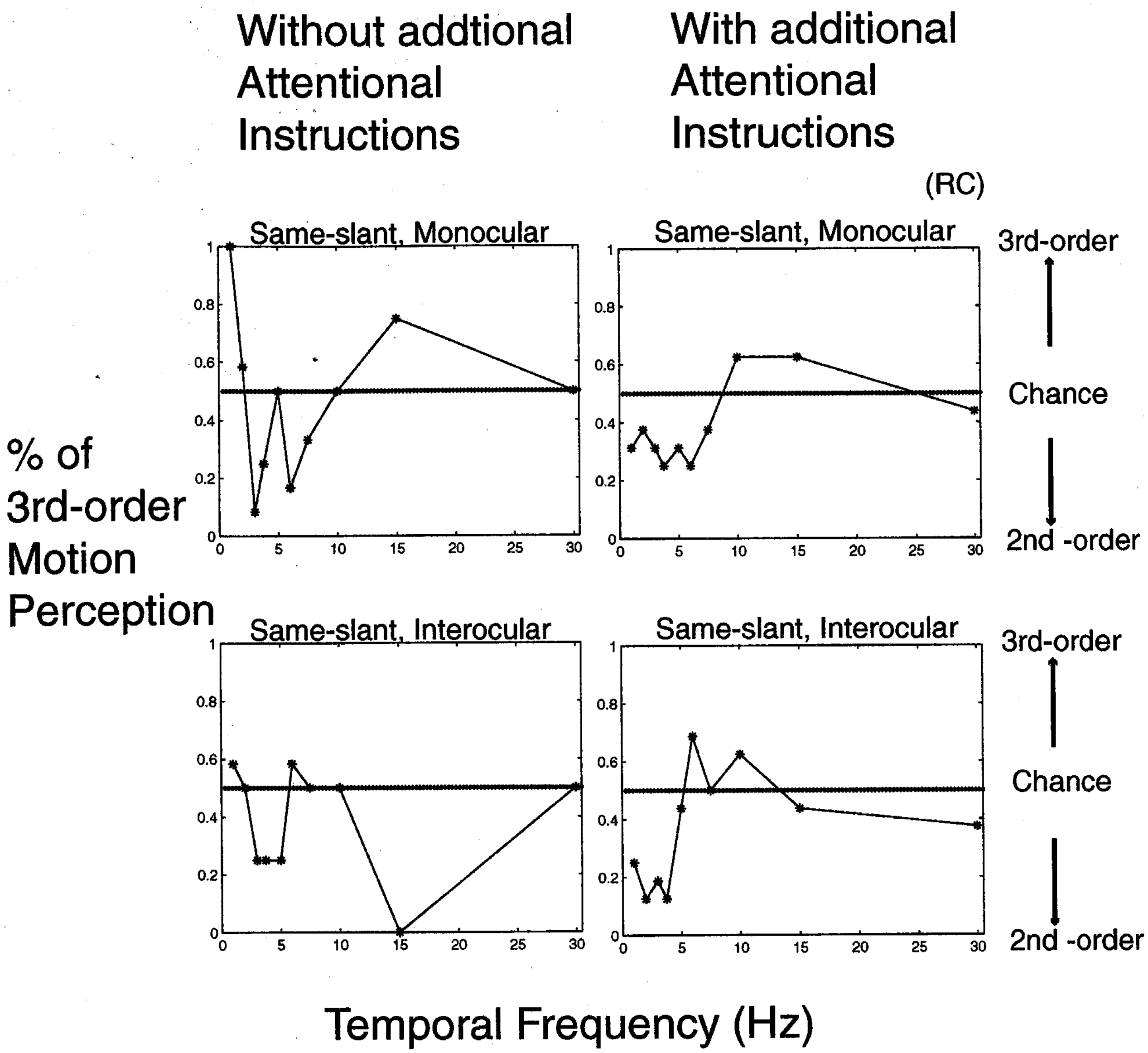


Figure 5.7: Perceived direction of motion before and after instructions to attend not to constrast but to slant. Subject RC.

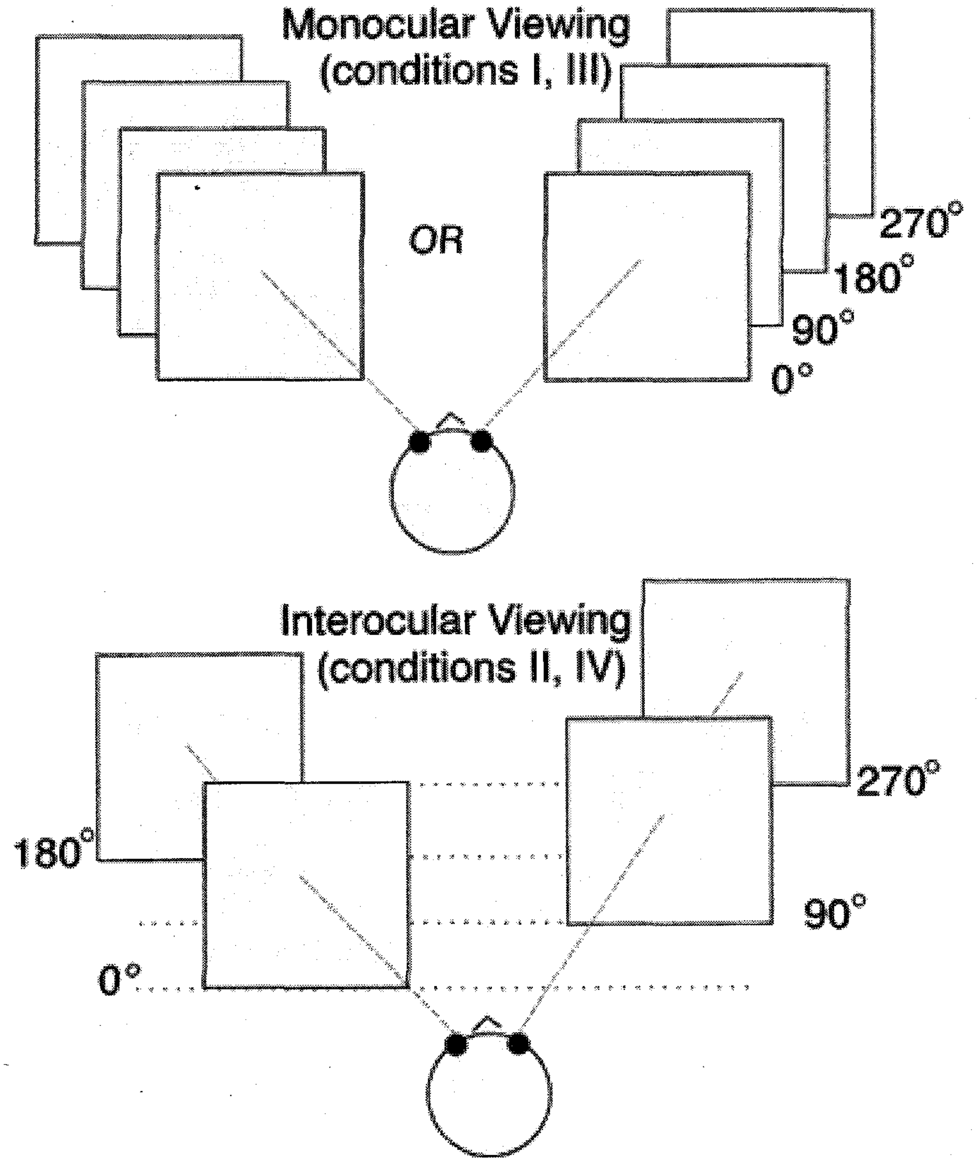


Figure 5.8: Illustration of monocular and interocular viewing conditions. In the monocular viewing condition, either one of the two eyes receives the entire stimulus. In interocular viewing, frames are presented alternatively to the two eyes. Consecutive frames in each eye are shifted 180^0. One eye alone receives no motion direction information, information must be combined from both eyes to yield a consistent motion direction.

5.7.1 Predictions

Orthogonal-slant grating patches, monocular viewing:

When observers attend to a particular grating slant, according to Lu & Sperling (1995a), they can perceive the third-order direction for temporal frequencies below 3 Hz. Between 3-12 Hz, observers can to perceive only second-order direction. Above 12 Hz, observers cannot perceive consistent motion. These predictions are shown in Fig. 5.9 (I).

Orthogonal-slant grating patches, interocular viewing:

According to Lu & Sperling (1995a), interocular viewing has no affect on the third-order computation, whereas it eliminates second-order motion perception. This gives us a way to eliminate the second-order direction from an ambiguous display with competing second- and third-order motion directions. The prediction is: observers perceive the third-order direction at low temporal frequencies ($< 3Hz$), and no consistent motion above 3 Hz.

Same-slant grating patches, monocular viewing:

The grating slant of patches s and v are the same. Therefore, by definition, observers pay equal attention to both grating slants. The prediction is: the third-order direction cannot be perceived, while the second-order direction is perceived for frequencies below 12 Hz.

Same-slant grating patches, interocular viewing:

The prediction is that performance will be at chance. Perception of the third-order direction is eliminated by the same-slant configuration; second-order motion perception is eliminated by interocular viewing.

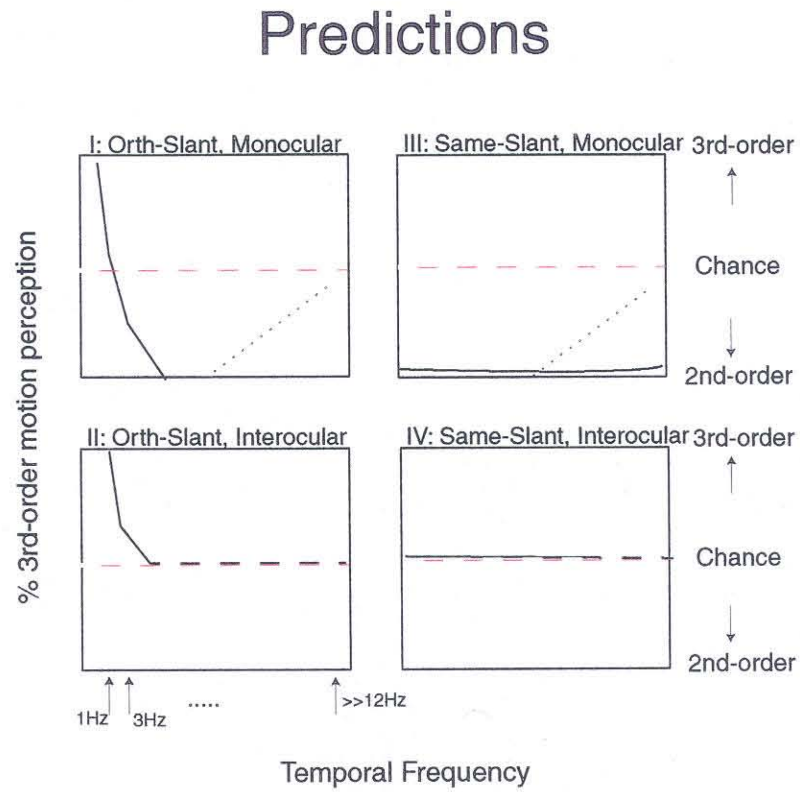


Figure 5.9: Predictions of perceived direction in four stimulus conditions. Predictions are based on three previous results: (i) Ho's (1998b) finding that two computations, a second- and a third-order motion computation are involved in the perception of texture-defined motion stimuli; (ii) the measurements by Lu & Sperling (1995b) of the temporal tuning functions of the second- and third-order motion systems; and (iii) the differential resistance of the second- and third-order systems to interocular viewing.

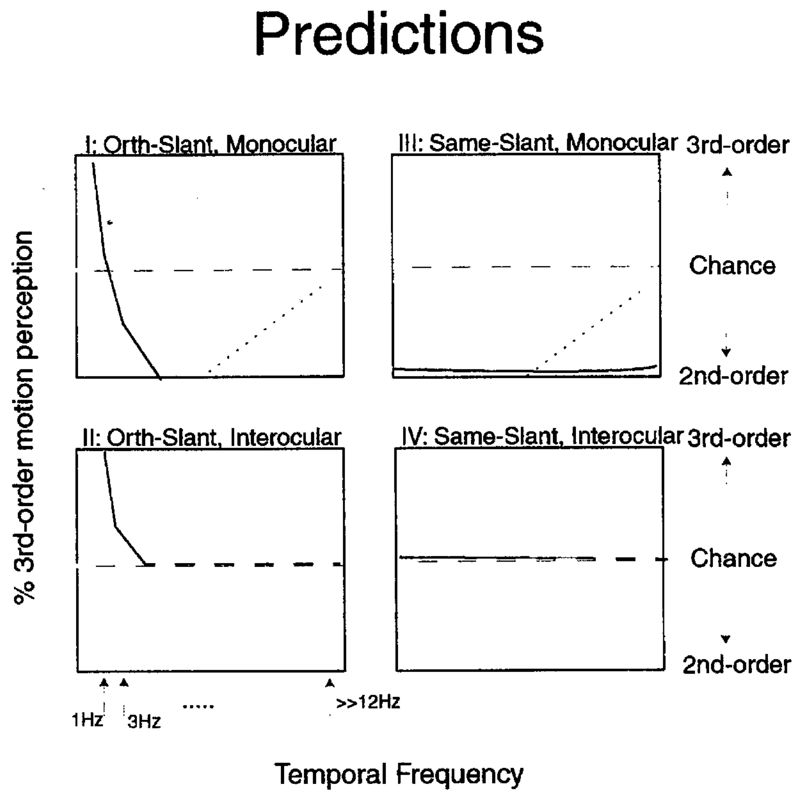


Figure 5.9: Predictions of perceived direction in four stimulus conditions. Predictions are based on three previous results: (i) Ho's (1998b) finding that two computations, a second- and a third-order motion computation are involved in the perception of texture-defined motion stimuli; (ii) the measurements by Lu & Sperling (1995b) of the temporal tuning functions of the second- and third-order motion systems; and (iii) the differential resistance of the second- and third-order systems to interocular viewing.

5.7.2 Results

Fig. 5.10 shows the combined data from five observers. Fig. 5.11 shows combined medium- and high-confidence reports of the second-order direction and the third-order direction, and the total low-confidence report (random guessing). When performance is close to chance in Fig. 5.10, data in Fig. 5.11 distinguishes between the case when observers guessed *randomly and the case when observers perceived similar amount of clear motion in each* direction.

Condition I (orthogonal-slant, monocular viewing). At frequencies below 3Hz, the perceived direction was mainly third-order. As frequency increased above 3Hz, the perception started switching to the second-order motion direction. At higher frequencies (5-15Hz), the perceived direction was mainly second-order. At temporal frequencies of 15-30Hz, performance fell to chance. From Fig. 5.10, at 3.75, 10, and 15Hz, the performance was quite close to chance (about 25% in the second-order direction). However, from Fig. 5.11, at 3.75Hz, there was about 50% of report of the second-order direction, and about 35% of report of the third-order direction. The fraction of random guessing was very low (about 15%). Therefore, at 3.75Hz, there were clear motion perceptions, sometimes in the second-order direction, sometimes in the third-order direction. At 10 and 15Hz, the fraction of reports of the second-order direction were about 15% and 10%, and the report of the third-order direction was about 25% and 15%. The fraction of random guessings were about 38% and 50%. Therefore, there was less clear perception of motion at 10Hz and 15Hz than at

3.75Hz, though the overall report of the third-order direction were about the same.

Condition II (orthogonal-slant, interocular viewing). From Fig. 5.10, relative to monocular viewing (Condition I), third-order perception was unaffected while reports of second-order motion direction decreased. Fig. 5.11 shows that the medium/high-confidence report of the third-order direction was not affected by changing viewing condition from monocular to interocular. The medium/high confidence report of the second-order direction was slightly decreased. At 15Hz, though, the total report of the third-order direction was about chance as shown in Fig. 5.10, about 30% of the total reports were medium-high confidence reports of second-order direction, and about 20% of the total reports were medium/high-confidence reports of the third-order direction, as shown in Fig. 5.11. Hence the apparent chance level at 15Hz in Fig. 5.10 was because observers perceived clear motion in one direction sometimes, and clear motion in another direction in the other times. The total amount of random guesses (Low confidence) slightly increased for temporal frequencies 7.5 Hz-10 Hz, but were otherwise similar to Condition I.

Condition III (same-slant, monocular viewing). Fig. 5.10 shows that relative to the orthogonal slant condition, reports of third-order direction greatly decreased. At frequencies below 3 Hz, observers perceived only the second-order direction. Where, in Condition I, the second-order direction was perceived, it was also perceived in Condition III. Fig. 5.11 shows that relative to orthogonal-slant, monocular viewing condition, same-slant configuration decreased the medium/high-confidence report of the third-order direction and increased the medium/high-confidence report of the second-order direction throughout the whole frequency region 1 - 30 Hz. Curiously, the total amount of random guessing was lowered *relative* to the orthogonal-slant condition.

Condition IV (same-slant, interocular viewing). Fig. 5.10 shows that the second-order direction perception was decreased above about 5 Hz as compared to the data in same-slant, monocular viewing. Comparing with data in Condition II (orthogonal-slant, interocular viewing), reports of third-order motion direction went from greater than 90% to less than 10% in the same slant viewing. The report of the third-order direction was close to chance for temporal frequencies of 7.5-15Hz. Fig. 5.11 also confirms the above phenomena. For temporal frequencies of 7.5-15Hz, the medium/high-confidence reports of the second-order direction, and of the third-order direction were about 35%-40%. This shows that the *apparent chance performance of motion detection* was not because observers could not perceive any motion, but rather that they perceived motion in the second-order direction sometimes, and in the third-order direction in the other times.

In summary, the perception of the second-order direction is indifferent to same-slant versus orthogonal slant stimuli; the perception of the third-order direction is eliminated in the same-slant stimuli. These results agree with the prediction that the third-order motion computation would be sensitive to the same-versus-orthogonal slant manipulation whereas the second-order motion computation would not be. Comparing results of monocular vs interocular viewing conditions, perception in the third-order direction showed no loss, or maybe even a little bit enhanced under interocular viewing condition. whereas perception in the second-order direction was decreased by interocular viewing condition, especially in the higher-frequency regions. However, in the low-frequency region, perception in the second-order direction was not affected.

Fig. 5.12 shows results from two observers for comparison with the combined results. The general trends are qualitatively the same.

