Visual Control of Posture and Visual Perception of Shape

Tjeerd Dijkstra

Visual Control of Posture and Visual Perception of Shape

een wetenschappelijke proeve op het gebied van
de Natuurwetenschappen

Proefschrift

ter verkrijging van de graad van doctor
aan de Katholieke Universiteit Nijmegen,
volgens besluit van het College van Decanen
in het openbaar te verdedigen
op maandag 17 januari 1994
des namiddags te 1.30 uur precies

door

Tjeerd Maarten Hein Dijkstra

geboren op 1 mei 1961 te Biak

Promotor: Prof. Dr. C.C.A.M. Gielen

This work was supported by the Dutch Foundation for Biophysics (NWO) and by ESPRIT Basic Research Actions 3149 and 6615.

CIP-GEGEVENS KONINKLIJKE BIBLIOTHEEK, DEN HAAG

Dijkstra, Tjeerd Maarten Hein

Visual control of posture and visual perception of shape /
Tjeerd Maarten Hein Dijkstra. - [S.l. : s.n.]. - Ill.
Proefschrift Nijmegen. - Met lit. opg. - Met samenvatting
in het Nederlands en Indonesisch.
ISBN 90-9006625-X
Trefw.: visuele perceptie / houdingscontrole.

aan Judy

aan mijn ouders

Corrected errors and typeset as a PDF file, 19 December 2000.

The Dutch and Indonesian summary, the words of thanks and the “flip the corners” images of my original thesis are not part of this reprint.

About the cover:
The figure on the cover is Saraswati, the Balinese deity of books, knowledge and learning [22]. She is depicted as a young woman, because science is always attractive and constantly rejuvenates itself. The swan on which she is seated is a symbol of wisdom. Her attributes are a chain, symbol of religion, a guitar, symbol of music and art and a palm-leaf book, symbol of knowledge and learning. The drawing was made by I Ketut Darwin.

About the corners of the pages:
When turned in rapid succession, the corners of the odd pages show a rotating section of a paraboloid and the derived velocity field in the image plane (compare with fig. 1–5). The paraboloid oscillates with a period of 20 double pages. There is a perturbation of 180 deg at page 101 (see chapters 2 and 3). The paraboloid is tessellated with a grid that is regular in the image plane. Because of this, one can detect the shape already without any movement. I tried to suggest the paraboloid by a random dot pattern, so that the shape cannot be detected without movement, but one does not get any 3D shape percept from that. This is probably because one has to bend the pages considerably in order to get them turning in rapid succession.
When turned in rapid succession, the corners of the even pages show the vector field of the double deformation (compare with fig. 1–7). The vector field oscillates with a period of 20 and oscillates in-phase with the paraboloid. One can easily observe the threefold symmetry of the double deformation.

Contents

Chapter 1

Introduction and summary

1.1 Introduction

What does it mean “to see”[1]? An obvious answer is: to know “what is where” by looking. This seems trivial to us because we see all the time: it is generally not a conscious act to see. But consider for a moment what is happening: light reflects from objects, enters our eye and is focused on the retina. There it interacts with photo pigments and generates neuronal activity in the layers of the retina. This neuronal activity is propagated to the visual cortex which is now thought to consist of more than 50 functionally different areas.

In a similar vein we can ask: what does it mean “to move”? Again, an obvious answer is: to cause something (or oneself) to change position by moving. And again this usually seems trivial to us because we move all the time, without giving this a moment thought. And just as for perception many processes are taking place: signals from central areas activate parts of our motor cortex and from there muscles are activated and finally the proteins in the muscles cause a movement.

I have deliberately contrasted perceiving and moving in the two paragraphs above. Clearly the two are not independent: we have to perceive a target in order to move to it, and we make movements in order to obtain a better view of them. Very striking examples of the interdependence[2] of visual perception and motor control are offered by infants. Infants of a few weeks old already use vision to regulate the amplitude of their arm movements: they use vision to learn to control their arm [81]. The fact that vision depends on the ability to move

---

[1] For a deeper answer to this question see the book by Marr [49].

[2] This interdependence is often called the action-perception cycle.

actively was shown in experiments with kittens: when kittens are passively moved from birth, they can see but their visual system performs worse than kittens who were allowed to move freely [30]. It seems that one has to move oneself in order to learn how to see.

An important question in the fields of visual perception and motor control is how the brain achieves its aims. This is an old question already posed by the ancient philosophers but a satisfactory answer has not been given yet, despite the much greater knowledge we have nowadays. There has been progress though, in the realization that the brain can be regarded at many different levels and that each level often supports many perspectives. One may regard the brain at the level of atoms and molecules e.g. the physics of nerve conduction or the force generation by proteins in the muscle. One may regard the brain at the level of individual cells e.g. by measuring the activity of neurons in response to visual stimuli. One may regard the brain at the level of groups of neurons or of complete areas e.g. by making brain scans. Finally, on the highest level, one observes the behaviour generated by the brain e.g. by showing visual patterns to human observers and by recording their response. The lowest level is usually within physics and chemistry, the middle levels are within physiology and the highest is usually within psychology[3]. Each level again supports many different perspectives. There is the perspective of the theoretician and of the experimentalist, of the person interested in normal behaviour and of the person more interested in pathology. Finally, and often overlooked, there is the perspective of common man.

The behavioural level adopted in this study is the most abstract level in the classification of the previous paragraph. The perspective used is a rigorous one: I tried to model the systems I studied by an explicit mathematical formulation. The attempt to do this high level modelling of a complex system is not new. The approach originated after the second world war under the name of cybernetics[4]. Its application to biological systems dates from the sixties. In this approach one identifies a measurable input (e.g. the size of the visual stimulus on the retina) and a measurable output (e.g. a movement of the observer in response to the stimulus). Note that both are time series and that the input is usually under experimental control. The power of cybernetics is that one can find the relations between input and output almost in an automatic way. Roughly, what one does is to use a random input series of sufficient length, the argument being

---

[3] See Riccio [60] for an equivalent scheme.

[4] What I mean here by “cybernetics” is a restricted type of cybernetics. This restricted cybernetics is also known under the names of “system identification” or “system theory”, see e.g. [47].

that all stimulus patterns occur because of randomness. Thus, by using random input one can measure the response to all possible stimulus patterns and one can completely describe the system under study.

The approach I used (the dynamic approach) differs considerably from the cybernetic approach[5], although I like the emphasis that cybernetics puts on building mathematical models. To understand the difference between the two approaches let us reflect on a central assumption of cybernetics that the characteristics of a system do not change when it is exposed to random input. In everyday life we are normally not exposed to random input (a house party being a possible exception) but we experience a slowly changing, clearly structured input. Our behaviour is adapted to this input. An important aspect that has been left out of cybernetics is something that might be termed very broadly as the "state of the mind". This is the idea that perception is not only specified by the current stimulus but that e.g. the previous percept or the expectations and attention of the observer also are important. The equivalent idea for motor control is that movement is not just dependent on the current motor command, but also on the previous command or on learned behaviour. Human observers will try to understand a stimulus, will try to make an internal model of the stimulus[6]. Human observers will use this internal model to perceive the stimulus better, e.g. by changing view position. Therefore, a complex interaction between stimulus and observer takes place. By using random stimuli cybernetics tries to get rid of the effects of this interaction. But in doing so it also gets rid of a lot of interesting effects which are relevant for human perception and motor control in natural settings.