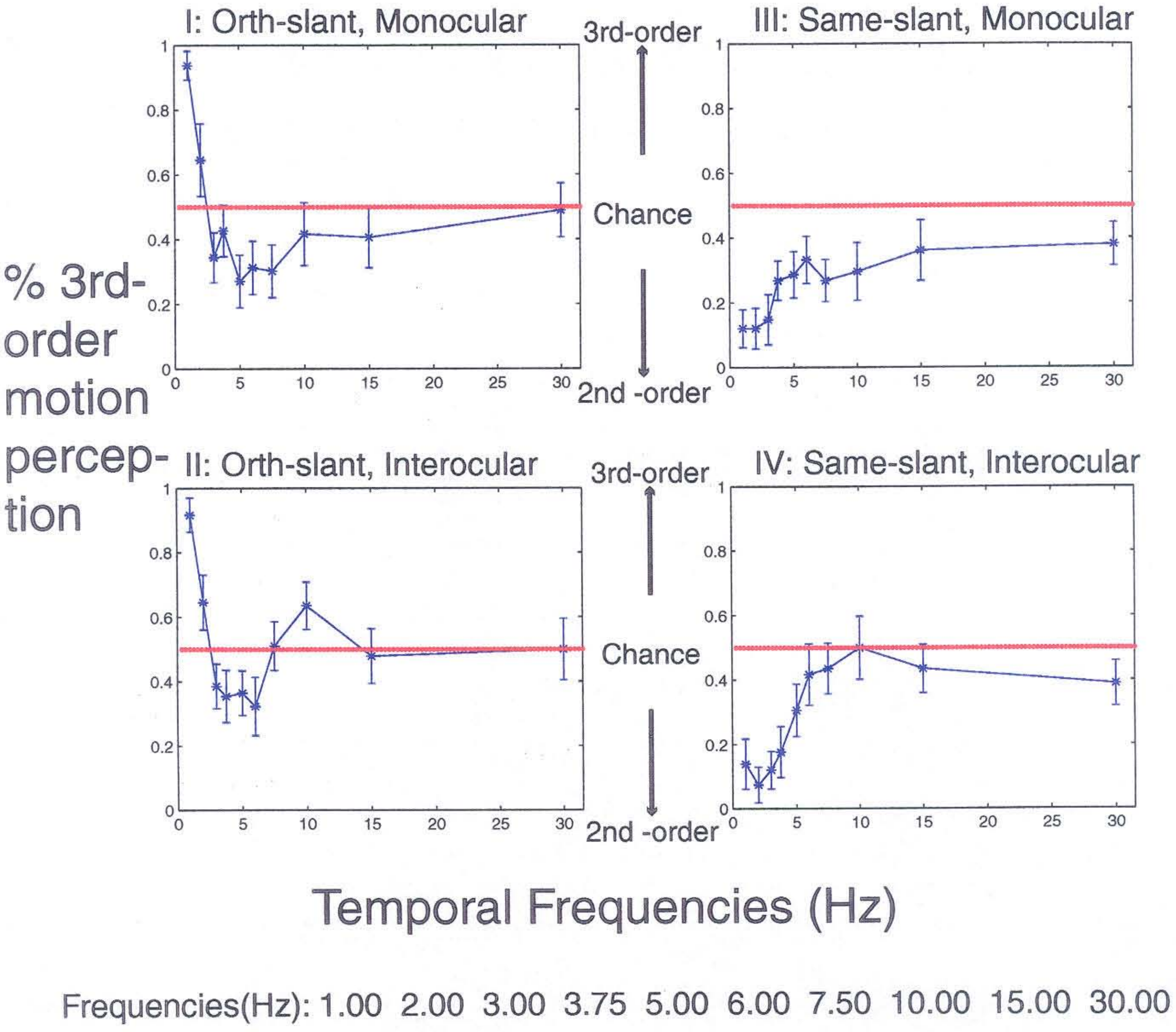


Figure 5.10: Combined results from the second- versus third-order ambiguous motion stimulus: Perceived direction. Combined data from 5 observers after black-white calibration and overall contrast reduction showing motion perceptions under four experimental conditions. Subjects: CH, TL, TC, RC, JG.

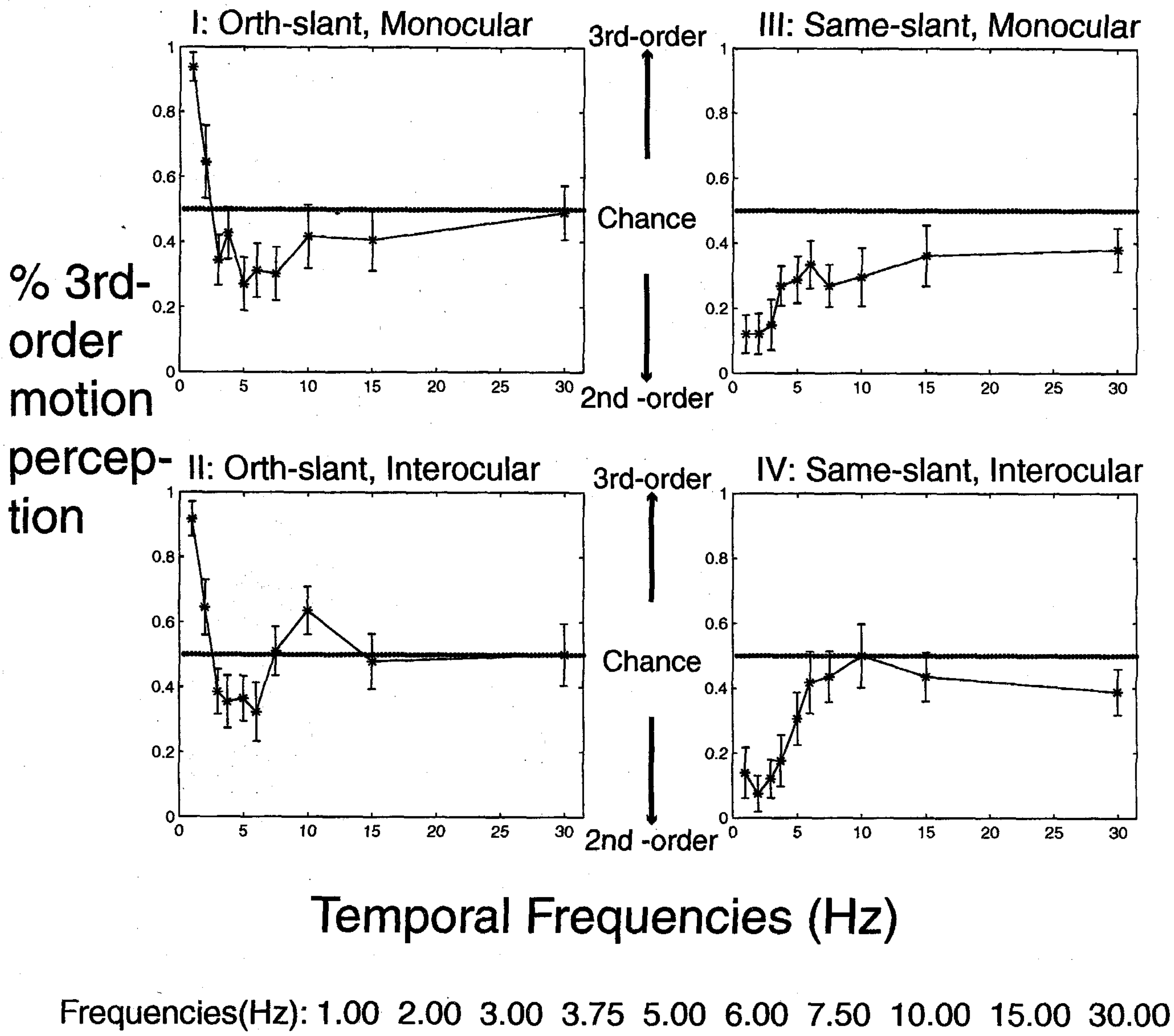


Figure 5.10: Combined results from the second- versus third-order ambiguous motion stimulus: Perceived direction. Combined data from 5 observers after black-white calibration and overall contrast reduction showing motion perceptions under four experimental conditions. Subjects: CH, TL, TC, RC, JG.

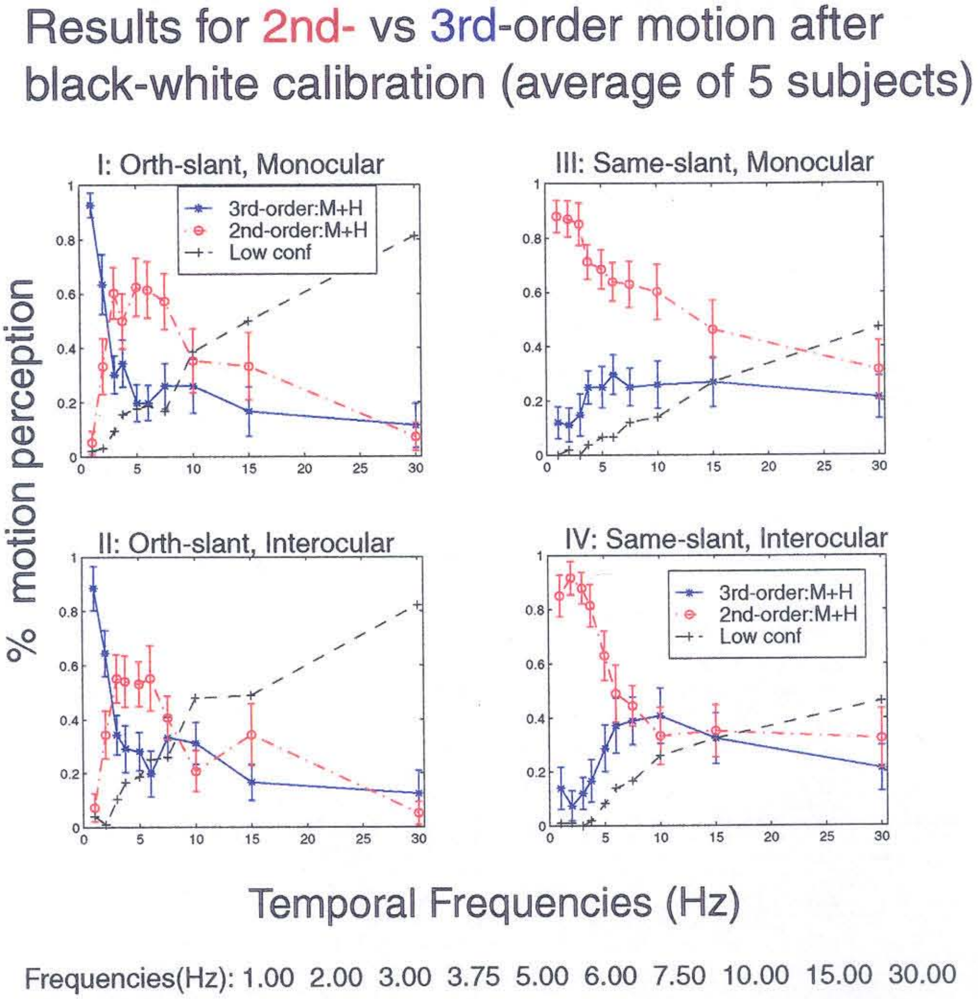


Figure 5.11: Combined results from the second- versus third-order ambiguous motion stimulus: Medium plus high confidence reports. Combined data from 5 observers after black-white calibration and overall contrast reduction, showing perception in the second- and third-order directions, with medium or high confidences. Total low confidence fraction is also shown to indicate the amount of random guesses. Subjects: CH, TL, TC, RC, JG.

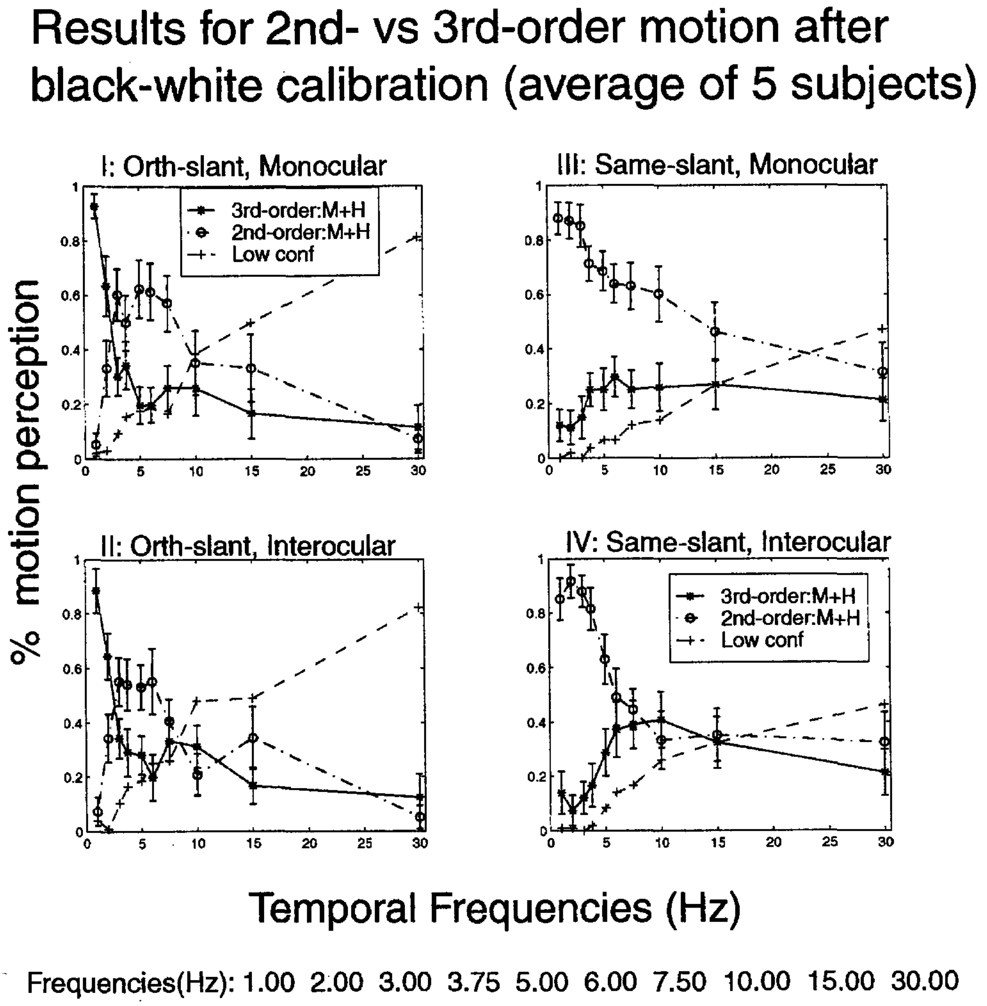


Figure 5.11: Combined results from the second- versus third-order ambiguous motion stimulus: Medium plus high confidence reports. Combined data from 5 observers after black-white calibration and overall contrast reduction, showing perception in the second- and third-order directions, with medium or high confidences. Total low confidence fraction is also shown to indicate the amount of random guesses. Subjects: CH, TL, TC, RC, JG.

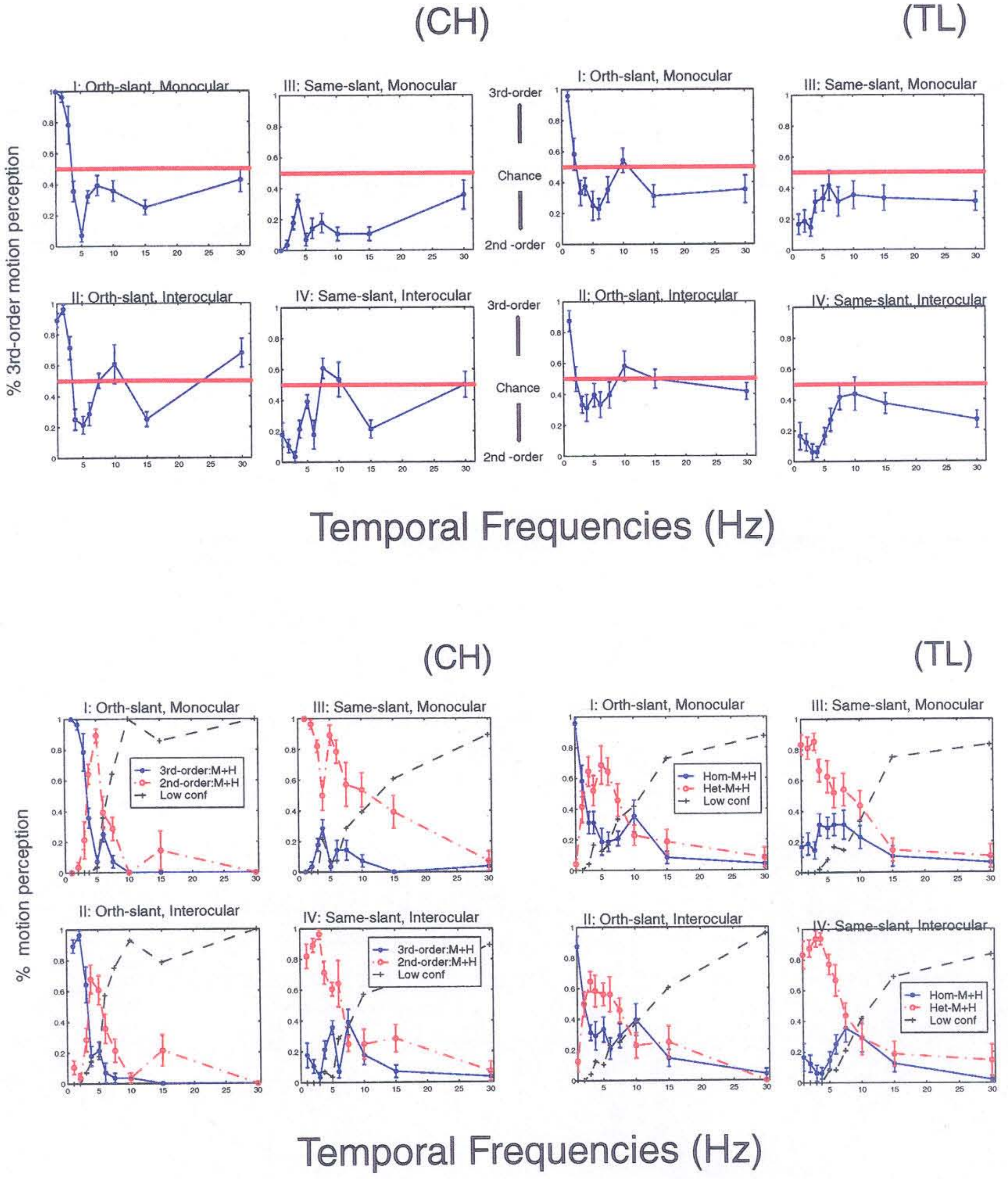


Figure 5.12: Individual data from the second- versus third-order ambiguous motion stimulus. **Upper row** Fraction of the third-order motion perception as temporal frequency varies. **Lower row** Motion perception of the third-order (medium and high confidences) and second-order (medium and high confidences). Total amount of low confidence (random guesses are also indicated. Three other observers behaved similarly. Subjects: CH, TL, TC, RC, JG.

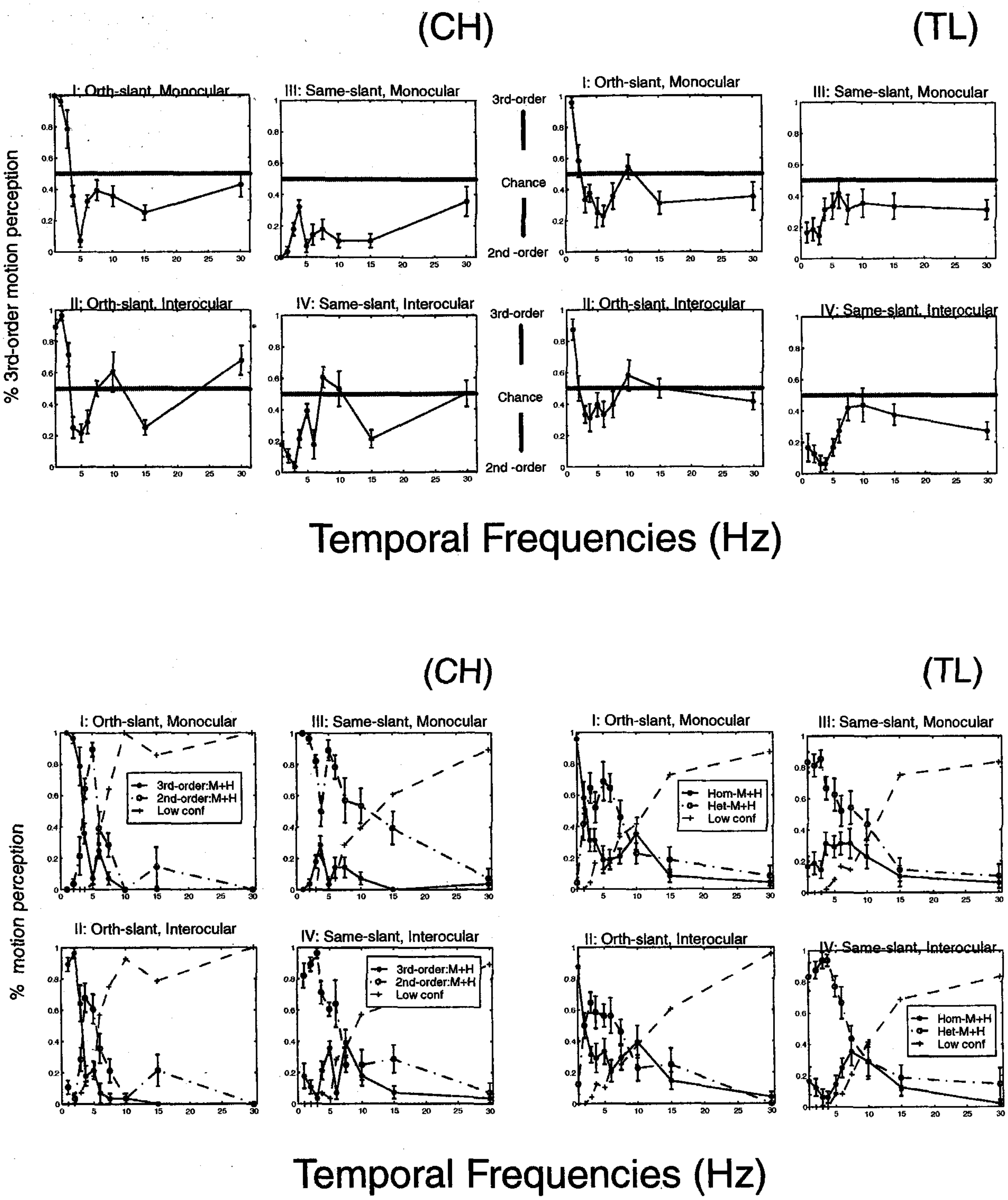


Figure 5.12: Individual data from the second- versus third-order ambiguous motion stimulus. **Upper row** Fraction of the third-order motion perception as temporal frequency varies. **Lower row** Motion perception of the third-order (medium and high confidences) and second-order (medium and high confidences). Total amount of low confidence (random guesses are also indicated. Three other observers behaved similarly. Subjects: CH, TL, TC, RC, JG.

5.8 Conclusion

5.8.1 Significant variables for determining motion direction in the same sequence of motion frames

Initial tendencies, attentional instructions, training, temporal frequencies, and viewing conditions are all factors for observers' selection of the second-order or the third-order motion directions. In general, the selection of second-order direction indicates computation of motion by a second-order motion mechanism, and the selection of the third-order direction indicates computation of motion by a third-order motion mechanism.

Given the same texture-defined motion stimulus, some observers showed initial preferences towards using a second-order motion system, and others towards using a third-order motion system. Attentional instructions induced the perceptions of some observers who initially preferred second-order system to use the third-order system. Referring to the model in Fig. 4.5, the attentional instruction increases the relative weight of the third-order motion system. Training further enhances this change of weight and improves the third-order motion perception.

For the same sequence of motion frames, at temporal frequencies above about 3 Hz, the third-order motion computation fails, and the second-order motion system becomes the only system for motion perception. Both the second-order and the third-order motion can be perceived well under monocular viewing condition. Under interocular viewing condition, the third-order suffers no loss while second-order motion is significantly decreased in some frequency region. Hence interocular viewing condition favors the third-order motion perception. Second-order motion follows Werkhoven *et al*'s (1993) energy model and is not sensitive to slant. Third-order motion direction is sensitive to slant. This is consistent with the properties of the third-order motion system as a pattern-tracking motion systems which computes spatio-temporal variations of selectively attended patterns where here, the pattern is slant.

5.8.2 Procedures for inducing different perceived directions in the same sequence of frames

For observers who perceive second-order motion initially, to change perception to the third-order direction, one can first try giving attentional instructions to change their perceived direction from the second-order to the third-order at low temporal frequencies. If this does not work, one can train observers by presenting observers with slower motion displays or *motion stimuli with* patches *s* colored. At medium temporal frequencies, where these observers can perceive second-order motion very well, one can present observers with motion stimuli under interocular viewing condition to reduce second-order perception.

For observers who perceive the third-order direction initially, to change perception to the second-order direction, one can increase temporal frequency to eliminate the third-order component and force observers to use second-order system. One can also give attentional instructions to induce observers to change the systems they use.

Chapter 6 Selecting First- and Third-Order Motion Systems

6.1 Introduction

It has been shown in the previous chapter that a pattern-tracking computation is an addition to, not a replacement of the motion-energy computation for texture-defined motion. Here we deal with well established first-order, luminance-defined motion stimuli. It has been amply demonstrated that luminance defined motion stimuli can be perceived by a low-level motion detectors. If there were only one system for the perception of motion in luminance-defined and texture-defined motion stimuli, the motion mechanism for first-order would be the motion energy mechanism. However, as will be shown here, there is also a competing pattern-tracking, third-order motion mechanism involved in the perception of motion of luminance-defined stimuli.

In this chapter, by studying the behavior of competition of first- and third-order motions in an ambiguous 'first-order' motion stimulus, it will be shown that third-order motion system can be used for motion perception in a first-order motion stimulus. Factors previously shown to affect the selection of systems between second- and third-order motions will also be examined here to see if they affect the selection of first- and third-order systems. If they do, it will provide an analogy for the case of texture-defined motion stimuli, and support the hypothesis that there is more than one system for motion perception in the texture-defined stimuli, as opposed to the single mechanism models (Johnston *et al.*, 1992; Taub *et al.*, 1997; Seiffert & Cavanagh, 1998).

6.2 Stimulus for Testing First- vs Third-Order Systems

We see lots of luminance-defined motion every day. With most ordinary stimuli, the first-order direction and the third-order direction are the same. The stimulus here was designed so that the first- and third-order directions are opposite to each other. The stimulus is an ambiguous motion display, very similar to the texture-defined motion version, except that mean luminance,instead of contrast, is varied across patches. The stimulus for the orthogonal-slant conditions (Conditions I, II) is as shown in Fig. 6.1. Four frames of a longer motion sequence of luminous-defined motion display are shown, each consists of a strip with 8 alternating patches. To adopt a terminology similar to that used in the case of texture-defined motion, the patches are analogously named s (standard) and v (variable). The mean luminance of patches s is greater than that of patches v. The motion direction defined by luminance is called the first-order direction while the motion direction defined by grating slant is called the third-order direction. The mean luminance of s is L_o while the mean luminance of v is $1.015*L_o$. The contrast m of both types of patches is 0.3. The stimulus in the same-slant conditions (Conditions III, IV) is similar to the one Fig. 6.1 except that all patches have the same slant.

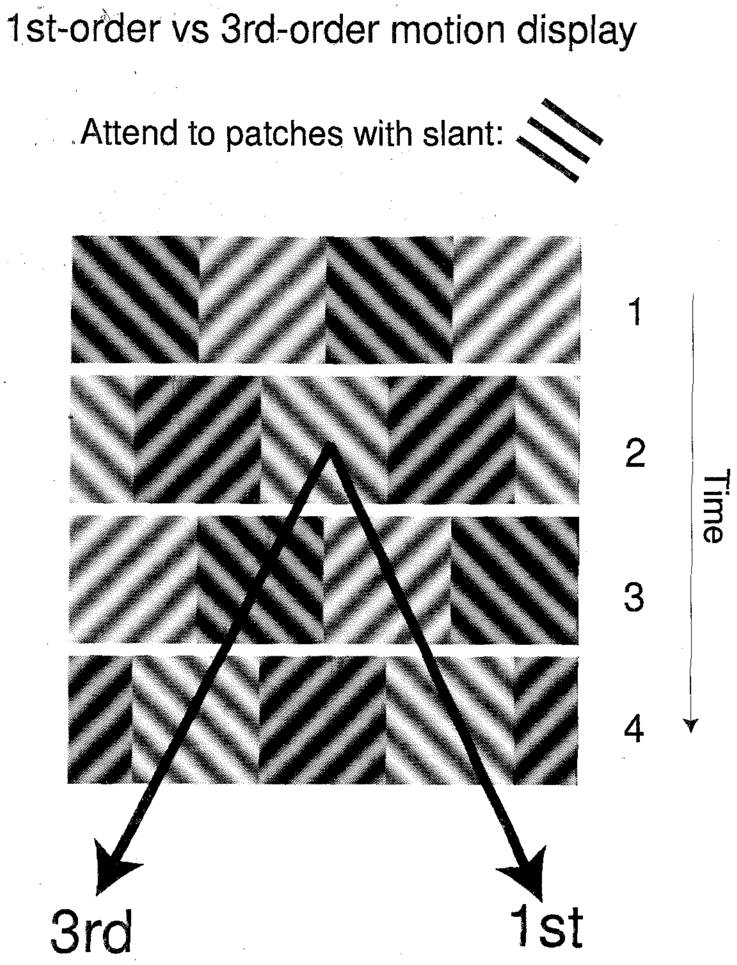


Figure 6.1: The luminance-defined motion stimulis. Four frames of a longer motion sequence of an ambiguous motion stimulus in which first- and third-order motions are in opposite direction. Each consists of a strip with eight alternating grating patches.

6.3 Initial Tendencies

As in the case of texture-defined, second- vs third-order motion stimuli, there were two initial tendencies for the same stimulus. Fig. 6.2 shows the initial tendencies for each observer at temporal frequency of 2 Hz, and with a luminance difference between patches s and v of 0.015 L_o. More observers could perceive third-order motion from the beginning than in the case of texture-defined motion stimuli. This might due to the fact that the observers previously had performed the tasks on texture-defined motion, hence the previous attentional instruction might carry over. This might also be due to the fact that there were no blank patches, which some observers found distracting. In the display for testing the second- vs third-order system, patches v existed only in the even-numbered frames, creating flickering that could attract involuntary attention.

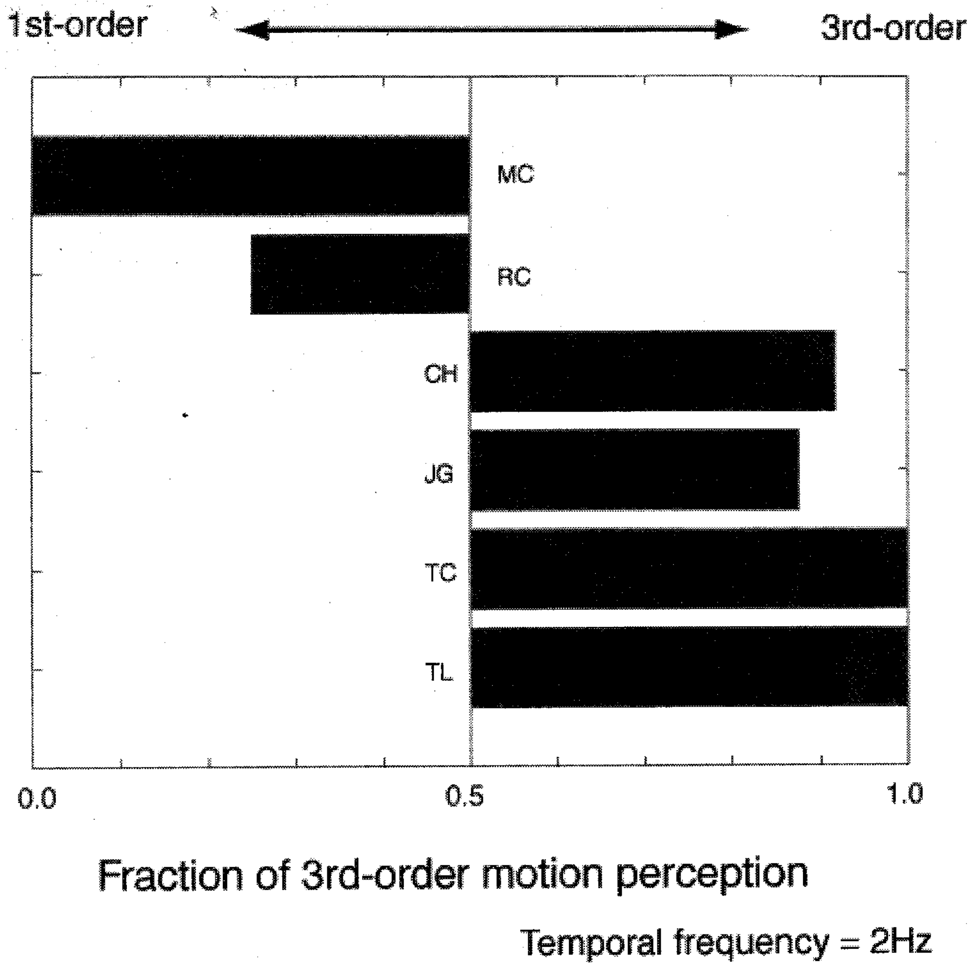


Figure 6.2: Direction of perceived of motion at temporal frequency of 2 Hz, before instruction to attend to a particular slant direction. Subject MC, RC, CH, JG, TC, TL.

6.4 Training

For observers who did not perceive motion in the third-order direction, attentional instructions and training with stimulus similar to that in Fig. 5.4 were used to help them perceive motion in the third-order direction. Fig. 6.3 illustrates effects of these training methods for one observer.

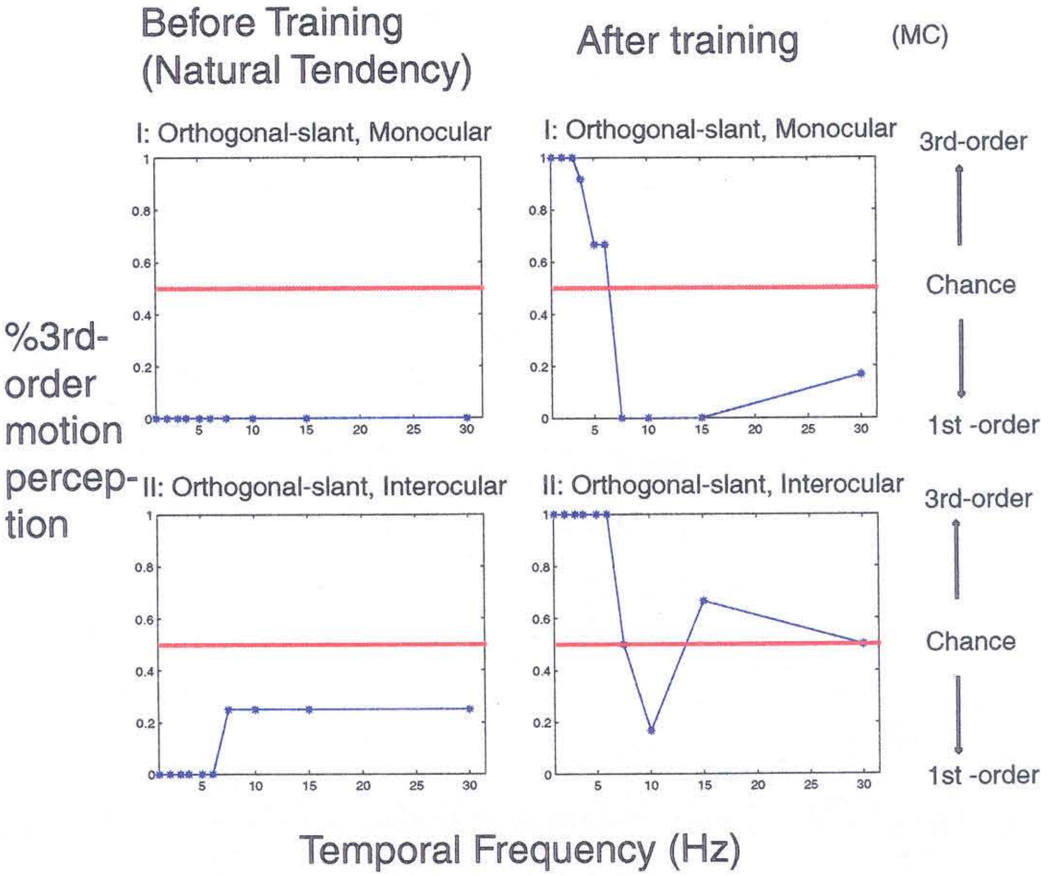


Figure 6.3: Luminance-defined ambiguous motion stimuli: Effect of training to attend to slant. Perceived direction of motion before and after instruction to attend to a particular slant direction. Before the attentional instruction, this observer's tendency was to perceive the first-order direction for all frequencies. After the attentional instruction, this observer perceived the third-order direction at temporal frequencies below 5Hz, and first-order direction at higher temporal frequencies. Subject: MC.

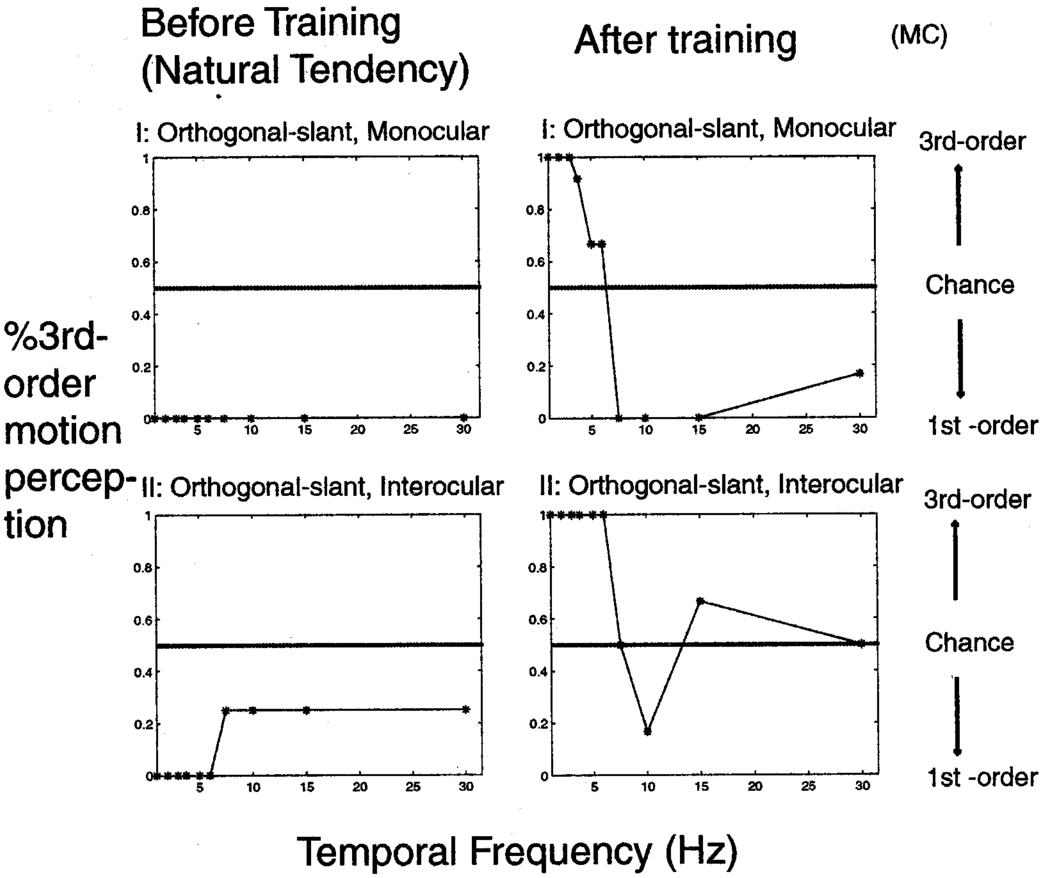


Figure 6.3: Luminance-defined ambiguous motion stimuli: Effect of training to attend to slant. Perceived direction of motion before and after instruction to attend to a particular slant direction. Before the attentional instruction, this observer's tendency was to perceive the first-order direction for all frequencies. After the attentional instruction, this observer perceived the third-order direction at temporal frequencies below 5Hz, and first-order direction at higher temporal frequencies. Subject: MC.