Of course these considerations about the shortcomings of cybernetics in modelling human behaviour are not new, but they have been considered as side issues and not suited for rigorous modelling. Using modern developments in the theory of dynamical systems, new directions have been proposed, where one tries to keep more of the complexity of biological organisms in the models. The success of my approach in modelling postural control shows that these ideas can be exploited successfully for the relatively high level task of upright posture. The success of the cybernetic approach has been mostly in modelling of peripheral systems, probably because on a low level we cannot change the characteristics of our behaviour depending on the input. For example, when light impinges on our retina, neurons in the visual cortex will become active, no matter what we do (provided we do not take very drastic measures). In general, the closer we come to the periphery, the harder it becomes to have a conscious influence on what

---

[5] More differences are discussed in e.g. [66].

[6] Mathematically, this could correspond to a dynamical system that has a fixed point under white noise input, but a limit cycle under periodic input.

happens. The question which approach is better, has to be settled in practice: both have their merits and depending on the system under study, one will be more suited than the other.

1.2 Visual control of posture

In chapters 2 and 3 of this thesis I consider the problem of the influence of vision on postural control (upright standing). Humans have only two legs and thus their upright posture is mechanically unstable. Without any form of control we would fall down. We use many sources of information about the orientation of our body relative to gravity to stabilise our posture. The insight that we use multiple sources of information, is a pretty recent one. Since the discovery of the balance organ in the inner ear, it was thought that only that organ is responsible for balance control. More recently, it has been established that vision is an important source of information for stabilisation too. When we move, the images of objects in the environment move over our retina. Because the balance organ measures acceleration, it cannot detect slow movements or movements at constant velocity. This problem for postural control is solved by the brain by assuming that the visual environment does not move. As a consequence, when the visual world moves over our retina we have the percept that we move ourselves. Examples of this are big screen cinemas and the departing train illusion: when the train on a neighbouring platform leaves, we think for a moment that our train is leaving. These illusions illustrate the usefulness of multiple sources of information: one source can sometimes be deceived but to deceive all of them is much harder. In the case of the departing train, a part of the inner ear which senses acceleration, would enable us to notice that it is not our train which departs.

In my experiments I used a moving room set-up to induce postural sway in standing human subjects (see figure 1–1). When subjects experience visual motion of the whole environment, they assume that they are moving (which is generally correct, except in my set-up). I did not move an actual room but simulated a moving wall by computer. This simulation was done by measuring the eye position of the subject and by using this to generate the geometrically correct view of the wall. The plus of computer generated stimuli is that I have great control over the timing and the spatial pattern. The minus is that the stimuli did not seem very realistic: all my computer could manage was a wall covered with 140 small bright dots. This was the only thing the subjects saw, the rest was dark, because otherwise this would give subjects a cue about their movement.

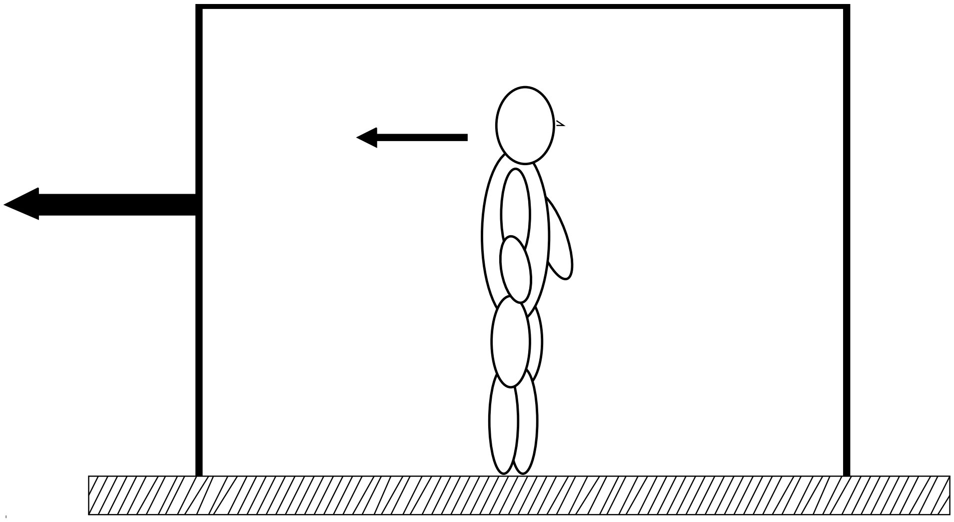


Figure 1–1 The moving room. When the room approaches, the subject sways back to avoid the percept of falling forward.

As I stated before I tried to model the postural control system by dynamical systems. As this is not a standard approach I developed a new method and some software for data-analysis. My method has a much greater generality than for the analysis of a moving visual stimulus and the postural response of subjects viewing this stimulus. Therefore, I will describe it separately from the results of the experiments, which I will discuss in the next paragraphs. The method is suited for exploring the timing relationships between two time series of oscillatory data. The ideal data for my analysis tools are two time series that not so much influence the amplitude of one another but influence the timing of one another. If the two time series are tightly locked, my method offers no advantage over cybernetics. Many phenomena in biology and physiology could be analysed with my method. The method might also be applicable to other fields e.g. economics. Many phenomena are of a cyclical nature (e.g. economic growth and inflation) and are influencing one another but are not tightly locked. It would be interesting to apply the method I developed for postural control to these type of data.

As an introduction to the methods which I used, let us consider the timing relationship between breathing and step cycle[7] during running: the amplitude of both is probably already maximum (otherwise you could run faster) and the timing is where all interesting dynamics happens. I usually run four steps in

---

[7] After writing this introduction, I found out that there is a considerable body of literature on this subject, e.g. [58, 7].