6.5 Procedures

Procedures are similar to those used for experiments for testing second- vs third-order systems in chapter 5 (section 5.6.1). Four conditions were studied: (I) orthogonal-slant, monocular viewing, (II) orthogonal-slant, interocular viewing, (III) same-slant, monocular viewing, and (IV) same-slant, interocular viewing.

6.6 Predictions

Since, according to Lu & Sperling (1995b), first-order motion has the same corner frequency as second-order motion, and it cannot be perceived under interocular viewing conditions as second-order motion cannot, and it does not take slant into account (since first-order motion is computed by a motion energy model similar to the motion energy model for the second-order motion), the predictions for the four conditions are exactly the same as those in the case of second- vs third-order competition, and can be summarized as in Fig. 5.9, with the first-order *direction replacing the role of the second-order direction.*

6.7 Results

Fig. 6.4 shows the combined results from five observers. Fig. 6.5 shows the combined medium- plus high-confidence reports of the first-order and third-order directions, and the total low-confidence reports (random guessing). When performance is close to 50-50 percent in each direction in Fig. 6.4, data in Fig. 6.5 distinguish between the case when observers guess randomly and the case when observers clearly perceive motion in opposite directions on different trials.

Condition I (orthogonal-slant, monocular viewing).

At low frequencies (< 5Hz), the perceived direction was mainly the third-order direction. At around 5Hz, the perception started switching to the first-order direction. At higher frequencies (5-15Hz), the perceived direction was mainly the first-order direction. As temporal frequency went to 30 Hz, performance fell towards chance.

To distinguish the nature of the near 50-50 performance at 5 Hz and 30 Hz, we use the data of Fig. 6.5. Fig. 6.5 shows that at 5Hz, 35% of the reports were medium/high-confidence in the third-order direction, and about 42% of the first-order direction. The percentage of random guessing was quite low (about 23%). Therefore, at 5Hz 77% of the the time, observers perceived motion clearly, sometimes in the first-order direction, other times in the third-order direction. At 30Hz, the combined medium and high-confidence reports of the first-order and the third-order directions were both less than 10%, while the fraction of random guessing was greater than 80%. Obviously at 30Hz, observers rarely perceived clear motion.

Condition II (orthogonal-slant, interocular viewing).

Fig. 6.4 shows that the reports of the third-order direction were similar in monocular and interocular viewing. This is true not only for the percent of third-order reports as a function of frequency but also the confidence with which these reports were made (as shown if Fig. 6.5. However, reports of the first-order direction were greatly decreased. The medium/high confidence report of the first-order direction was significantly reduced in the whole frequency range. Fig. 6.5 shows that at intermediate frequencies, observers perceive clear motion in opposite directions on different trials; at 30 Hz, more than 80% of the time, observers say they are just guessing.

Condition III (same-slant, monocular viewing).

Fig. 6.4 shows that with same-slant viewing, in the entire frequency range, reports of the third-order direction always occur on less than 40At frequencies up to 15 Hz, observers perceived the first-order direction with medium or high confidence more than 2/3 of the trials. In changing from orthogonal slant to same slant stimuli, the main effect on the results was the enormous reductions of reports of the third-order direction at frequencies below 5 Hz.

Fig. 6.5 shows that at 1-3.75Hz, 20% to 35% of the total reports were medium/high confidence reports of the third-order direction. So the 50-50 performance in Fig. 6.4 for frequencies 1-3.75 Hz was due to clear perception of motion in both the first- and the third-order directions on different trials. At 30Hz, Fig. 6.5 shows that there was about 45% of the reports were medium/high-confidence reports of the first-order direction. Compared

Fig. 6.4 shows that the reports of the third-order direction were similar in monocular and interocular viewing. This is true not only for the percent of third-order reports as a function of frequency but also the confidence with which these reports were made (as shown if Fig. 6.5. However, reports of the first-order direction were greatly decreased. The medium/high confidence report of the first-order direction was significantly reduced in the whole frequency range. Fig. 6.5 shows that at intermediate frequencies, observers perceive clear motion in opposite directions on different trials; at 30 Hz, more than 80% of the time, observers say they are just guessing.

Condition III (same-slant, monocular viewing).

Fig. 6.4 shows that with same-slant viewing, in the entire frequency range, reports of the third-order direction always occur on less than 40At frequencies up to 15 Hz, observers perceived the first-order direction with medium or high confidence more than 2/3 of the trials. In changing from orthogonal slant to same slant stimuli, the main effect on the results was the enormous reductions of reports of the third-order direction at frequencies below 5 Hz.

Fig. 6.5 shows that at 1-3.75Hz, 20% to 35% of the total reports were medium/high confidence reports of the third-order direction. So the 50-50 performance in Fig. 6.4 for frequencies 1-3.75 Hz was due to clear perception of motion in both the first- and the third-order directions on different trials. At 30Hz, Fig. 6.5 shows that there was about 45% of the reports were medium/high-confidence reports of the first-order direction. Compared to Condition I (orthogonal-slant, monocular viewing), at 30Hz, there were many more medium/high-confidence reports in Condition III although the total reports of the third-order direction were the same, as shown in Fig. 6.4.

Condition IV (same-slant, interocular viewing).

The percentage of reports of the first-order direction was greatly reduced in the frequency region 3-15 Hz as compared to the data in Condition III (same-slant, monocular viewing). Comparing with the data in Condition II (orthogonal-slant, interocular viewing), reports of the third-order direction were greatly reduced. In Fig. 6.4, reports of the third-order direction were close to 50-50 performance at frequencies from 3Hz to 30Hz. Fig. 6.5 shows that at 3Hz, about 42% of the reports were medium/high-confidence reports of the third-order direction, about 45% of the reports were medium/high-confidence reports of the first-order direction, only about 15% of the reports were low-confidence reports. Therefore, at 3Hz, observers perceived clear motion in one direction sometimes and in the other direction the other times. At 10Hz and 30Hz, the percent of low-confidence reports increased to about 38% and 55%, respectively.

Fig. 6.6 shows individual data from two observers for comparison with the combined results. The general trends are qualitatively the same although there are parametric differences between observers.

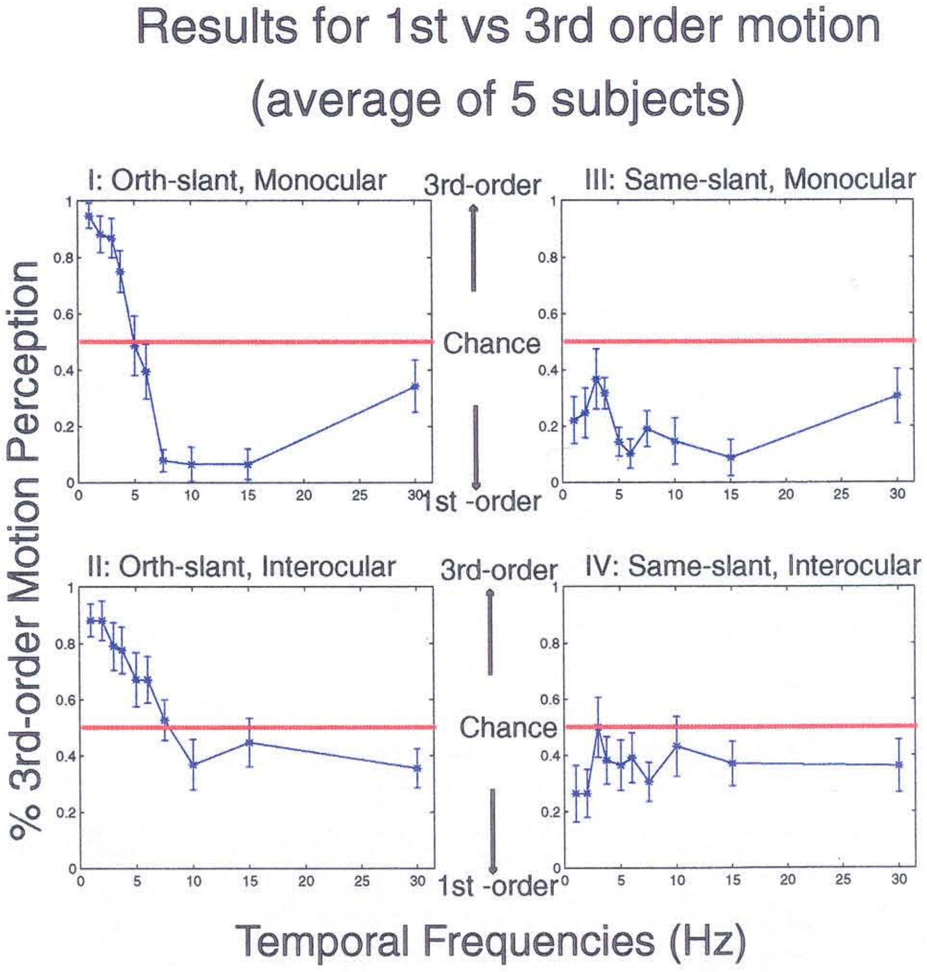


Figure 6.4: Combined results from the first- versus third-order ambiguous motion stimulus: Perceived direction. Data from 5 observers showing reports of directions under four experimental conditions. Subjects: CH, JG, MC, RC,TL.

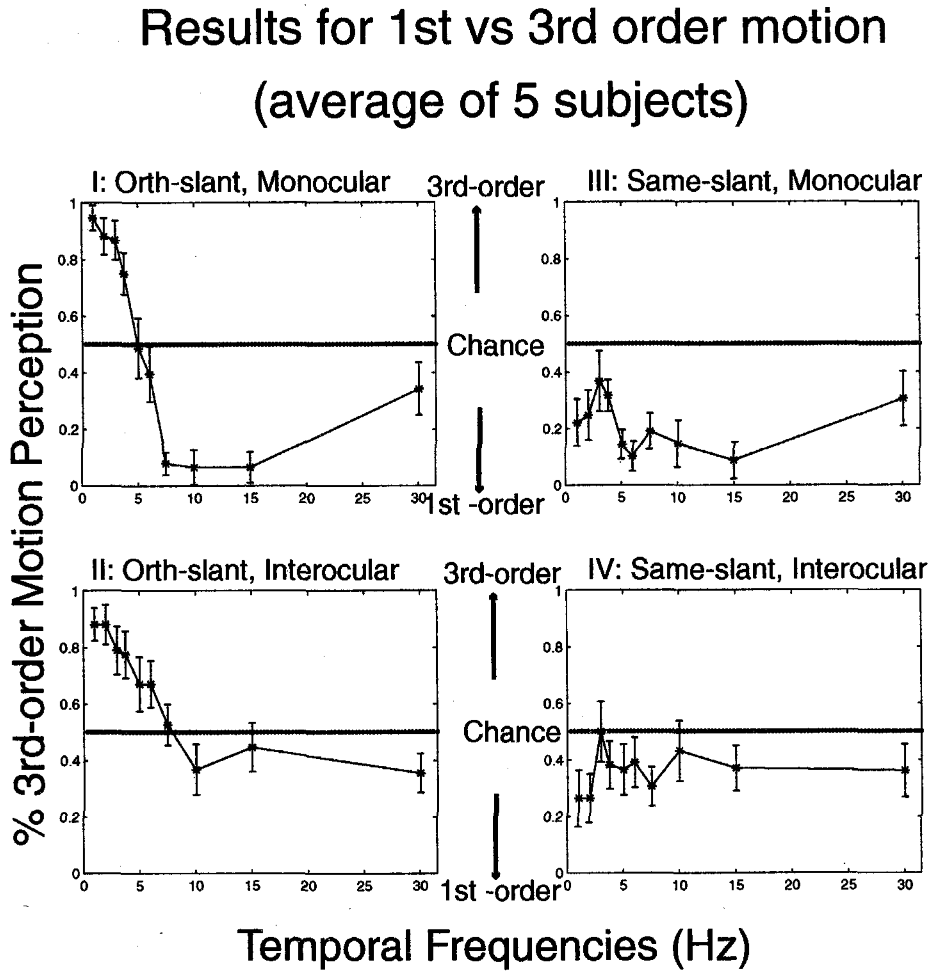


Figure 6.4: Combined results from the first- versus third-order ambiguous motion stimulus: Perceived direction. Data from 5 observers showing reports of directions under four experimental conditions. Subjects: CH, JG, MC, RC,TL.

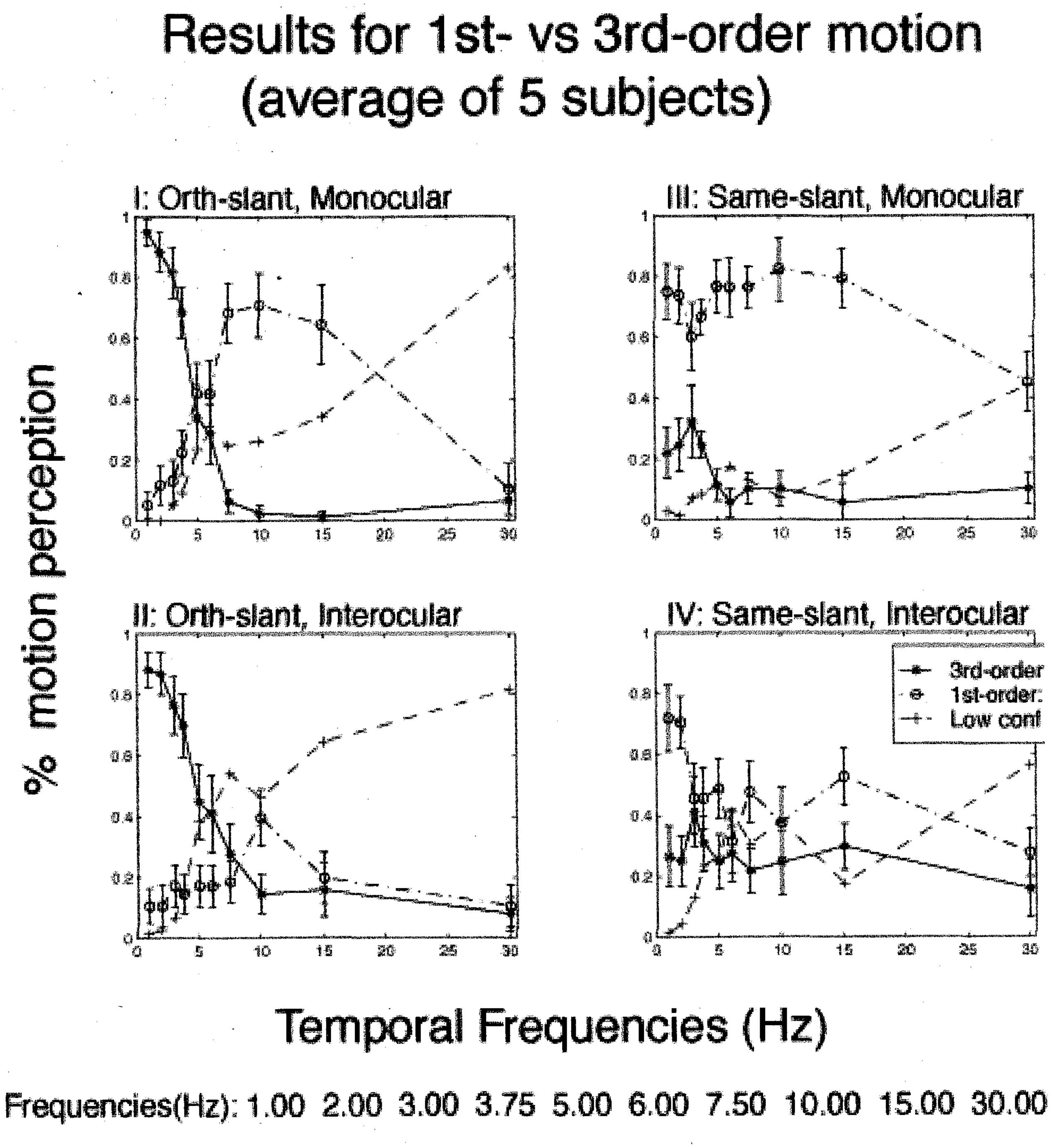


Figure 6.5: Combined results from the first- versus third-order ambiguous motion stimulus: Medium plus high confidence reports. Data from five observers showing reports of the first- and third-order directions with medium plus high confidences. Total low confidence fraction is also shown to indicate the amount of random guessing. Subjects: CH, JG, RC, TC, TL.

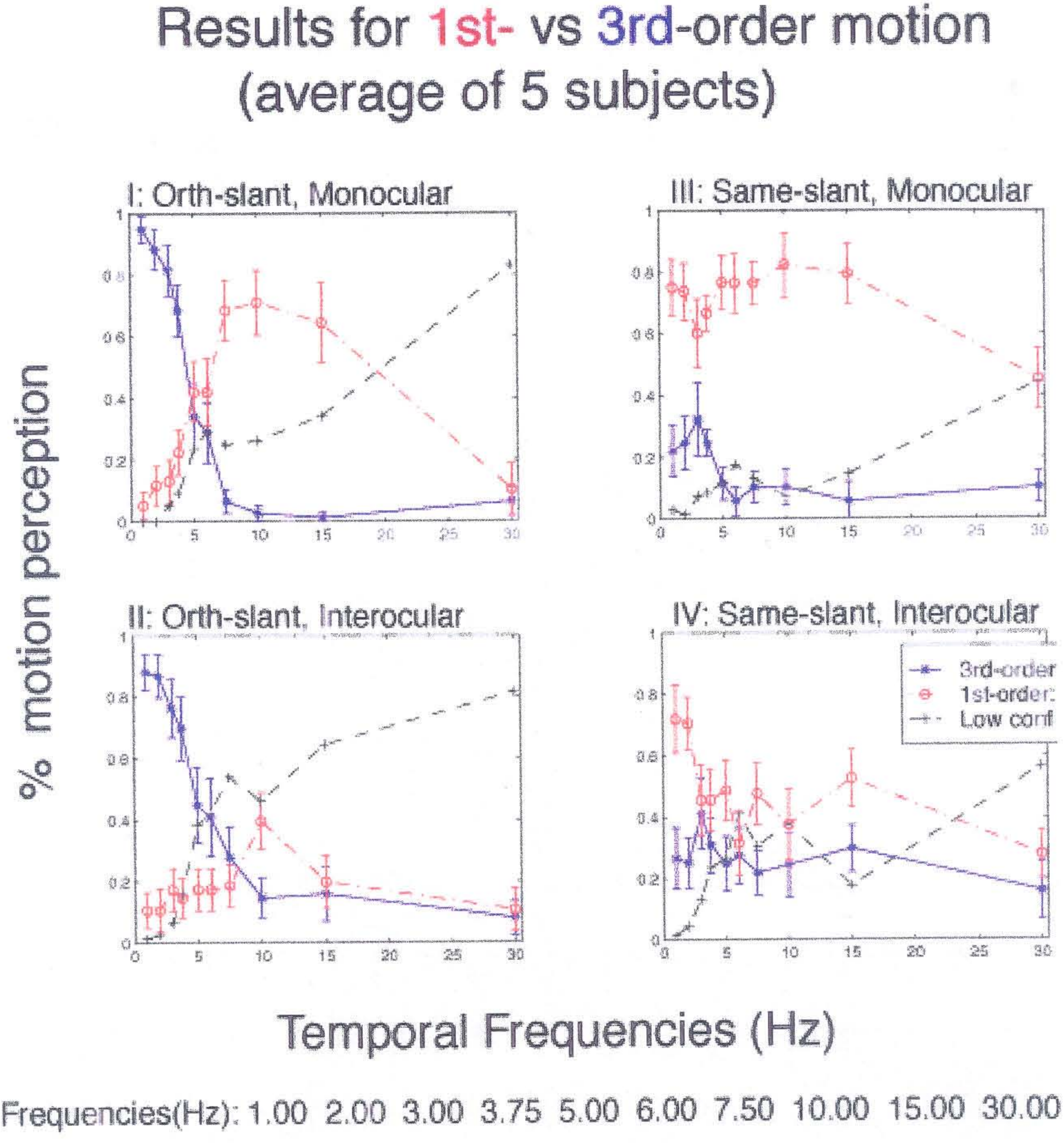


Figure 6.5: Combined results from the first- versus third-order ambiguous motion stimulus: Medium plus high confidence reports. Data from five observers showing reports of the first- and third-order directions with medium plus high confidences. Total low confidence fraction is also shown to indicate the amount of random guessing. Subjects: CH, JG, RC, TC, TL.

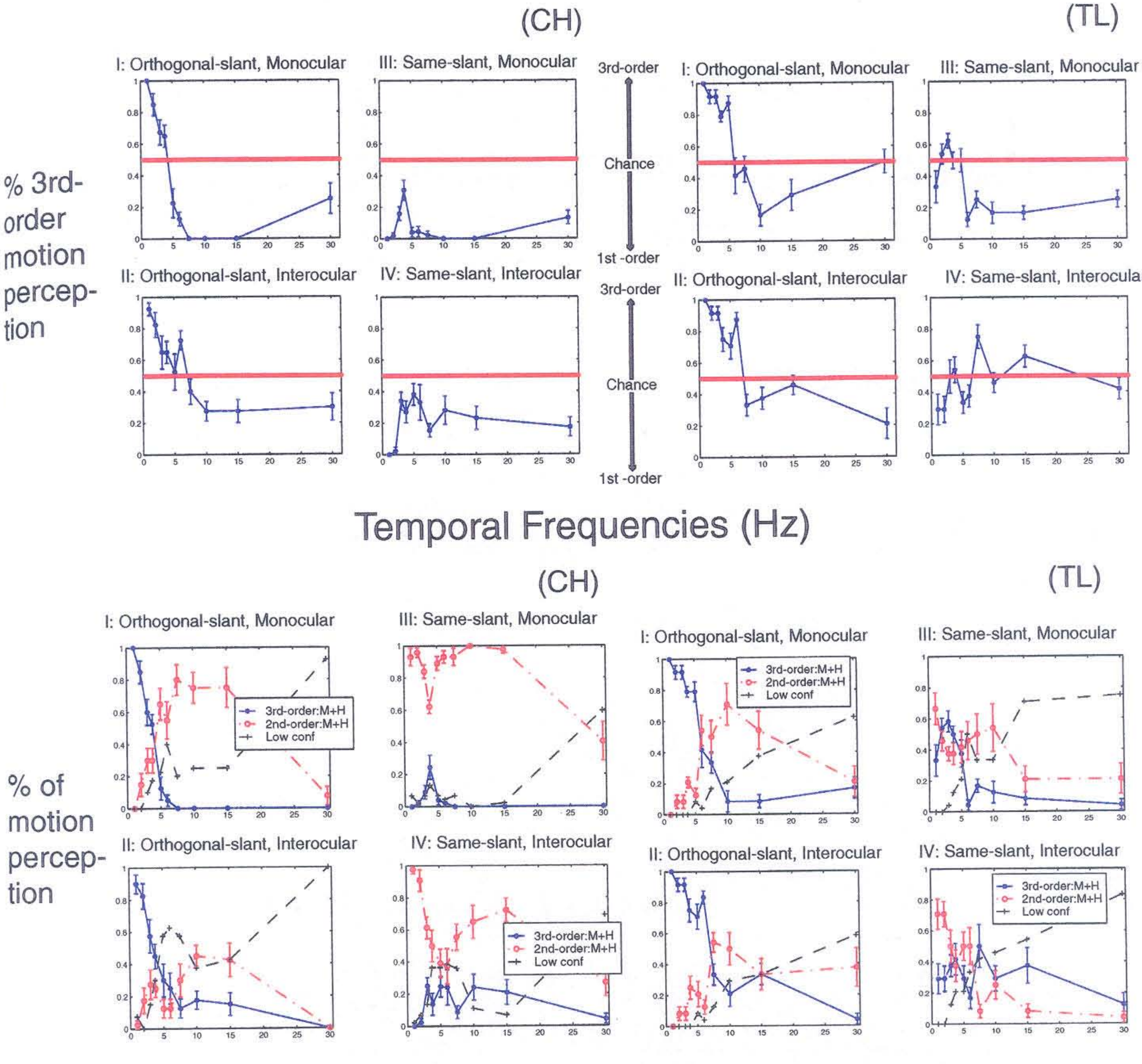


Figure 6.6: Individual data from the first- versus third-order ambiguous motion stimulus. **Upper row** Fraction of report of the third-order direction as temporal frequency varies. **Lower row** Reports of the third-order direction (medium and high confidences) and the first-order direction (medium and high confidences). Total amount of low confidence (random guesses are also indicated. Three other observers behaved similarly. Subjects: CH, JG, MC, RC, TL.

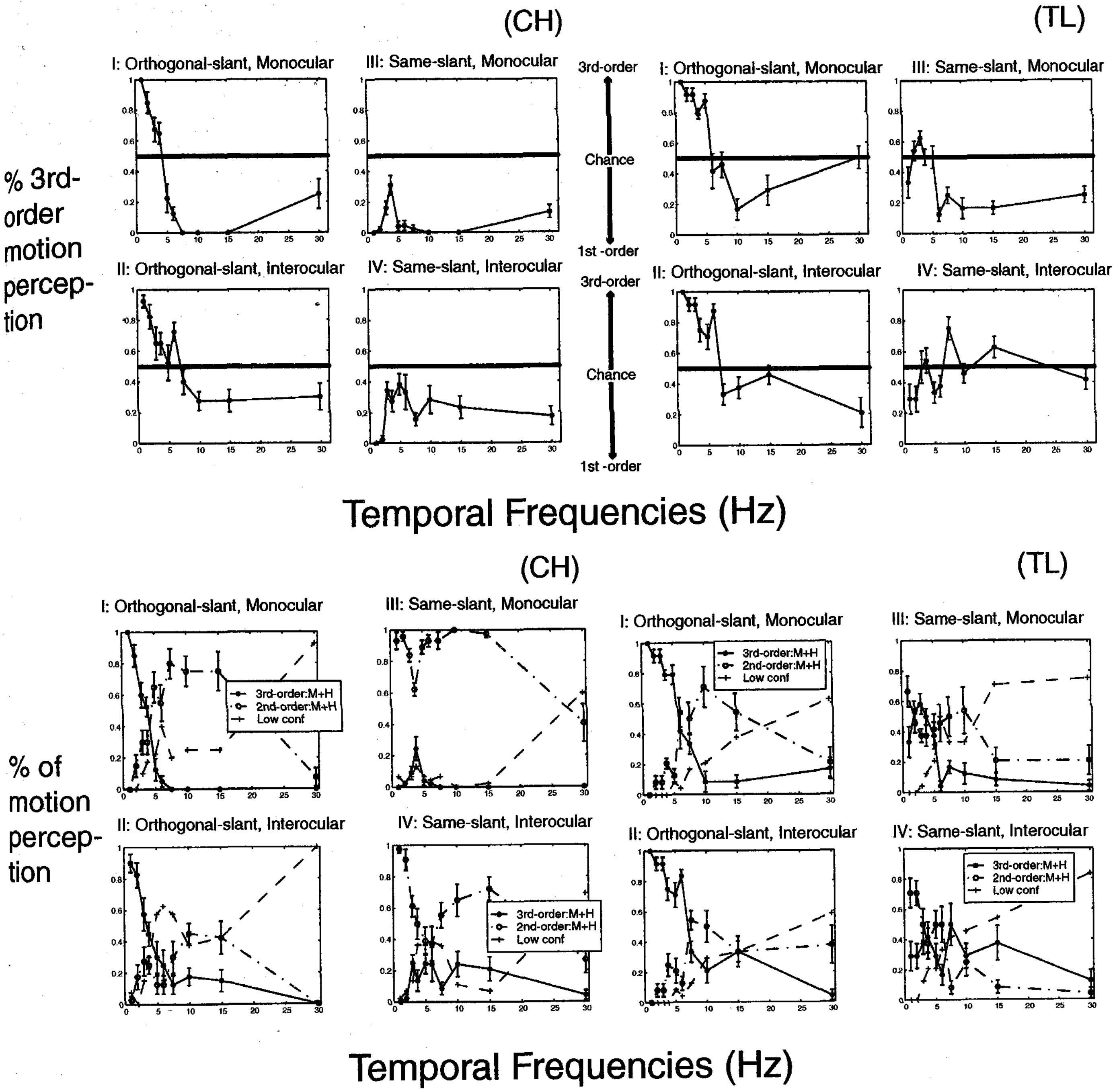


Figure 6.6: Individual data from the first- versus third-order ambiguous motion stimulus. **Upper row** Fraction of report of the third-order direction as temporal frequency varies. **Lower row** Reports of the third-order direction (medium and high confidences) and the first-order direction (medium and high confidences). Total amount of low confidence (random guesses are also indicated. Three other observers behaved similarly. Subjects: CH, JG, MC, RC, TL.

6.8 Conclusions

The selection of the first-order direction indicates computation of motion in a first-order motion system (except at temporal frequencies less than 5Hz, see Chap. 7) and the selection of the third-order direction indicates computation of motion in a third-order motion system. The evidence for these computations being performed by first- and third-order motion systems is the predicted results with manipulations of temporal frequency, monocular vs interocular viewing, and the effects of instructions. Competition between the first- and third-order systems is qualitatively similar to that between the second- and third-order systems. Initial tendencies, attentional instructions, training, temporal frequencies and viewing conditions are all important factors for the selection of either first- or third-order motion systems in the ambiguous first- vs third-order motion stimulus of Fig. 6.1. Furthermore, the first-order motion system is also insensitive to slant as predicted by standard motion analysis models, while slant is an important cue for the third-order, pattern-tracking motion system.

Chapter 7 Selecting Motion Systems

7.1 Conclusions

In the previous two chapters, various factors affecting the selection of first-, second-, and third-order motion systems were discussed. This chapter summarizes the results and discusses about any differences between the results and the predictions.

7.1.1 Texture-Defined Motion: Second- versus Third-Order

Factors influencing competition of two systems	Influence
Initial Tendency (at 2Hz)	3/7 observers second-order, 4/7 observers third-order
Attentional Instruction + Training	All perceived third-order for freq $< 3Hz$
Temporal Freq	$< 3Hz$: third-order $> 3Hz$: second-order
Monocular vs Interocular viewing	interocular favors third-order

Dual tasks:	Interference
Letter task + Letter task	Yes
Letter task+ 2nd-order motion task	No
Letter task+ 3rd-order motion task	Yes
2nd+2nd-order motion tasks	Yes
3rd+3rd-order motion tasks	Yes

7.1.2 Luminance-Defined Motion: First- versus Third-Order

Factors influencing competition of motion systems	Influence
Initial Tendency (at 2Hz)	2/6 observers first-order, 4/6 observers third-order
Attentional Instruction + Training	All perceived third-order for freq$< 5Hz$
Temporal Freq Manipulation	$< 5Hz$: third-order
	$> 5Hz$: first-order
Monocular vs Interocular viewing	interocular favors third-order

7.1.3 Summary

In texture-defined motion, both second-order and third-order motion systems may be used to extract motion direction. The second-order motion system does not share bottleneck resource with letter recognition while the third-order motion system does. In luminance-defined motion, both first-order and third-order motion systems may be used to extract motion direction. An illustration of the architecture of the motion system is shown in fig. 7.1

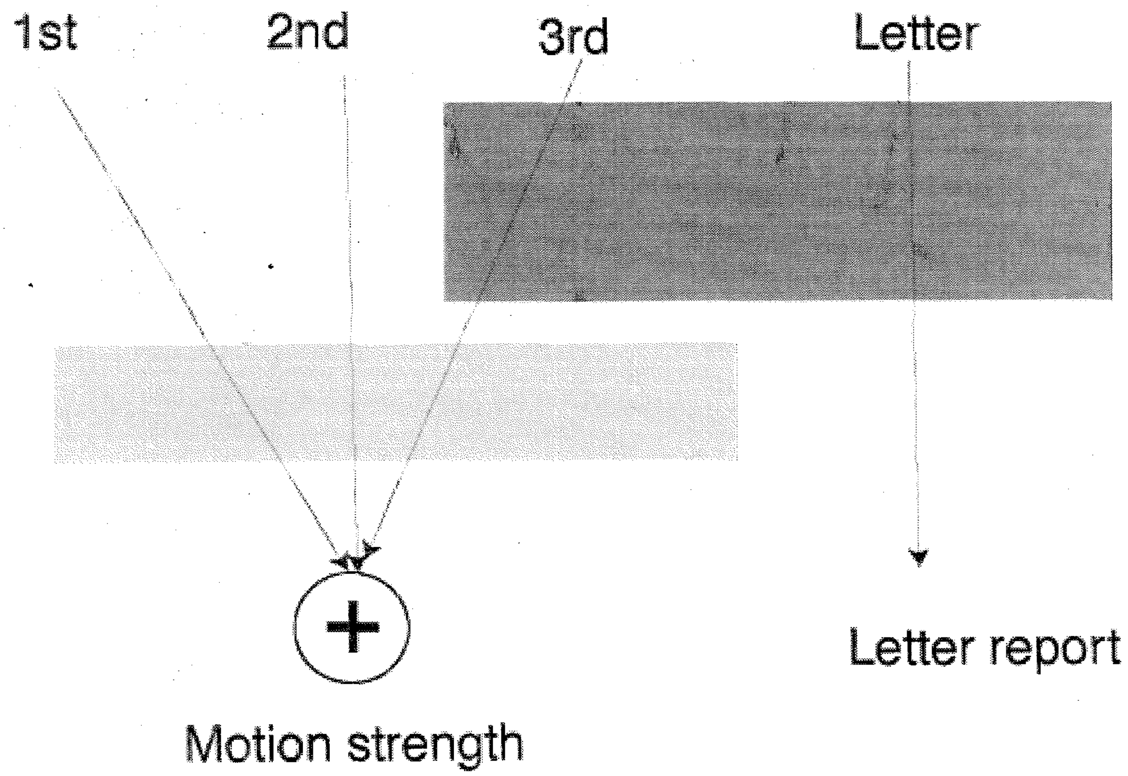


Figure 7.1: Summary of architecture of motion systems.

Training and instructions to attend to slant changed the perceived motion direction from the second- or the first- order to the third-order direction, while increasing temporal frequency changed perceived motion direction from the third-order to the second- or the first-order. Interocular (in contrast to monocular) viewing reduced the probability of perceiving motion in first- and second-order directions, the reduction was greaer for the first-order direction. Interocular viewing (relative to monocular) did not significantly reduce the tendency to perceive motion in the third-order direction. Fig. 7.2 shows a summary of these results.

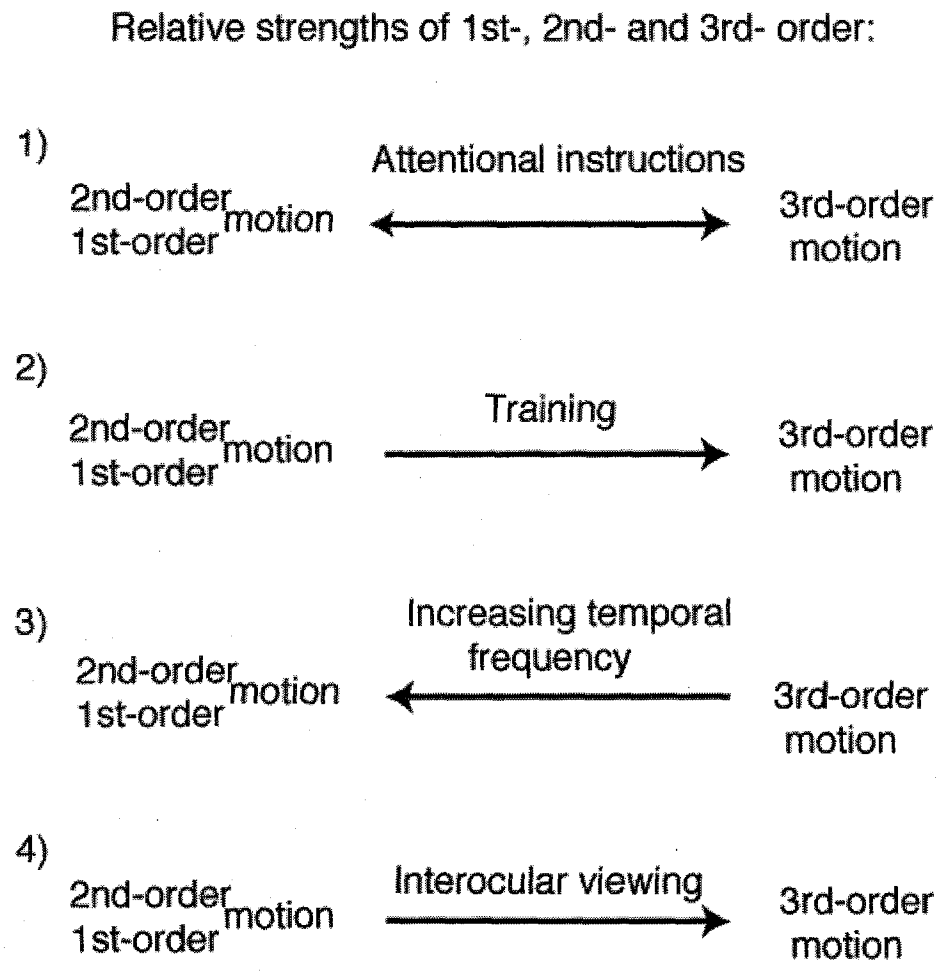


Figure 7.2: Summary of factors affecting the selection of motion systems.