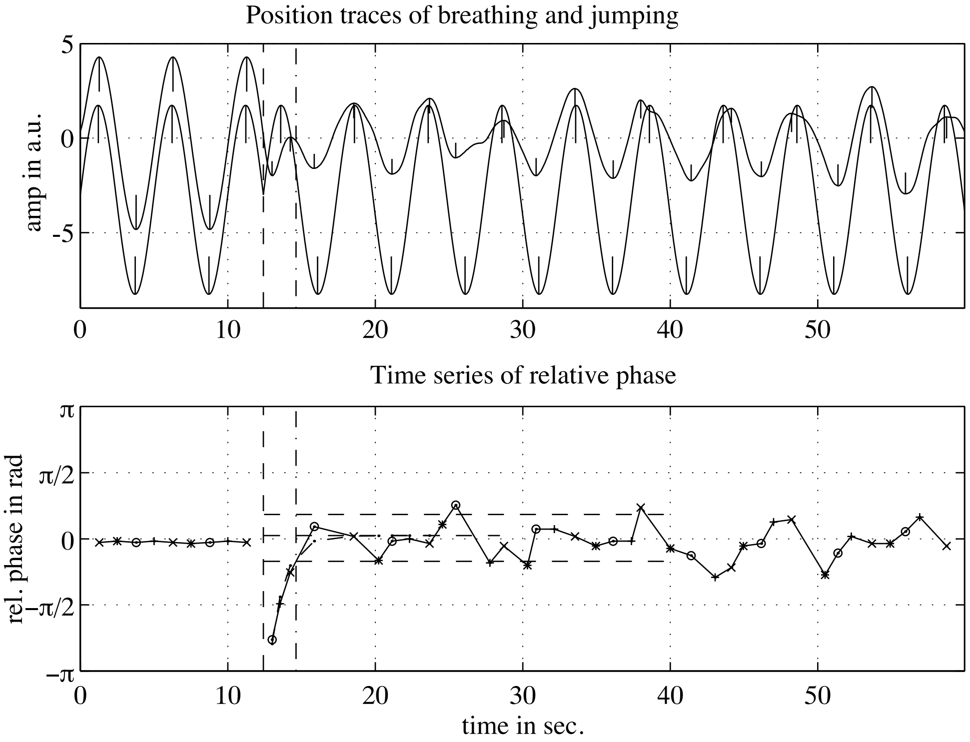


Figure 1–2 Example of two time series with high temporal stability. Trajectories of two oscillators are plotted in the upper panel and the derived time series of relative phase is plotted in the lower panel. I simulated the data by integrating a driven linear second order oscillator. The noise is switched on after the perturbation.

one breathing cycle. Of course, I do not run precisely four steps: from breathing cycle to breathing cycle there is some variability. When I get tired this variability grows. Also when I start running, I easily avoid objects: when I have to change rhythm, e.g. to jump over a pool, I quickly recover my rhythm. When I get tired this recovery from changes of rhythm is slower. Thus we have discovered two measures of *temporal stability*: the variability in timing of the step cycle compared with the breathing cycle, and the rate of recovery after a change of rhythm. These two quantities are the central issue in chapter 2. Especially important is the fact that these quantities both measure temporal stability. Thus when some parameter is manipulated (say, the amount of training) we expect both quantities to covary. Now, when I continue running I get even more tired and occasionally I go to five steps in one breathing cycle. Of course, I will try to stick with 4 to 1 but every now and then I do not manage and I change to 5 to 1 for one breathing cycle. The fixed relationship (when I am still quite fit) is called *absolute coordination* and the variable relationship, with switches of rhythm is called *relative coordination* [82]. Finally, when I am really at the limit (and hopefully in sight of the finish) there is no strong relation between running and breathing any more. This is termed *absence of coordination*. This breakdown of a fixed time relationship between two rhythms is the central issue of chapter 3.

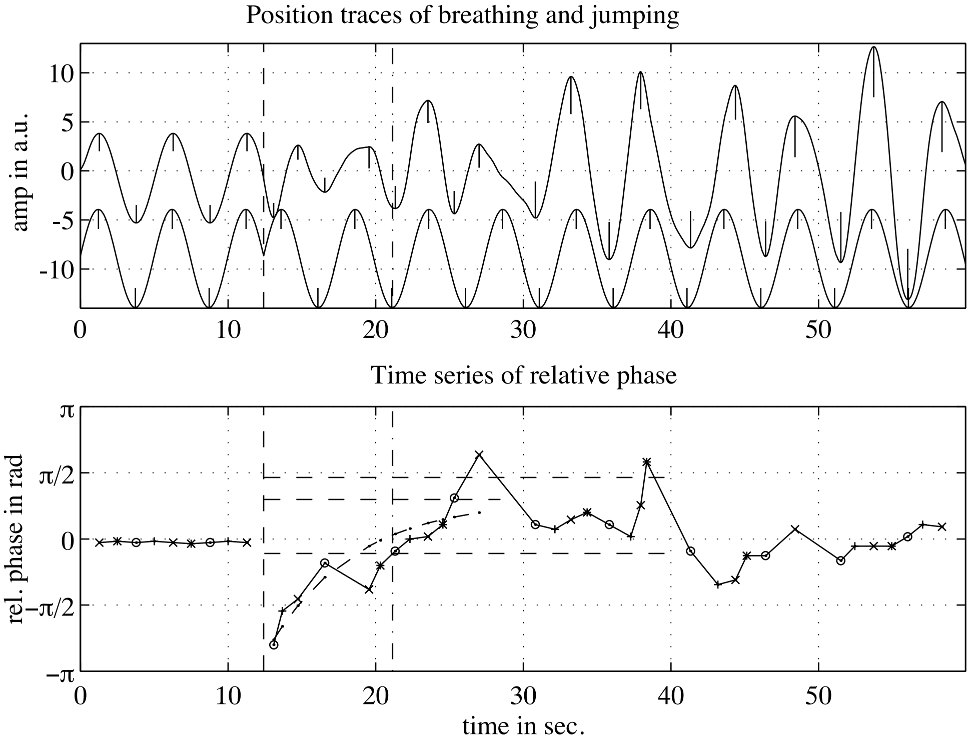


Figure 1–3 Example of two time series with low temporal stability.

In figures 1–2 and 1–3 I have illustrated the measures of temporal stability. In figure 1–2 top row, we see two oscillatory trajectories. They show 1 to 1 locking, because both time series have the same number of cycles. Thus they do not describe my running (which is 4 to 1) but more the jumping of a kangaroo. For the purpose of this illustration, this change is just a matter of convenience. Also for convenience, I have chosen one of the trajectories without noise (say, the breathing) and one with noise (the jumping). In the first 12 seconds I show a deterministic timing relationship, to show how the trajectories look without noise. Both breathing and jumping are exactly in-phase. Obviously the timing relationship is easily described in this case. At time 12 seconds there is a sudden change of breathing and then we observe a noisy time relationship: sometimes jumping is ahead of breathing, sometimes it is delayed. In the bottom row of figure 1–2, I have plotted this time relationship. Positive means that jumping is ahead of breathing, negative means that it is the other way around. Now, directly after the sudden change in breathing we see that jumping is very much delayed relative to breathing. It takes our electronic kangaroo a few seconds to return to in-phase breathing and jumping again. In figure 1–2, I show the kangaroo in a fit state: the relative timing does not fluctuate very much and it returns fast after the sudden change. Thus temporal stability is high. This should be contrasted with figure 1–3, where I show the kangaroo in a tired state. Here, the relative timing fluctuates considerably and it returns slowly after the sudden change. Thus temporal stability is low.