7.2 Discussion

7.2.1 Motion Energy vs Pattern-Tracking

It has been established for more than a decade that second-order motion computation involves a motion-energy mechansism with a non-linear preprocessing(Chubb & Sperling, *1988*; Chubb & Sperling, *1989*; Werkhoven *et al.*, *1993*). Seiffert and Cavanagh *(1998)* claimed that second-order motion was not computed by a motion energy model because they found that their texture-defined (second-order) motion stimuli was sensitive to position rather than velocity and suggested a pattern-tracking or attention-based model for second-order motion. However, they did not provide any reason to explain why the motion energy model failed. In fact, I have shown in an earlier article (Ho, 1998b), as well as in this thesis, that texture-defined motion stimuli can be perceived by either a second-order, *motion energy mechanism*, or by a third-order, pattern-tracking mechanism. So Seiffert & Cavanagh's result is correct but is only half of the whole story. As further shown in chapter 5, second-order computation is not sensitive to grating slant and follows the motion energy model (Werkhoven *et al.*, 1993), whereas the third-order computation involves tracking slants.

I have also shown in an earlier article (Ho, 1998b) (as well as shown here in chapter 3), that texture-defined motion task interfere with another texture-defined motion task both in the case when the second-order system is used and even more so when the third-order motion system is used. Therefore it indicated that attention is required for texture-defined motion. Only that in the case of second-order motion, the attentional resource required is different from the attentional resource for the letter recognition. The attentional requirement was further proven in the Structure-Attention Mapping method and model (Ho, 1998a) (as described in chapter 4) for the pathways of the tasks conducted in the previous experiments.

Seiffert and Cavanagh (1999) showed that texture-defined motion (at temporal frequency of 0.2 Hz) was a serial, attention-demanding process. At 0.2 Hz, what Seiffert and Cavanagh observed was probably the third-order motion (the temporal frequency issue was discussed in chapter 5) instead of the second-order motion. So what they actually showed was that the third-order motion was attention-demanding. Their results serve as evidence to support the earlier discovery of Ho (1998b) that there exists another mechanism, the pattern-tracking mechanism, apart from the motion energy mechanism, that is responsible for motion perception in texture-defined motion. In addition to the discovery of this alternative mechanism for motion perception of texture-defined motion, my result (1998b) also reconciled the apparent contradiction with the previous findings (Chubb & Sperling, 1988; Chubb & Sperling, 1989; Werkhoven *et al.*, 1993) by the proposal of two concurrent motion systems for the same motion stimuli.

7.2.2 Motion Perception in First- and Second-Order Direction in Interocular Viewing

Alternative third-order motion computation

It was observed in chapter 5 that reports of the second-order direction were almost unaffected by changing viewing condition from monocular to interocular at temporal frequencies of 1-5 Hz as shown in Fig. 5.10.

Lu & Sperling (1995b) proposed that third-order motion is both bottom-up and top-down. At low temporal frequencies, even when no top-down attention is paid to any particular patterns, the salience map for third-order motion is assumed to mark the bottom-up salient patterns (the higher contrast patches) as figure, and the rest as ground. In another words, although there is no attentional instruction to attend to a particular pattern, salience is still computed by the bottom-up process. This kind of computation extracts motion in the same direction as does the second-order motion system, but it probably is carried out in higher-level brain areas.

It was also observed in Chapter 6 that reports of the first-order direction were almost unaffected by changing the viewing condition from monocular to interocular at the two lowest temporal frequencies, 1 and 2 Hz, as shown in Fig. 6.4. Here, even without attentional instructions, observers might still choose to attend to high luminance or to low luminance. As the stimuli for luminance-defined motion were not black-white calibrated (see Appendix), it is probable that the luminance minima of the of stimulus were more salient than the maxima. Therefore, motion direction might be computed by the third-order computation. This third-order computation would yield the same motion direction as the first-order computation. However, unlike the first-order computation, it would be completely binocular, and it would have a corner frequency of 3-4 Hz instead of 10-12 Hz.

Other observations

For interocular viewing, as compared to monocular viewing, reports of both the first- and second-order directions decreased somewhat for temporal frequencies of about 4-15Hz. The decrease with interocular viewing was greater for first- than for second-order. Possibly, second-order motion is not eliminated in interocular viewing, although it is significantly impaired. First-order motion, which is probably computed at a lower level in the visual system, is more nearly eliminated in interocular viewing.

There is also a provocative observation in this study that whether under monocular or interocular viewing, there is often a bump at around 6-15 Hz for the ambiguous texture-defined motion display, and around 2-5 Hz, and 5-10 Hz for the ambiguous luminance-defined motion display. These bumps may due to alias paths.

Interocular vs monocular low-level motion

The debate on whether low-level motion perception is binocular or monocular has been carried on for more than a decade. It has been long believed that early motion is monocular because of a long standing failure to achieve dichoptic motion perception in random dot displays (Braddick, 1980). However, more recently, Carney & Shadlen (1986) appeared to show that short-range motion was binocular. This position was criticized (Georgeson & Shackleton, 1992) and replied (Carney & Shadlen, 1992; Carney & Shadlen, 1993). Lu & Sperling, (1995b) using a test-pedestal paradigm and five frames for motion perception, showed that under interocular viewing, the frequency tuning functions of first and second-order motion were similar to that of third-order motion, and hence concluded that first- and second-order motions were monocular. Recently, results on second-order motion (Carney, 1993) showed that under the same test-pedestal paradigm but with an extended time of display (several seconds instead of five frames), first- and second-order motion could be perceived. Derrington and Cox (1998) also confirmed this result.

From a physiological point of view, layer $4c\alpha$ (which receives afferents from the LGN) of V1 is the farthest area into the cortex where there is complete segregation between cells from two eyes. It has also been found that most cells responding to motion directions are in the areas beyond layer $4c\alpha$ of V1. Therefore, it is almost certain that the first- and second-order motion are computed in brain areas where binocular cells are present, and where monocular cells also may be present.

Gradation of binocular and monocular low-level motion

Derrington and Cox (1998) found that two out of three observers performed worse with dichoptic than with monocular presentation, particularly at high temporal frequencies. They proposed that the reasons for the decrease in performance in low-level motion was due to the existence of monocular motion mechanism and the effect of imperfect alignment of the visual axes to the target patches.

Suppose there were a monocular motion mechanism, and the ratio of binocular motion cells to monocular cells increases from V1 up to the higher areas of cortex. It is possible that interocular viewing may decrease the perception of the first- and second-order motion to a certain extent, but not completely eliminate it. Suppose there were more noise and higher thresholds for the binocular cells and for the monocular cells, under a integration-firing model, the binocular motion cells might not be able to fire after the first cycle of motion frames, but would be able to fire after a few more iterations. Therefore, the decrease in sensitivity as compared to monocular cells can be compensated by extending the time of display. Furthermore, since it was proposed that first-order motion is computed in a lower visual area, area V1, in the cortex than second-order motion, which is proposed to be in area V2, first-order motion should suffer more loss than second-order motion. This is also seen in the this study. So one way to reconcile the differences between the experiments that support binocular low level motion and those that support monocular low level motion is as proposed above: dichoptic or interocular viewing decreases early motion to a certain extent, and the higher in the cortex the motion is computed in, the less it is affected. Hence, the third-order or feature-tracking type of motion is not affected at all, whereas the second-order motion is affected to some extent, and first-order motion is affected to a large extent.

Constant time vs constant frames (5 frames) seems to be a critical difference in the experiments supporting arguments of the side that low-level motion can be binocular and the side that low-level motion is strictly monocular. If the number of binocular cells is small, and there are some other reasons, like more noise is induced and etc, the sensitivity of binocular motion cells is lower, it is possible that under 5 frames condition, the motion signal is insignificant.

Binocularity in this study

The experiments in this study are not intended to investigate the binocularity of low-level motion. While no conclusion can be drawn on this issue, it was observed that third-order motion is not affected by interocular viewing condition, second-order was affected to some extent, first-order motion is affected to an even larger extent. Hence in the consideration of selecting motion systems, interocular viewing condition favors the third-order motion system more than the second- or first-order motion system.

Appendix A Black-White Calibration

Recent findings (Scott-Samuel & Georgeson, 1995; Nam & Chubb, 1998; Smith & Ledgeway, 1997; Lu & Sperling, 1999) found that non-linear processing of second-order stimuli produces artifactual first-order components. The visual system is more sensitive to negative contrast (" black") than to positive contrast ("white"). Here, we used a method proposed by Lu & Sperling (1999) to compensate for this kind of distortion.

As shown in Fig. A1, the odd-numbered frames are to be calibrated. The texture in these frames would have the same mean luminance if there werek no distortion. The even-numbered frames carry a small difference in luminance. If, due to distortion, there were a first-order component in the odd-numbered frames, a motion signal would be created by combining it with the real luminance component of the even-numbered frames. If there were no first-order component, the even-numbered frames alone could only produce counterphase first-order signals, the odd-numbered alone could only produce counterphase second-order signals. Since it was shown by a phase-test (Lu & Sperling, 1995b) that first- and second-order motions are computed in separate channels, it is assumed that they do not combine to form a motion signal.

Lu & Sperling (1999) suggested multiplying the negative contrast by a factor α to counter the black-white asymmetry. In the calibrating procedure in this study, α was varied from 0.4 to 1.1 and observers' motion perceptions were measured. When α is too small, motion perception was in the "α-too-small" direction. When α was too large, motion perception was in the "α-too-large" direction. The α that consistently eliminated apparent motion was selected as multiplication factor in the experiments. This α was called the "null point". As shown in chapters 5 and 6 in this study, a third-order motion can be perceived be both texture-defined and luminance-defined motion. Therefore, if temporal frequency is less than 5Hz, motion perception can be due to a third-order motion computation. Therefore, in this study, we added another constraint on the temporal frequency, such that measurements were made at a temporal frequency of about 8-10Hz, depending on whether observers could perceive any motion from a texture-defined motion stimulus as shown in Fig. 5.1 and a luminance-defined motion stimulus as shown in Fig. 6.1 at that frequency or not.

Lu & Sperling (1999) found that the null points were usually less than 1, and the average was about 0.75. Results for this study are shown in Fig. A2. For orthogonal-slant stimuli, the null points were from 0.68 to 1.1. For same-slant stimuli, the null points were from 0.75 to 1.1.

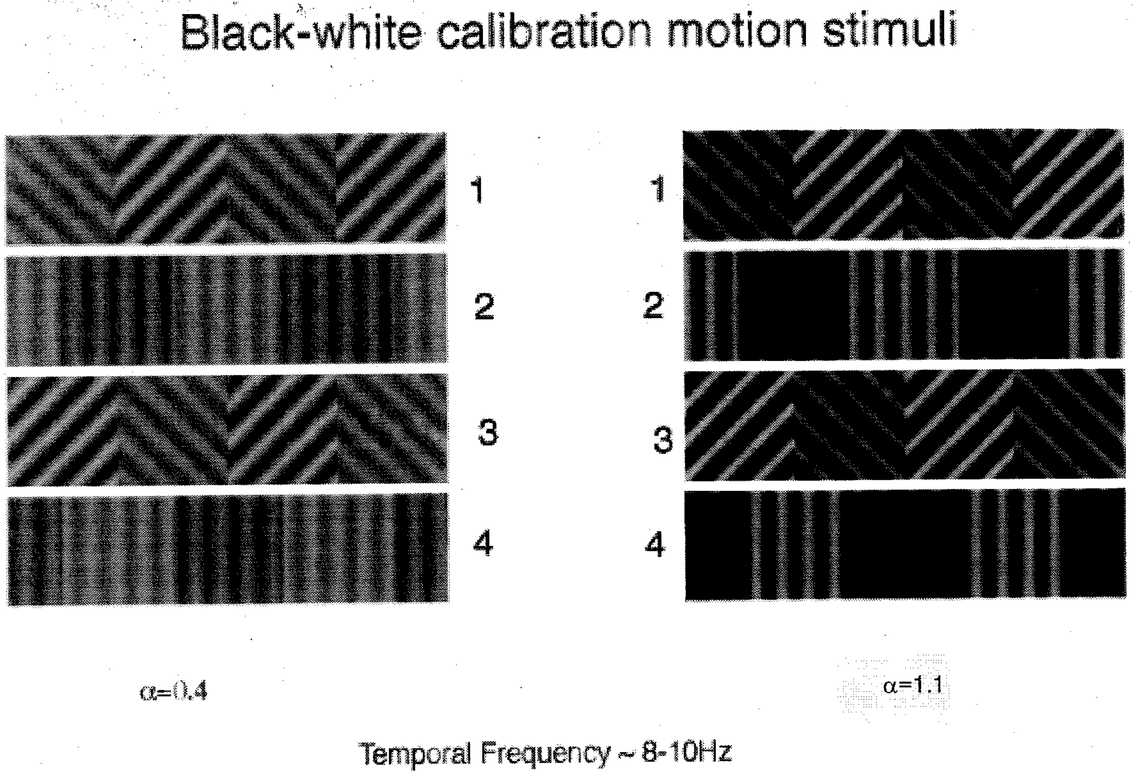


Fig. A1: Calibration motion frames for two values of α, 0.4 and 1.1.

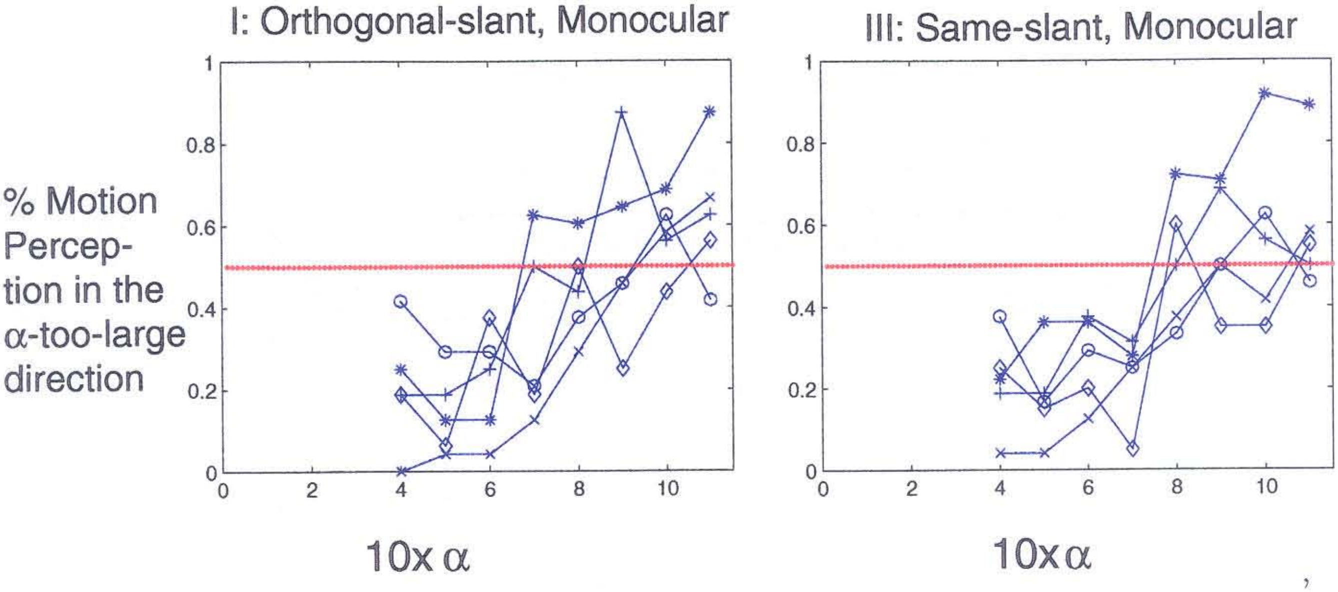


Fig. A2: Results of calibration of four observers. Motion in the α-too-large direction. X-axis is α. Y-axis is the performance on motion perception. Chance is at 0.5. For orthogonal-slant stimulus, observers' null points are 0.68, 0.7, 0.8, 0.9, 1.05. For same-slant stimulus, observers' null points are 0.75, 0.78, 0.8, 0.9, 0.9.

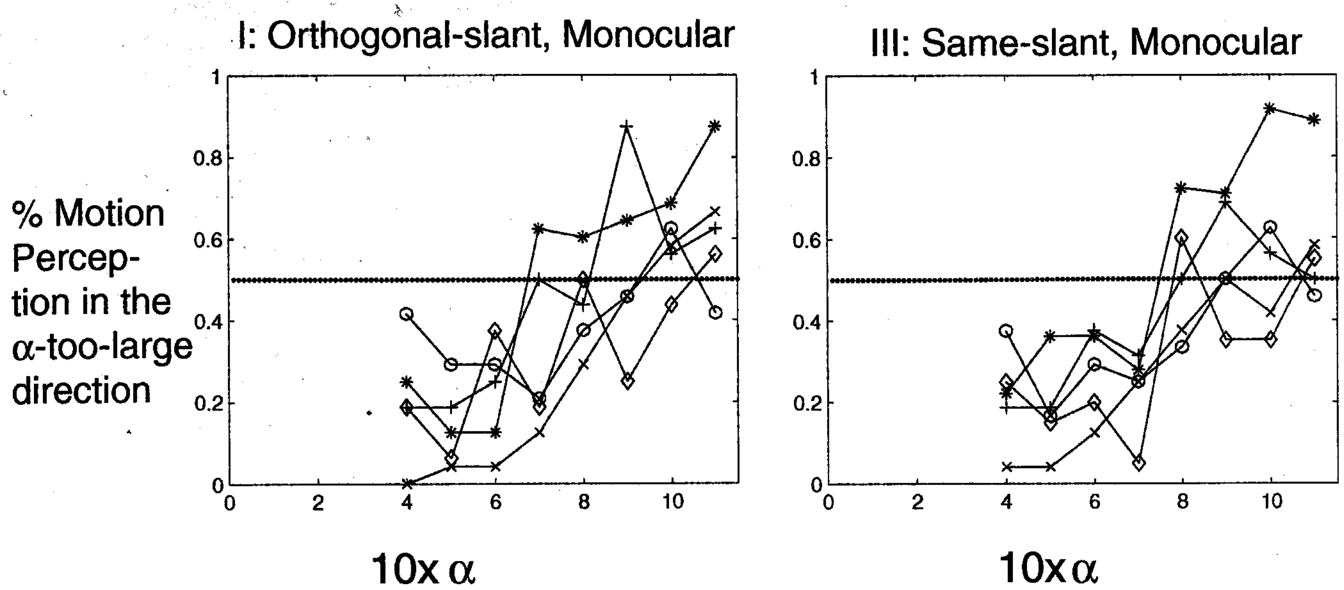
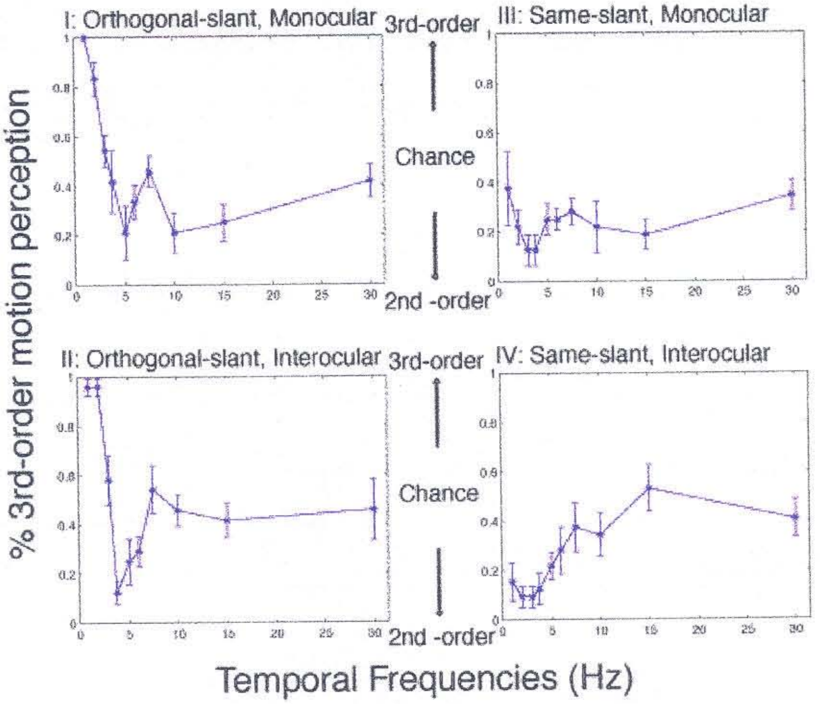


Fig. A2: Results of calibration of four observers. Motion in the α-too-large direction. X-axis is α. Y-axis is the performance on motion perception. Chance is at 0.5. For orthogonal-slant stimulus, observers' null points are 0.68, 0.7, 0.8, 0.9, 1.05. For same-slant stimulus, observers' null points are 0.75, 0.78, 0.8, 0.9, 0.9.