In both figures 1–2 and 1–3 I have illustrated absolute coordination, where both time series have the same number of cycles. In figure 1–4 I have illustrated relative coordination. In this case our electronic kangaroo is breathing faster than he can jump. In the bottom row of figure 1–4 this is illustrated by the time series disappearing at the top and reappearing at the bottom. Note the two time series have a preferred time difference. When they have this difference, the jumping frequency is almost equal to the breathing frequency: a long period of this behaviour occurs around time 20 seconds. Then, the jumping cannot follow the faster breathing anymore (around time 25 seconds) and jumping misses a few cycles compared to breathing.

After this introduction to the methods which I have developed for the analysis of postural control, I will now describe the results of the experiments I performed. In the first experiment (chapter 2) I manipulated the mean distance between a simulated wall and the subject. Clearly, distance has an influence on the stability of posture: for a fixed amount of sway of the subject, nearby objects will lead to larger retinal motion than objects farther away. For example, construction workers, working on a high building have much more sway. Clearly, the nearest visual landmarks e.g. neighbouring buildings, are far removed. In the experiment I moved the wall periodically with a small amplitude (4 cm) and a low frequency (0.2 Hz) in forward/backward direction. This is so slow that the movement of the wall is hardly noticed by the subjects. Subjects swayed back and forth periodically and usually in phase with the stimulus, thus behaving as our electronic kangaroo of the previous paragraph. From the trajectories, I extracted the so-called time series of relative phase which describes the differences in timing

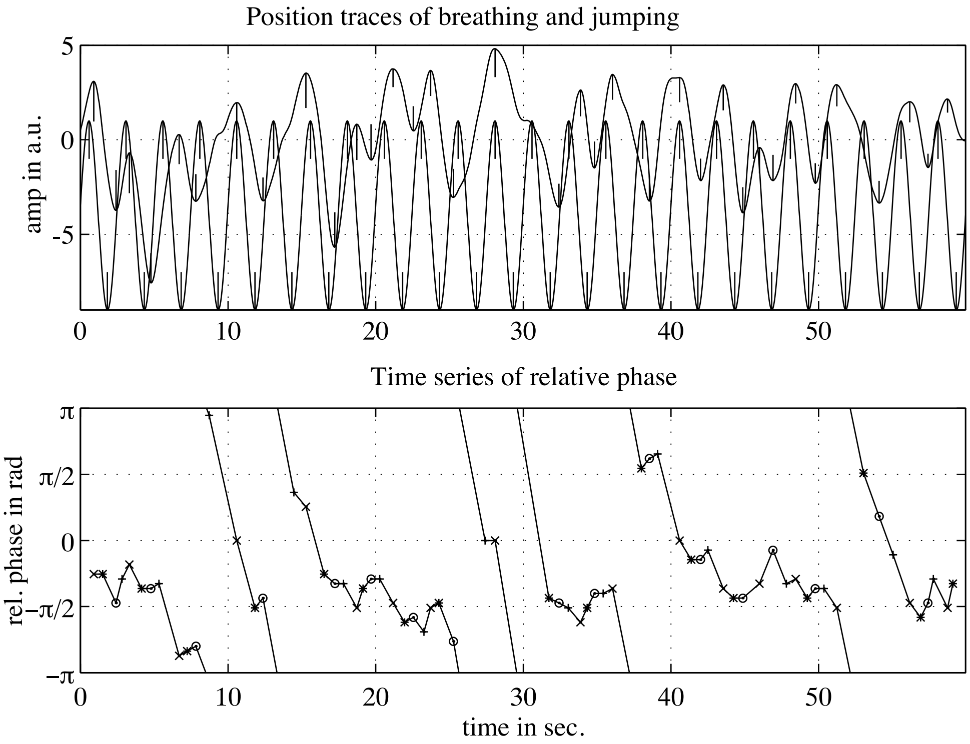


Figure 1–4 Example of two time series with relative coordination.

between the movement of the wall and the postural response. The time series of relative phase is not constant: it fluctuates around the mean value. The amount of fluctuations can be quantified by the angular deviation, which is one of the measures of temporal stability discussed in the previous paragraph. At random intervals I changed the direction of movement of the wall (see figure 1–2): the subjects were out of phase and it took them some time to regain their normal in-phase behaviour. The time to go back to normal behaviour is called the relaxation time and it is the second of the measures of temporal stability discussed above. Remember that I changed the mean distance to the wall: I found that temporal stability is high when the wall is near, and that it is low when the wall is far away. Moreover, I found a significant correlation between the two measures of temporal stability, angular deviation and relaxation time. The results showed that a linear dynamical model (which is discussed in section 3.2) successfully described the timing aspects of postural control in a moving visual environment. One detail of the model, the amplitude of postural response, was not correctly predicted. Therefore, I performed a second experiment, which I will describe in the next paragraph.

In the second experiment (chapter 3) I varied the frequency of the oscillating wall, keeping its velocity constant. There are several reasons why changing frequency is interesting. First the inner ear (balance organ) is a high-pass filter i.e. it works only for relatively high frequencies (this is one of the reasons the train illusion works, the equivalent with departing planes would probably not). Second, as postural control on a stable support is largely achieved from the ankles[8] it is mechanically difficult to produce high frequencies. Most importantly, we found in chapter 2 that postural control does not consist just of movements compensating for currently perceived motion. It seemed that subjects somehow "learned" the dynamics (this is something cybernetics would have trouble with). By learning I mean that subjects perceive the movement of the environment and then actively produce the compensating movement themselves. For higher frequencies I found that subjects could not follow the rhythm. More interestingly, sometimes they did not just give up, but tried to follow the rhythm, leading to relative coordination (without consciously knowing that they did).

The research reported in these two chapters could be extended in many ways. Below I will give some ideas which seem especially promising to me. First, one could try to develop some of the methods in a diagnostic tool for medical practice. Balance disorders can be caused both by problems with the inner ear and after amputation of part of a leg. In the second case patients have to learn to walk again with their prothesis. This training takes place in the hospital, where the environment is safe and predictable. It turns out that a number of patients still have trouble with walking after returning home. Part of the problem might be the rough diagnostic tools of the medical staff. There are some simple balance tests that measure sway amplitude, but they are not necessarily good predictors of stability. My concept of temporal stability comes closer to a good performance measure, because it includes stability when an unexpected event happens (in the relaxation time).

A second idea is to change the structure of the visual environment. This was the original idea for the project, but I never got around doing it (sorry, Stan). I always used a wall because this has a fixed distance to the subject, (contrary to e.g. a tunnel) and this is easier in the mathematical models. A subject only perceives a moving image of the environment and has to extract the movement of the environment from the image. The subject could make use of particular

---

[8] There are other so-called "strategies" for postural control, but this distinction is not so important for our purposes.

differential invariants (I will introduce these in chapter 5) and one could devise stimuli to test which one(s) is used.

A third idea would be to continue on the hypothesis that subjects learn the dynamics. This could be tested by using stimuli that do not change phase, as in my experiments, but change frequency. When subjects have learned the dynamics better they might be faster in adapting to frequency changes.

Yet another problem is that not all subjects are affected in the same way by my stimulus: some subjects just did not sway coherently with the visual motion in my set-up. This is not just a problem of my set-up, but also occurs in other experiments [4]. I had the impression that part of the problem lies in the fact that subjects are not relaxed: just talking to them about anything sometimes made a considerable difference[9].

1.3 Visual perception of shape

In the chapters 4 and 5 of this thesis I consider the problem of visual perception of the shape of a three-dimensional (3D) surface. Human observers use many sources of information for the perception of the shape of an unknown object. An important source is the motion of the image of the object on our retina when we move relative to the object. The distribution of velocities caused by this movement, is different for different object shapes. In figure 1–5 left panel I have sketched a sphere covered with dots (the sphere itself is invisible, imagine it to be of glass). Also sketched is the projection of the sphere on an image plane (say, the screen of your computer) which looks just like a cloud of dots. By a small rotation of the sphere the dots in the image sweep out a short path: the so-called *velocity field* (see figure 1–5 right panel). In figure 1–6 I have plotted this velocity field again in frontal view. Looking only at the field of velocities it is hard too see what object shape generates this field. But when I would show the small vectors of the velocity field as moving dots on a computer screen, all observers accurately perceive the shape of the object. Obviously, our brain is capable of extracting the shape of an object from a field of velocities very accurately.

The problem of the visual perception of 3D shape from visual motion, which I consider in chapters 4 and 5 is complementary to the problem I consider in chapters 2 and 3. There, I look at the problem of how visual motion influences the perception of self-motion. These two problems illustrate nicely the dual nature of visual motion: it can be caused by movement of oneself, in which case

---

[9] This effect is also used by hospital personnel when they measure your heart rate (they do not want the measurement to influence your heart rate).

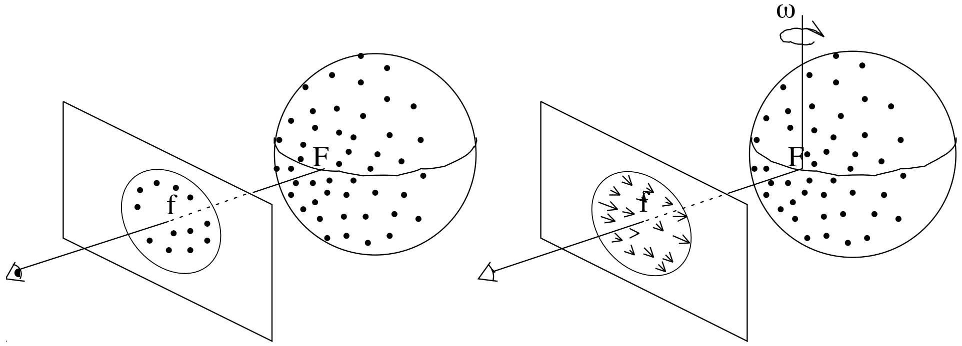


Figure 1–5 The velocity field. In the left panel a sphere and its image is plotted, in the right panel the sphere is rotating around a vertical axis (with rotation rate ω) and this gives rise to a field of velocities in the image. The fixation point is denoted by F.

one would like to know e.g. the direction of self-motion, or it can be caused by movement of objects, in which case one would like to know e.g. the shape of the object. Besides complementarity, there are many differences between chapters 2 and 3 on one hand and chapters 4 and 5 on the other. The first two chapters are commonly classified as part of motor control whereas the last two chapters as belonging to psycho-physics and computational theory. The first two chapters focus on the temporal characteristics (using dynamical systems), whereas the last two focus on the spatial characteristics (using differential geometry).

How could the brain extract shape from a field of velocities? Looking again at the plots of the velocity fields in figure 1–5 we note that the distribution of velocities is smooth i.e. any velocity vector resembles its neighbours. When we would observe a random distribution of velocities (e.g. the velocities of the snow flocks in a snow storm) we would not perceive any shape (which is the veridical percept for the snow storm). The smoothness of the velocity field is caused by the smoothness of the shape: if it has edges, like e.g. a cube, the velocity field would be discontinuous (in the first derivative). I restrict my attention to smooth objects, mainly because these are easier from a mathematical viewpoint.

So we have singled out the mathematical problem which the brain solves in the perception of shape. The solution takes as input a smooth distribution of velocities and returns as output the shape of the object. This problem is known as the "shape-from-motion" problem. Besides the shape of the object, some more things enter in the mathematics too, making matters much more complicated. First, the object moves with an unknown velocity relative to us. Second, it is at an unknown distance and further we might look at the object from an angle

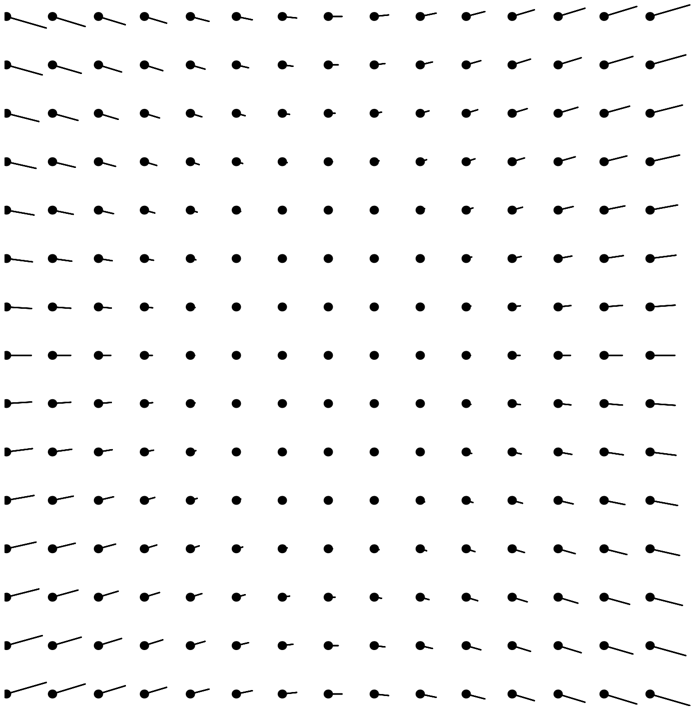


Figure 1–6 A frontal view of the velocity field of a rotating sphere.

(like viewing a painting from an oblique angle). There are several ways the brain might cope with these extra complications. Maybe the brain calculates these quantities also from the distribution of velocities. A straightforward solution of the shape-from-motion problem with all these quantities is quite complicated [84] and it seems unlikely that the brain would actually use this. Another possibility would be to construct shape measures that can be extracted directly from the velocity field. This is the approach I take in chapter 5. Finally, we could make use of other (non-visual) information too. For example, when we are moving, we could use knowledge of our velocity of self movement in the task of perceiving the shape of objects. This question is tested experimentally in chapter 4.