Results before the black-white calibration are shown in Fig. A3 and Fig. A4. They are qualitatively similar to the results after black-white calibration and overall contrast reduction as shown in Fig. 5.10 and Fig. 5.11 .

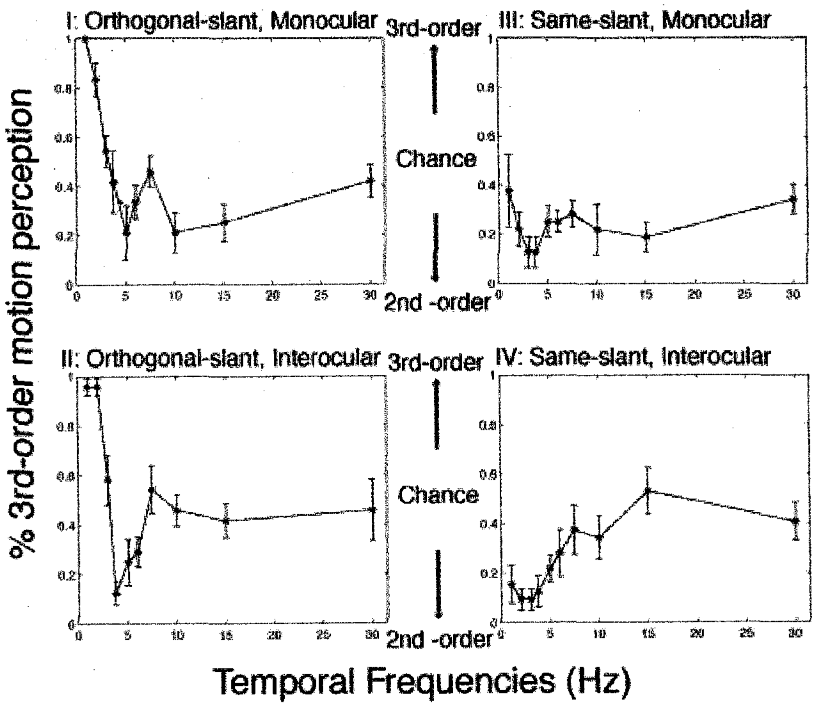


Results for 2nd vs 3rd order motion before black-white calibration (average of 2 subjects)

Frequencies(Hz): 1.00 2.00 3.00 3.75 5.00 6.00 7.50 10.00 15.00 30.00

Fig. A3: Averaged results from 2 observers for the 4 conditions before performing black-white calibration and overall contrast reduction. **Condition I Orthogonal-slant under monocular viewing:** At low frequencies ($< 3Hz$), the motion perception is mainly third-order. At higher frequencies(3-6Hz), the perception is mainly second-order, with a maximum second-order perception at around 4.5Hz. As temporal frequency increase beyond 5Hz, second-order motion perception decreases to about chance levels at around 18.75Hz. **Condition II Orthogonal-slant under interocular viewing:** The third-order motion perception at low frequencies is not affected. The second-order motion perception is affected and shifted towards chance level. **Condition III Same-slant under monocular viewing:**third-order motion perception is greatly decreased. Even at low frequency ($< 3Hz$), observers perceive second-order motion only. The perception of second-order motion is generally unaffected. As compared to Condition I, second-order motion is not sensitive to orientation while third-order motion is. **Condition IV Same-slant under interocular viewing:**second-order motion perception is decreased at around 3-7Hz as compared to Condition III. Comparing with Condition II, third-order motion

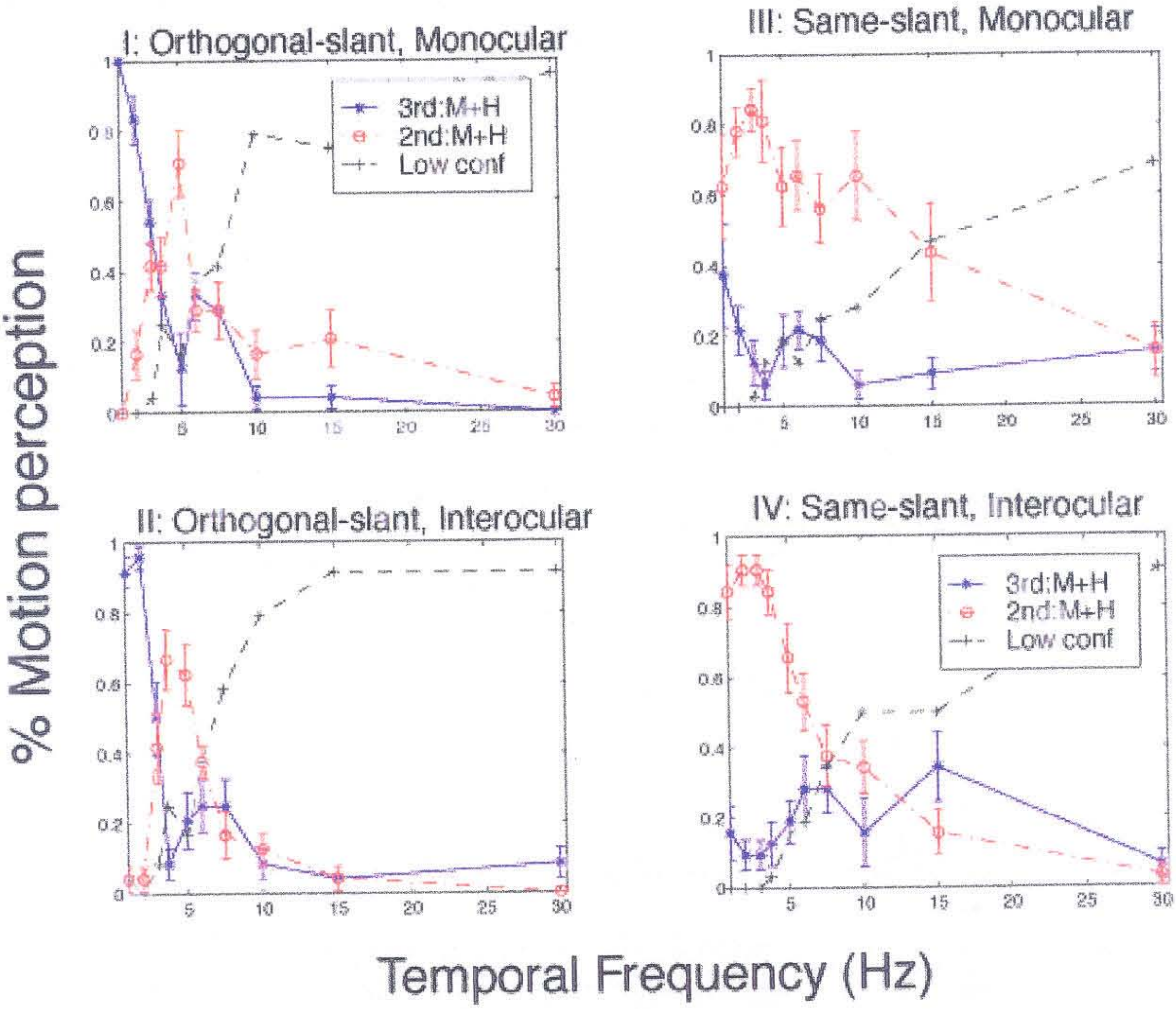
Results for 2nd vs 3rd order motion before black-white calibration (average of 2 subjects)



Frequencies(Hz): 1.00 2.00 3.00 3.75 5.00 6.00 7.50 10.00 15.00 30.00

Fig. A3: Averaged results from 2 observers for the 4 conditions before performing black-white calibration and overall contrast reduction. **Condition I Orthogonal-slant under monocular viewing:** At low frequencies ($< 3Hz$), the motion perception is mainly third-order. At higher frequencies(3-6Hz), the perception is mainly second-order, with a maximum second-order perception at around 4.5Hz. As temporal frequency increase beyond 5Hz, second-order motion perception decreases to about chance levels at around 18.75Hz. **Condition II Orthogonal-slant under interocular viewing:** The third-order motion perception at low frequencies is not affected. The second-order motion perception is affected and shifted towards chance level. **Condition III Same-slant under monocular viewing:**third-order motion perception is greatly decreased. Even at low frequency ($< 3Hz$), observers perceive second-order motion only. The perception of second-order motion is generally unaffected. As compared to Condition I, second-order motion is not sensitive to orientation while third-order motion is. **Condition IV Same-slant under interocular viewing:**second-order motion perception is decreased at around 3-7Hz as compared to Condition III. Comparing with Condition II, third-order motion perception is eliminated, confirming the conclusion from Condition I and III that third-order motion is sensitive to orientation while second-order motion is not. Subjects: JG, CH. Subjects AL, TC showed similar behavior.

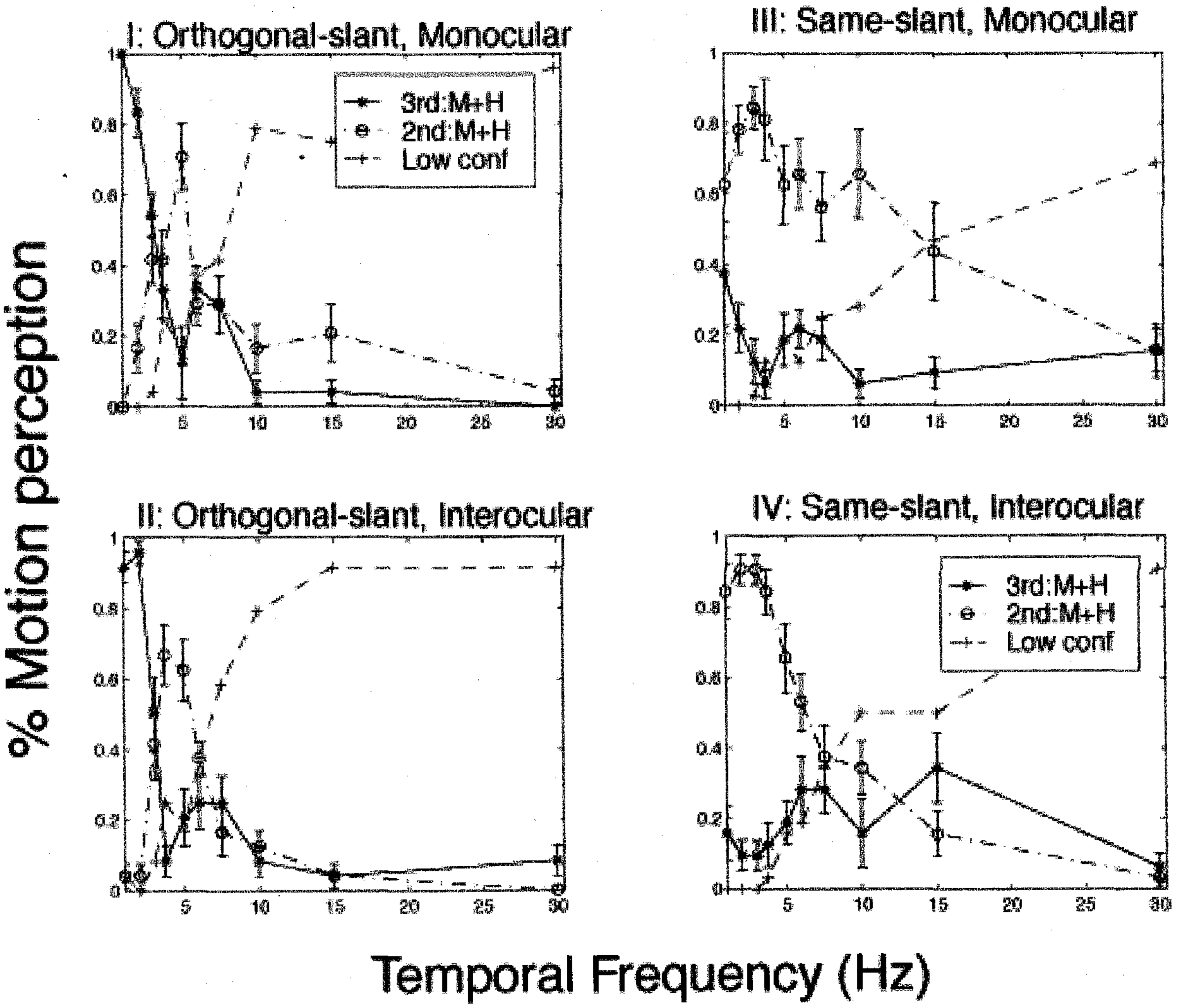
Results for 2nd vs 3rd order motion before black-white calibration (average of 2 subjects)



Frequencies(Hz): 1.00 2.00 3.00 3.75 5.00 6.00 7.50 10.00 15.00 30.00

Fig. A4: Combined results from 2 observers showing second- and third-order motion perception with medium or high confidences. Total low confidence fraction is also shown to indicate the amount of random guesses. Subject: CH, JG.

Results for 2nd vs 3rd order motion before black-white calibration (average of 2 subjects)



Frequencies(Hz): 1.00 2.00 3.00 3.75 5.00 6.00 7.50 10.00 15.00 30.00

Fig. A4: Combined results from 2 observers showing second- and third-order motion perception with medium or high confidences. Total low confidence fraction is also shown to indicate the amount of random guesses. Subject: CH, JG.

Appendix B Some Final Considerations

Single vs Multiple/Distributed Attentional Resources

Despite William James' famous statement that "everyone knows what attention is", the definition of attention is actually quite different for different people, making scientific discourse on this subject difficult. Some psychologists equated attention to "consciousness". For them, "the union of soul" often leads to the conclusion of one undifferentiated attention. Other psychologists, like Sperling (1986) and Nakayama (1990), and physiologists (Reynolds & Desimone, 1999), prefer the definition of resource allocation, competition for the limited resources or the bottlenecks of resources. In this thesis, I have adopted the information processing approach to define attention. In this view, attention is seen as a process of resource allocation. Since the computational resources are distributed in the brain, attention needs not be unitary and controlled by a single central processor, but is distributed throughout the brain as multiple computational resources that could work in parallel simultaneously.

Rensink (1999) found that in a visual search with blanks between frames, if targets changed, one could attend to 5 objects, whereas if distractors changed, one could attend to about an average of 1.4 and up to 1.6 objects. He concluded that one could only attend to one. Obviously, the conclusion and result does not completely coherent, since if one can only attend to one object, why did the result showed one attend to 1.4 objects? The redefinition of "object" to be the union of several items cannot really explain his other experimental result when observers could attend to 5 objects. In the experiment of changing distractors, attention can be drawn to those changing distractors. Therefore, searching becomes difficult. In fact, this experiment is actually an evident that "object" is not as good a measure of attention as resource allocation.

One classic experiment to test how many items one can perceive is the whole-report task (Sperling, 1960), letters or digits are presented briefly in a matrix, and observers are asked to report as many as they can. The result is that observers can report 4 or 5 items. This is an evidence for the hypothesis that one can perceive more than one thing, apart from being the evidence for the definition of attention as resource allocation and that the resources have limited capacity. The even more striking partial-report task (Sperling, 1960) which gives observers cues for which line of letters to attend to, shows that in a short period of time, our visual system actually holds all letters. This iconic memory decays very quickly. Under the resource allocation definition, it is not surprising at all that different tasks require different amount of attention, and therefore, the number of items one can attend to is different for different tasks.

Among those who believe attention as the resource allocation, there is still the question: is there a central resource for all tasks, or is it possible that different tasks use different resources. This issue is best exemplified by a task that involves different modalities, such as visual, auditory and sensory motor. While it is difficult to read a book and listen to speech at the same time, it is quite easy to read and listen to music at the same time. Why? Is there one attentional resource that is shared by all modalities or is there an independent attentional resource for each modality?

On the one hand, Duncan (1997) showed that an attention-demanding visual task does not interfere with an attention-demanding auditory task while those two tasks interfere within their modalities. On the other hand, Driver and Spence (1994) showed that attending to an visual object produced only a very modest degree of bias in the system toward processing sound coming from the same location. Within the visual modality, Duncan and Humphrey (1989) found that under the visual search paradigm, task difficulty increased with increased similarity of targets to non-targets and decrease with decreased similarity between target and nontargets, producing a continuum of search efficiency. Within visual modality, Carney, Shadlen and Eugene (1987) showed that low level motion information and color information are processed parallelly. Ho (1998b) showed that a low level motion (second-order motion) is processed parallelly with shape.

While it is common to 'enjoy' listening to music and read a book at the same time, it is difficult to listen to a radio talk show and read a book simultaneously. This is because while the auditory system and the visual system do not interfere much with each other in the former case, the interpretation of the semantic meaning of what you hear and what you read does require the same computational resource, attributing to the interference in the latter case.

Therefore, it is useful to consider attention as a process of resource allocation that is happening everywhere throughout the perceptual, cognitive and motor hierarchy in the brain. Each brain area or circuit in this hierarchy has its own finite informational processing resources and hence its attentional bottleneck. In this hierarchy, there is convergence and divergence of information flow. Competiton for processing resources (or allocation of attention) occurs at the point of convergence produces interference. When we are listening to music (without doing much intellectual interpretation of the structure of the piece) and read a book, the auditory system's resource is allocated to process the sound of music and the visual system's resource is utilize to process the word patterns in the book. These two perceptual processes are distinct, utilizing mostly different resources, so competition and interference is minimum. However, to interpret the deeper meaning of the text or the messages from a radio will require the involvement of reasoning faculty in the brain, therefore we can either read or listen to the radio's message, but not both simultaneously. Therefore, there is really no inconsistent between Duncan *et al*'s (1997) and Driver and Spence's (1994) results.

Pashler (1998) noticed that while many perceptual tasks did not interfere, many simple sensorimotor tasks interfered (Pashler, 1994; Pashler, 1998) using a response time dual-task paradigm.

One issue to consider in terms of dual-task paradigm is that, there are two different ways of measuring interference, one is to measure response time, the other is to measure performance (percentage correct). These two methodologies yield different results. A bottleneck model is often considered to consist of three stages: (1) perceptual analysis, (2) decision/response selection, (3) response production. Interference in performance indicates competition for resource in stage (1) perceptual analysis. Interference in response time may indicate competition for resources in (1) or (2). Interference in perceptual tasks can be measured by either response-time paradigm or by performance paradigm. However, it will be meaningless to measure performance in sensorimotor tasks. Reponse-time paradigm is usually used to measure interference in sensorimotor tasks.