The first question I consider in the visual perception of shape is whether subjects directly use information about their own movements in the perception of shape. If observers would do this, the solution to the shape-from-motion problem is easy (assuming the object itself is stationary): the ratio of observer velocity and retinal velocity of a dot gives the distance to the dot and by knowing the distances of all dots we can extract shape. I tested this by showing planar or curved surfaces (segments of spheres) to actively moving observers. The observers moved in left/right direction in front of a big screen. I constructed the objects by computer and by feedback of head position, just as I did in the experiments on postural control. The objects were covered with dots, just as in figure 1–5. The task of the subject was to say whether the surface was planar, convex or concave. When the surfaces were very curved, subjects always perceived the correct shape. When the curved surfaces were not so curved, subjects often responded planar, making an error. For planes subjects mostly responded planar but sometimes concave or convex. Thus, I could use the percentage of correct responses as a criterion for the performance of a subject. In the condition where the subjects are moving themselves, they can use all information about their own velocity in perceiving. I compared the performance in this subject movement condition with the performance in other conditions, where subjects had less or no knowledge about the movement of the object relative to themself.

In the subject movement condition I measured and stored the head movements of the subjects. I used these measured head movements to recreate exactly the same retinal stimulus for a stationary observer. I did this by fixing the head of the subject using a chinrest and by playing the stimulus as a movie. I showed two types of stimuli to the stationary subjects. (1) Translating objects, where the object moved relative to the subject just as in the subject movement condition and where the subjects had to make eye movements to fixate the object. (2) Rotating objects, where the retinal motion of the object was the same as in the previous conditions if fixation was perfect and where subjects only had to look ahead in order to fixate the object. Thus I had three conditions[10]: a condition of self movement where subjects had all information about their own movements, a translation condition where subjects only had information about their eye movements and a rotation condition where subjects had no other information besides the visual stimulus itself. The results showed that performance was best in the rotation condition, where subjects had no information concerning the velocity of the object. Thus, subjects do not use knowledge of their velocity directly in shape perception. If they had, they would have done best in the condition where they were moving themselves. Instead, subjects seemed to use information about self movement to fixate better. The experiments also hinted that subjects might use the direction of the velocity of self movement (i.e. moving to the left or to the right) to establish whether the object is concave or convex. I will come back

---

[10] Actually I had six conditions because all movement conditions were shown with a large field of view and a small field of view. For simplicity, I only discuss the main results for a small field of view.

to this in the next paragraphs when I discuss the model of chapter 5.

The second question I consider in the visual perception of shape is the shape-from-motion problem. I delayed the discussion of what I mean by “shape” or, more precisely, what quantities I use to characterise shape. Intuition can help us here: we call both a small and a large sphere of being of the same shape, they only differ in size. Such a size independent descriptor of shape is the *shape index*. The novelty of my approach is that I showed that the shape index can be calculated directly from the velocity field by making some reasonable approximations. To understand what I mean by *direct*, we must consider the complete solution of the shape-from-motion problem. The solution is to write as many equations as unknowns and to solve this system of nonlinear coupled equations by algebraic means. The solution returns all of the unknown parameters as a big chunk. For humans this might not be optimal: the different parameters that enter in the equations are mathematically all equivalent but they are quite different in an ecological sense. For the velocity the observer might use other information as well: in chapter 4 I present evidence that observers can use the direction of their movements in the perception of shape. Further, distance and orientation of the surface change because of the movement, whereas the shape is a property inherent to the object i.e. it does not change. Returning to my approach, by direct calculation I mean that I do not calculate the other non-shape parameters, like the velocity of the observer. The key to my solution lies in the concept of *invariance* under rotations of the image plane. Invariance exploits the idea that there is no preferred direction in the image plane. Thus I write all my equations in such a way that they are independent of the orientation of the coordinate system in the image plane.

The velocity field generated by movement of a smooth object can be decomposed according to order. Order zero is an approximation of the velocity field by a constant, meaning that all vectors of the velocity field of the zeroth order term are parallel. Also it can be shown that the zeroth order term is already an invariant, the so-called translation. In figure 1–7 I have plotted the invariant decomposition of the velocity field of the sphere of figure 1–5 with each row representing a different order. On the top row one sees the translation which gives the mean of the velocity field generated by the moving sphere. This invariant of the velocity field is related to the distance to the object: an object that is twice as far removed will have a translation which is twice as small. Of course the translation also depends on the velocity of the observer but so do the other invariants, only the dependence on the distance is different among them. The first order consists of four quantities: the linear spatial variations of the horizontal component of the velocity field in the horizontal and the vertical direction and the linear spatial

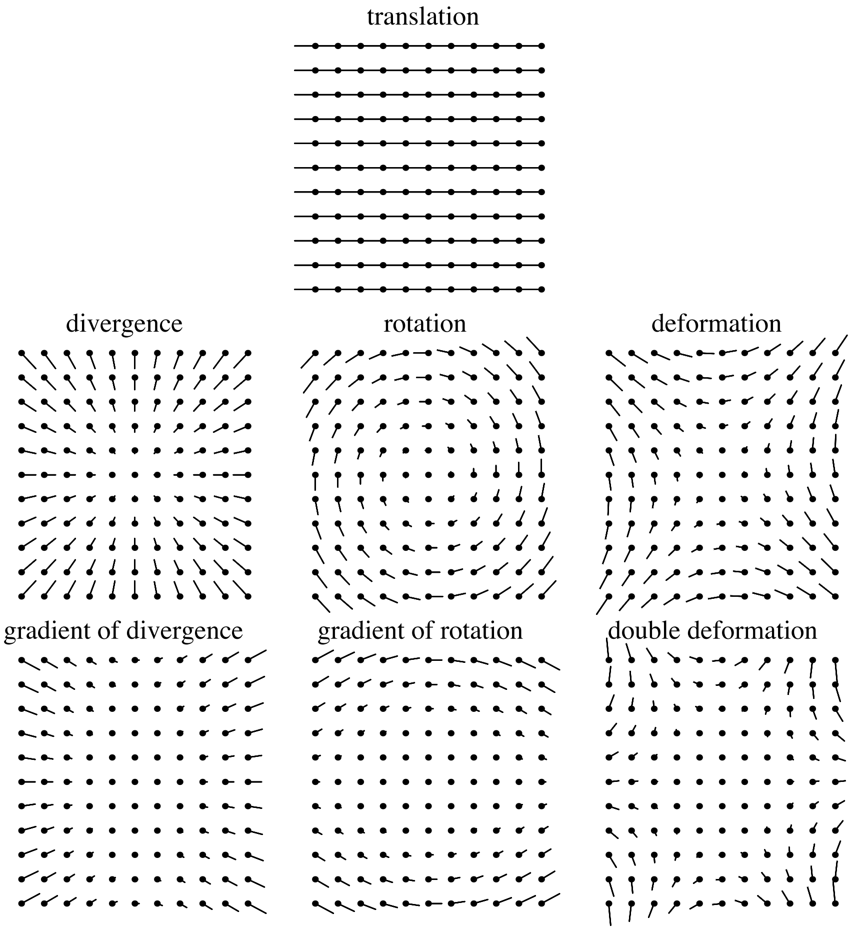


Figure 1–7 The decomposition in invariants of the velocity field of a rotating sphere. For clarity, the invariants are not to scale.

variations of the vertical component in the horizontal and the vertical direction. These four quantities are defined relative to a coordinate system. Turning the page by a quarter turn, horizontal becomes vertical and vice versa. Clearly, my description of the linear spatial variations is not invariant[11]. It turns out that certain combinations of linear spatial variations are invariant. These are plotted in fig. 1–7, middle row. From left to right we have the divergence, the rotation and the deformation. The divergence is equivalent to an expansion/contraction, the rotation to a rotation around an axis perpendicular to the image and the

---

[11] Thus it is variant but nobody ever uses this term.

deformation is equivalent to a contraction in one direction and an expansion in the orthogonal direction. Similarly, we can decompose the quadratic variations of the velocity field. The second order invariants are the gradient of the divergence, the gradient of the rotation and the double deformation. These invariants are plotted in fig. 1–7, bottom row. The vector field of the gradient of the divergence shows a contraction on the left side and an expansion on the right side: thus the divergence is not constant, but changes from left to right, resulting in a gradient. The vector field of the gradient of the rotation shows a clockwise rotation on the top side and an anticlockwise rotation on the bottom side: thus the rotation changes from top to bottom, resulting in a gradient. The double deformation is hard to understand intuitively, which is probably the reason why I am the first to use it for shape perception.

I discovered that the absolute value of the shape index can be calculated from the length of the double deformation and the length of the sum of the two gradients (gradient of divergence and gradient of rotation). I could not recover the sign of the shape index (i.e. whether the object is convex or concave) from the velocity field alone, because I effectively used parallel projection. The sign might come from active movements of the observer. I described some evidence for this in chapter 4. Further I discovered that some other shape characteristics, like the direction of maximal curvature, can also easily be derived from the invariants. In the calculation I had to make an approximation equivalent to using parallel projection. This approximation is reasonable for everyday vision but can be made to cause a breakdown of shape perception in experimental situations. It would be interesting to test whether this breakdown does actually occur in human vision. Further, my way of calculating the shape index is easily implemented on computer and thus could be used in image processing. For example, one might use it for extracting the shape of an object that is moving relative to a camera, in those cases where the movement of the camera is not known.

After a successful solution of a problem, at least 10 new questions jump to ones mind and my research is no exception. I see several ways the research of chapter 4 could be extended. On the experimental side it would be interesting to measure eye movements (rotations of the eyes in their sockets) during general head movement. Although this seems easy, it has actually not yet been done, mainly for technical reasons. By measuring eye movements we could strengthen the conclusions of the study considerably, because the explanation is based on the hypothesis that fixation of the object is different in the different movement conditions. With an eye to the theory developed in chapter 5 we could try to test the hypotheses derived from my theory. Especially interesting is to show slanted curved objects and to add a component to the velocity in forward/backward direction.

The theory of chapter 5 could be extended in several ways. One might try to drop the assumption of a smooth vector field and try to obtain the differential invariants directly from the brightness distribution of a moving object. In everyday life (also a robot's life) we do not see velocity fields as sketched in figure 1–5 but we see the surfaces of objects. Different parts of the object have a different brightness, but the brightness distribution varies smoothly over the surface. When the surface moves, this distribution moves across our retina. To obtain a velocity field from such a changing distribution is not always possible, because the surface might be uniformly white. Also it is not necessary because we are not interested in the velocity field itself but in the differential invariants.

A further (rather straightforward) extension would be to go to third order invariants. These are related to the gradients of curvature and maybe to the parabolic curves on the surface (these curves separate concave and convex parts of a surface). Yet another extension would be to try to model the receptive fields of cells in the visual cortex. There are some interesting mathematical problems in such an attempt. When one would take the vector fields of the pure invariants as receptive fields, one would face the problem that e.g. the detector for the gradient of divergence (lower left panel of figure 1–7) is sensitive to a pure translation (upper panel of figure 1–7). This is so because the vector fields of the zeroth and second order are not orthogonal. In general, one can show that all invariants of the same order are orthogonal and that all invariants of even order are orthogonal to all invariants of odd order. But invariants of different odd and different even orders are not orthogonal. Thus, one would have to devise a different decomposition such that all invariants are orthogonal.

A very interesting extension of both perceptual studies would be to build a dynamical model of shape perception: this would be a linkage between the field of visual perception with the modelling ideas behind the first two chapters. Such a dynamical model of shape perception might cure a major shortcoming of the theory proposed in chapter 5 viz. that it is instantaneous: the solution is only about the current velocity field. Of course, when we observe a moving object the velocity field changes in time. Observers could (and in fact do) use information over time to obtain a better idea of the shape of the object. Realising this, there are many ideas from dynamics one could develop for perception. A very important concept from dynamics is stability. This concept is relevant for perception too, because we know that some percepts are more stable than others. For example, the Necker cube (a line drawing of a cube) allows for two interpretations: which side is front and which is back is not specified by the drawing and thus it is up to us to add this information. When viewing the Necker cube for a prolonged period of time one usually has several reversals between the two interpretations: the percept is not stable. By adding other information, e.g. shading, we can make the Necker cube much less ambiguous and one has only rarely a reversal: the percept is more stable. Thus, the number of depth reversals per time unit could be used as a stability measure, comparable to the fluctuations in relative phase which we used as stability measure in the studies on postural control. This extension of the theory of chapter 5 would be a major one, because one really adds new concepts to the theory.

Chapter 2

Visual control of posture: influence of distance

abstract

I investigated the relationship between a moving visual environment and the induced postural sway experimentally by varying the mean distance to a sinusoidally moving wall. I analysed the temporal relationship (1) in terms of the fluctuations of relative phase between visual and sway motion and (2) in terms of the relaxation time of relative phase as determined from the rate of recovery of the stable relative phase pattern following abrupt changes in the visual motion. The two measures are found to converge to a well-defined *temporal stability* of the action-perception cycle. This stability decreases as the distance of the visual scene from the observer increases. This and the increase of mean relative phase with visual distance are consistent with predictions of a model proposed by Schöner. However, the amplitude of visual sway decreases little as visual distance increases in contradiction to the Schöner-model and suggests that sway is actively generated. The visual expansion rate is found to decrease strongly with visual distance. This leads to the conclusion that postural control in a moving visual environment cannot be understood simply in terms of minimisation of retinal slip, and that dynamic coupling of vision into the postural control system must be taken into account.