It seems that the sensorimotor tasks are significantly different from perceptual tasks. I would consider the problem as perception is going up the cortical hierarchy, while action is going down the hierarchy. Consider one of the central executive system(s) in the brain (as shown in fig. A5) that is responsible for a number of cognitive functions, including perceptual and sensorimotor functions. For perceptual tasks, most of the computations are carried out in the earlier processing stages. The central system is not responsible for processing most information, but only taking them in. So when there are more than one *perceptual tasks being processed, performance will not be affected, while response time* may be affected. If the bottleneck resources for processing two perception tasks are non-overlapping, say neuronal pathways, P1 and P4 (both are responsible for some visual tasks), in Fig. A5, there is no interference in performance dual-task experiments, while it may be interference in response time dual-task experiments depending on the nature of the decision process at the final bottleneck. Moreover, two tasks from different modalities may or may not interfere under the performance dual-task paradigm, depending on the resource they require. For instance, P2 (visual task) and P7 (auditory task) will not interfere, but P2 (visual task) and P5 (auditory task) will interfere. So there is actually no myth in whether there are different attentions for different modalities, and no contradiction between Duncan *et al*'s (1997) and Driver and Spence's (1994) results. If the pathways of two tasks overlap, even if they are responsible for tasks with input from different modalities, there will still be interference. Of course, the more different the tasks are, the less possible the pathways responsible for those tasks will overlap. Sensorimotor tasks, on the other hand, can be very different. They are mostly generated from some higher bottlenecks. Therefore interference is often found.

When the dual-task paradigm is used to probe competition and interference of attention, it is easy to obtain false interferene results. If the instructions are not carefully given, or the observers are not trying hard to do the tasks, dual-task paradigm can produce spurious interference results. In fact, logically speaking, to prove the absence of another attentional resource, one has to exhaust all possibilities. To prove the existence of another resource, on the other hand, one counter example is enough. So, undifferentiated attention is a position that is almost impoissible to prove. Experiments that claimed that visual attention is unitary and undifferentiated often compare tasks that have a common resource bottleneck somewhere.

The theory of "multiple, distributed" instead of "single, unitary" in the consideration of attention seems to take on even to the consideration of consciousness. While people used to believe in one unitary consciousness, Zeki (1998) proposed that consciousness itself is a modular, distributed system. He drew results from physiology that the parallel, multistage systems of the visual brain are both processing and perceptual systems (where conscious perceptions of certain attributes were formed), and do not project to a unique common area. He also supported this position by studies of diseased human brains that showed that activity in separate processing-perceptual system, especially those of color and motion, could lead to the perception of the relevant attribute even when the other processing systems were inactive and that activity in individual processing-perceptual systems had a conscious experience as a correlate.

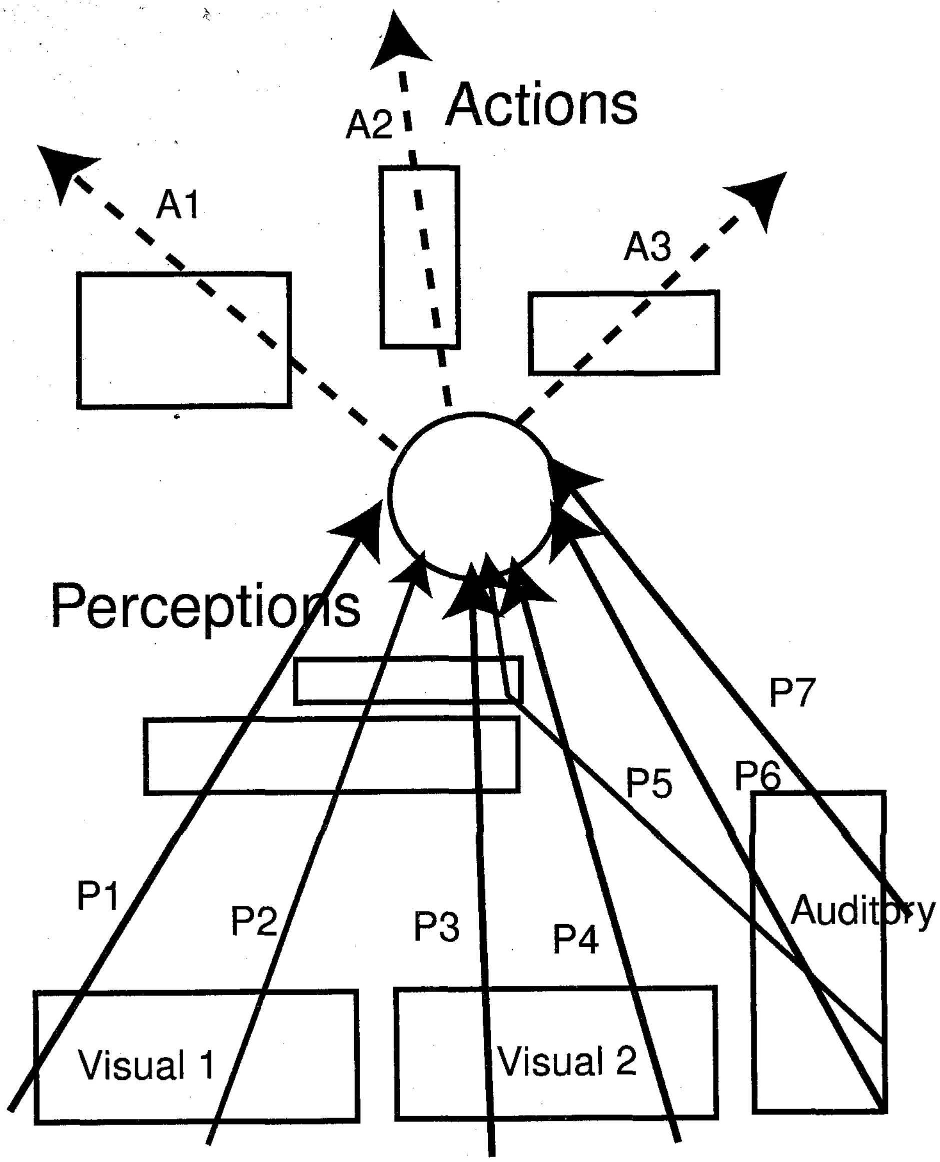


Fig. A5: Perception tasks and action tasks going into and out of one of the bottleneck(s) in the higher brain area for congnitive functions.

Attention vs Will

When we are asked to attend completely to something, how much attention can we actually put into it? While attention is described as the bottleneck resource, it is also often used to refer to the mechanism that change the weight (Ho, 1998b; Ho, 1998a), gain control, or increase contrast (Reynolds *et al.*, 1997). While it has been standard practice in psychophysics and physiology to call this kind of phenomena "attention", it may be confusing to have different definitions of attention. Under this kind of terminology, there are various unspecified definitions of attention: voluntary, sustained, top-down vs involuntary, transient, bottom-up. When one volunatarily decides to attend to object 1, a salient object 2 suddenly appearing in the visual field can draw the attention away from the object 1. In fact, this is a trick magicians often use. When audience are paying attention to the hand where some magic is supposed to happen, magician waves a handkerchief in the other hand, and the audience's attention will be drawn to the handkerchief. So even when we attend to something, we may not be actually attending to it.

Therefore, it will be easier to refer attention to the resource allocation. This allocation can be caused by a bottom-up process, say, some salient objects in the visual field, or by a top-down process, which is better described as "will". The input units compute the bottom-up saliency. The "will" manipulates the weight (or, gain control, neuronal interactions), the output unit take in all these weighted inputs and compute the result using a decision process. The output is the resource. The final result reflects the resource allocation for the inputs. These kind of competition probably exists in many levels in the brain. The inputs that win in all the competitions through the whole pathway reach the perception.

This way, attention refers only to one concept. Bottom-up or top-down are only the causes of attention. It may be easier to call the change in weight (gain control, neuronal interaction, and etc), "will", and the actual resource allocation a task determined by natural weight is "bottom-up attention", and if this resource is given after changing the weight with will, the "top-down, selective attention".

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**(2), 284–299.

Albright, T. 1982. Form-cue invariant motion processing in primate visual cortex. *Science*, **255**, 1141–1143.

Albright, T. 1984. Direction and orientation selectivity of neurons in visual area MT of the macaque. *J. of Neurophysiology*, **52**, 1106–1130.

Blasdel, G.-G, Lund, J.-S, & Fitzpatrick, D. 1985. Intrinsic connections of macaque striate cortex: axonal projections of cells outside lamina 4C. *J. Neurosci.*, **5**, 3350–3369.

Boulton, J. C., & Baker, C. L. 1993. Dependence on stimulus onset asynchrony in apparent motion: Evidence for two mechanisms. *Vis. Res.*, **33**, 2013–2019.

Braddick, O. J. 1974. A short range process in apparent motion. *Vis. Res.*, **12**, 43–77.

Braddick, O. J. 1980. Low-level and high-level processes in apparent motion. *Phil. Trans. R. Soc. Lond. B*, **290**, 137–151.

Carney, T. 1993. Evidence for an early motion system which integrates information from the two eyes. *Vis. Res.*, **33**(14), 1977–1996.

Carney, T., & Shadlen, M. N. 1986. Mechanisms of human motion perception revealed by a new cyclopean illusion. *Science*, **232**(4746), 95–97.

Carney, T., & Shadlen, M. N. 1992. Binocularity of early motion mechanisms – comments on Georgeson and Shackleton. *Vis. Res.*, **32**(1), 187–191.

Carney, T., & Shadlen, M. N. 1993. Dichoptic activation of the early motion system. *Vis. Res.*, **33**(14), 1977–1996.

Carney, T., Shadlen, M. N., & Eugene, S. 1987. Parallel processing of motion and colour information. *Nature*, **328**(13), 647–649.

Cavanagh, P. 1992. Attention-based motion perception. *Science*, **257**, 1563–1565.

Cavanagh, P., & Mather, G. 1989. Motion: The long and short of it. *Spatial Vision*, **4**, 103–129.

Chubb, C, & Sperling, G. 1988. Drift-balanced random stimuli- A general basis for studying non-fourier motion perception. *J. Opt. Soc. Am. A*, **5**(11), 1986–2007.

Chubb, C, & Sperling, G. 1989. Texture interactions determine perceived contrast. *Proc. Natn. Acad. Sci. U.S.A.*, **86**, 2985–2990.

Derrington, A., & Cox, M. 1998. Temporal resolution of dichoptic and second-order motion mechanisms. *Vis. Res.*, **38**(22), 3531–3539.

DeYoe, E A, & Van Essen, D C. 1988. Concurrent Processing streams in monkey visual cortex. *Trends Neuroscience*, **11**, 219–226.

Dobbins, A. C., Jeo, R. M., Fiser, J., & Allman, J. M. 1998. Distance modulation of neural activity in the visual cortex. *SCIENCE*, **281**(5376), 552–555.

Driver, J., & Spence, C. J. 1994. *Spatial synergies between auditory and visual attention.* Cambridge, MA: MIT Press. Pages 311–331.

Duncan, J. 1993. Similarity between concurrent visual discriminations - dimensions and objects. *Perception & Psychophysics*, **54**(4), 425–430.

Duncan, J., & Humphreys, G. W. 1989. Visual search and stimulus similarity. *Psychological Review*, **96**(3), 433–458.

Duncan, J., S., Martens, & R., Ward. 1997. Restricted attentional capacity within but not between sensory modality. *NATURE*, **387**(6635), 808–810.

Georgeson, M. A., & Shackleton, T. M. 1992. No evidence for dichoptic motion sensing - A reply to Carney and Shadlen. *Vision Res.*, **32**(1), 193–198.

Graziano, M.-S.-A., Andersen, R.-A., & Snowden, R.-J. 1994. Tuning of MST neurons to spiral motions. *J. of Neurosicence*, **14**(1), 54–67.

Ho, C. E. 1996. Architecture of motion pathways for second-order motion. *CNS Technical Report*, CNS–TR–96–02.

Ho, C. E. 1998a. Architecture of Cortex Revealed by Divided Attention Experiments. *Proc. of Theoretical Aspects of Neural Computation (TANC'97)*, 165–179.

Ho, C. E. 1998b. Letter recognition reveals pathways of second-order and third-order motion. *Proc. Natl. Acad. Sci. USA*, **95**, 400–404.

Ho, E.-C., & Koch, C. (1996)*Inv-Ophth-V* **37**, 3 Ho, C.-E. (1996) *CNS Technical Report* **CNS-TR-96-02**.

Ho, E. C., & Koch, C. 1996. The role of attention in second-order motion. *IOVS*, **37**(3).

Johnston, A., McOwan, P. W., & Buxton, H. 1992. A computational model of the analysis of some first-order and second-order motion patterns by simple and complex cells. *P. Roy. Soc. Lond. B Bio.*, **250**, 297–306.

Kanwisher, N.-G., Yin, C., & Wojciulik, E. 1997. *Repitition Blindness.* Cambridge, MA: MIT Press.

Koch, C., & Ullman, S. 1985. Shifts in selective visual attention: towards the underlying neural circuitry. *Human Neurobiology,* **4**, 219–227.

Livingstone, M, & Hubel, D H. 1987. Psychophysical evidence for separate channels for the perception of form, color, movement, and depth. *J. Neurosci.*, **7**, 3416–3468.

Lu, Z. L., & Sperling, G. 1999. The amplification principle in motion perception. *IOVS*, **40**(4), 199.

Lu, Zhong-Lin, & Sperling, George. 1995a. Attention-generated apparent motion. *Nature*, **377**, 237–239.

Lu, ZL, & Sperling, G. 1995b. The functional architecture of human visual motion perception. *Vision Research*, **35**(19), 2697–2722.

Maunsell, J, & Van Essen, D.-C. 1983a. Functional properties of neurons in middle temporal visual area of the macaque monkey I: Selectivity for stimulus direction, speed and orientation. *J. of Neurophysiology*, **49**(5), 1127–1147.

Maunsell, J, & Van Essen, D.-C. 1983b. Functional properties of neurons in middle temporal visual area of the macaque monkey II: Binocular interaction and sensitivity to binocular disparity. *J. of Neurophysiology*, **49**(5), 1148–1167.

Nakayama, K. 1990. *The iconic bottleneck and the tenuous link between early visual processing and perception.* Cambridge, MA: Cambridge University Press. Pages 411–422.

Nam, J. H., & Chubb, C. 1998. The mechanism that detects texture-defined motion is highly sensitive to luminance-defined motion. *IOVS*, **38**(4), 1106–1106.

Pashler, H. 1994. Dual-task interference in simple tasks: data and theory. *Psychollogical Bulletin*, **116**(2), 220–244.

Pashler, H. 1998. *The Psychology of attention.* Cambridge, MA: MIT Press.

Ramachandran, V S, & Gregory, R L. 1978. Does colour provide an input to human motion perception? *Nature*, **275**, 55–56.

Reichardt, W. 1957. Autokorrelationsauswetung als Funktionsprinzip des Zentralnervensystems. *Z. Naturforschung Teil*, **B**(12), 447–457.

Rensink, R. A. 1999. The magical number one, plus or minus zero. *IOVS*, **40**(4), 279B239.

Reynolds, J. H., & Desimone, R. 1999. Competitive mechanisms subserve attention in macaque areas V2 and V4. *J. Neurosci*, **19**(5), 1736 1753.

Reynolds, J. H., Pasternak, T., & Desimone, R. 1997. Attention increases contrast sensitivity of cells in macaque area V4. *IOVS*, **38**(4), 3206–3206.

Rockland, K.-S. 1985. A reticular pattern of intrinsic connections in primate area V2 (area 18). *J. Comp. Neurol*, **235**, 467–478.

Scott-Samuel, N. E., & Georgeson, M. A. 1995. Does early non-linearity account for second-order motion? *Perception*, **24**, 104.

Seiffert, A. E., & Cavanagh, P. 1998. Position displacement, not velocity, is the cue to motion detection of second-order stimuli. *Vis. Res.*, **38**(22), 3569–3582.

Seiffert, A. E., & Cavanagh, P. 1999. Serial Search for texture-defined motion. *IOVS*, **40**(4), 805.

Shipp, S, & Zeki, S. 1989a. The organization of connections between areas V5 and V1 in macaque monkey visual cortex. *Eur. J. Neurosci.*, **1**, 309–332.

Shipp, S, & Zeki, S. 1989b. The organization of connections between areas V5 and V2 in macaque monkey visual cortex. *Eur. J. Neurosci.*, **1**, 333–354.

Smith, A. T., & Ledgeway, T. 1997. Separate detection of moving luminance and contrast modulations: Fact or artifact? *Vision Research*, **37**, 45–62.

Sperling, G. 1960. The inforamtion available in brief visual presentations. *Psychological Monographs: General and Applied*, **74**(498), 1–29.

Sperling, G., & Dosher, B.-A. 1986. *Strategy and optimization in human information processing.* Wiley, NY: (Wiley-Interscience Publication). Pages 2–1 – 2–65.

Sperling, G., & Melchner, M.-J. 1978. The attention operating characteristic: Examples from visual search. *Science*, **202**, 315–318.

Sperling, G., Budiansky, J., Spivak, J.-G., & Johnson, M.-C. 1971. Extremely rapid visual search: The maximum rate of scanning letters for the presence of a number. *Science*, **174**, 307–311.

Taub, E., Victor, J. D., & Conte, M. M. 1997. Nonlinear preprocessing in short-range motion. *Vis. Res.*, **37**, 1459–1477.

Ungerleider, L G, & Mishkin, M. 1982. *Two cortical visual systems.* Cambridge: MIT Press.

Vaina, L.-M., & Cowey, A. 1996. Impairment of the perception of 2nd-order motion but not first-order motion in a patient with unilateral focal brain-damage. *Proc. Roy. Soc. B*, **263**, 1225–1232.

Van Santen, J. P. H., & Sperling, G. 1984. Temporal covariance model of human motion perception. *J. Optical Society America A*, **1**(5), 451–473.

Van Santen, J. P. H., & Sperling, G. 1985. Elaborated Reichardt detectors. *J. Optical Society America A*, **2**, 300 321.

Werkhoven, P., Sperling, G., & Chubb, C. 1993. The dimensionality of Texture-defined Motion: a Single Channel Theory. *Vision Res.*, **33**(4), 463–485.

Werkhoven, P., Sperling, G., & Chubb, C. 1994. Perception of apparent motion between dissimilar gratings-spatiotemporal properties. *Vision Res.*, **34**(20), 2741–2759.

Wilson, H.-R. 1994. The role of second-order motion signals in coherence and transparency. *1994 Higher-order processing in the visual system, Ciba Foundation Symposium*, **184**, 227–244.

Wilson, H.-R., Ferrera, V.-P., & Yo, C. 1992. A psychophysically motivated model for two-dimentional motion perception. *Visual Neuroscience*, **9**, 79–97.

Zeki, S. 1990. The motion pathways of the visual cortex. *Pages 321 345 of:* Blakemore, C. (ed), *Vision: Coding and Efficiency.* Cambridge, England: Cambridge University Press.

Zeki, S. 1998. Paralle processing, asynchronous perception, and a distributed system of consciousness in vision. *Neuroscientist*, **4**(5), 365 372.

Zeki, S, & Shipp, S. 1988. The functional logic of cortical connections. *Nature*, **335**, 311–316.