This chapter is an expanded version of the paper "Temporal stability of the action-perception cycle for postural control in a moving visual environment" by Tjeerd Dijkstra, Gregor Schöner and Stan Gielen, Experimental Brain Research 97, 477–486 (1994). The data have been reanalysed for this chapter and some results differ from the published paper: in particular less

2.1 Introduction

The contribution of vision to postural stability has been a topic of research for decades [17, 43, 4]. In these studies it was established how the amplitude and frequency content of sway depends on the presence and nature of visual information. More recent research [76, 77] investigated how the temporal structure of postural sway depends on the temporal structure of visual motion. From this work we know that spontaneous sway takes place in a frequency range below about 1 Hz and that frequencies of visual motion must be below 0.5 Hz in order to induce coherent sway. The theoretical view was proposed that the postural control system can be characterised as a linear second order low-pass system with a cutoff frequency at approximately 0.5 Hz [77].

Studying the relative timing of postural sway and motion of the visual scene is crucial if we are to understand how the postural control system is coupled to vision. Unfortunately, there are a number of technical difficulties when this action-perception relationship is to be analysed precisely. In experiments in which an actual moving room is put into motion, little control of its precise timing is possible [41, 70]. However, these stimuli do have the advantage of being realistic which, among other things, is reflected in the small amplitudes of room motion that are sufficient to induce postural sway: 3 mm for a viewing distance of 30 cm for Lee and Lishman and 2.5 cm for a viewing distance of 2 m for Stoffregen. These amplitudes are so small that they are generally not consciously perceived by the subjects. When visual scenes are generated by computer and are displayed on a screen, control over their timing is excellent. However, it is difficult to simulate realistic three dimensional (3D) scenes. For instance, sway of an observer parallel to the screen should lead to displacements of the scene on the optic array of the observer that are larger for objects nearby and smaller for objects far away. This requires measuring the eye position of the observer (i.e. the position of the optic centre of the eye in 3D space) and a fast update of the visual scene by feedback of the eye position of the observer.

Here I report on experiments in which I solved this problem without going way beyond perceptual thresholds. Computer displays of fronto-parallel walls were calculated in real-time based on feedback information from the measured position of the eyes of the observer such as to generate displays that represent walls with a consistent depth. These walls were sinusoidally moved in fore/after direction

---

data are excluded from analysis and the correlations between stability measures are higher. Part of this work has been presented at the International Conference on Event Perception and Action, Amsterdam 1991 and at the International Conference on Posture and Gait, Portland 1992.

in order to induce postural sway of the subject. The situation of a sinusoidally moving wall was chosen because this is interpreted easily in terms of the Schöner-model[1]. It should be noted that neither the model nor the experimental set-up are restricted to this type of stimulus. An additional advantage of this set-up over previous ones, which employed force plates or other apparatus to measure posture, is that this set-up gives the position of the eye in 3D space. This allows us to calculate precisely the optic flow on the optic array of the subjects caused by head movements. In the experiment I varied the mean distance between subject and wall as the experimental variable. Distance is known to play an important role in postural control [41, 55]. Furthermore, in the Schöner-model, distance determines the temporal stability, contrary to e.g. the frequency or the amplitude of the stimulus which have no influence on stability (see eqs. 3.11 and 3.12).

With this set-up I address a number of questions about the temporal relationship of postural sway and wall movement in fore/after direction that are posed by the Schöner-model (see section 3.2). This model characterises the postural state controlled by the nervous system in terms of the position of the eye in the physical environment. Many sensory processes other than vision may contribute to determine this postural state, such as proprioception, vestibular information and pressure sensing in the foot. The idea is that all of these sensory processes contribute to a dynamical system such that the equilibrium postural state (upright posture) is an attractor solution. It is important to note that this dynamical system is not the same as the physical system of the passive biomechanics involved in posture. Instead, the control properties of the nervous system including reflex loops and active control systems, are cast into the form of a dynamical system. The visual influence on postural control is parametrised by the expansion rate of the fixated object on the retina. This expansion rate is formally equal to the inverse of "time-to-contact" independently of object size and distance [42, 75]. The parametrisation of visual influence by the expansion rate is not essential: other parametrisations, like the spatial mean of the optic flow field, will have the same temporal characteristics and lead to the same model predictions. They only result in a rescaling of the coupling constant, which captures the strength of the visual influence on the postural control system.

In the model the assumption was made that the non-visual contributions can be represented by a linear, second order dynamical system. Additive coupling (via a coupling constant) to the expansion rate leads to a linear driven oscillator.

---

[1] I wanted to stick with the structure and order of the original articles. Therefore, the description of the model is given in the next chapter (see section 3.2).

When the visual surround is oscillating, the model predicts that vision makes two contributions to the dynamics (see eq. 3.4). First, it stabilises posture by enlarging the effective damping. The size of this extra stability depends on the distance to the wall and on the coupling constant. Second, vision drives the dynamics with the frequency of the oscillation. The effective amplitude of the drive is the amplitude of the movement of the visual surround divided by the mean distance to the surround. It is important to note that this model *tends* to minimise expansion rate, but does not always do so, depending on the frequency of the visual drive and on the stability (see eqs. 3.8 and 3.9). Driving frequencies that are very different from the eigenfrequency lead to large phase delays and thus to large expansion rates. Low stability leads to more variability in phase delay and thus to larger mean expansion rates. The hypothesis of minimisation of retinal slip has been put forward to explain the observed increase of RMS amplitude of postural movements when the distance to a static surround is increased [41, 55]. These last authors assume a threshold for detection of retinal slip and assume that the postural control system minimises the supra threshold retinal slip. The Schöner-model is consistent with this increase in RMS because the stability in the model decreases with increasing distance. The Schöner-model can be regarded as an extension of the minimisation hypothesis to a dynamical context. The main difference between the hypothesis of retinal slip minimisation and the Schöner-model is that the latter explicitly models the intrinsic dynamics of the postural control system.

The experimental results obtained by van Asten et al. [76, 77] and by Berthoz et al. [4] are compatible with the Schöner-model and a related model had already been proposed by these authors. The Schöner-model stresses the *temporal stability* of the relationship between visual motion and postural sway. Stability is postulated to underlie both the persistence of phase locking between stimulus and response in the face of fluctuations as well as the return to phase locked behaviour following an external perturbation. This postulate and a number of other predictions cannot be tested on the basis of these older data because methods to measure the stability of the relative timing of stimulus and sway were not implemented and no perturbations of that relative timing were performed. Here I calculate two measures of temporal stability from the data: (1) the angular deviation of relative phase evaluated from a time series of relative phase and (2) the relaxation time obtained by determining the time it takes the system to recover its stable relative timing pattern after an abrupt phase shift of the sinusoidally moving visual array. I manipulated the temporal stability by varying the distance between the eye and the visual scene. I tested the concrete model predictions that as visual distance increases:

- temporal stability decreases and both measures of temporal stability covary (see eqs. 3.11 and 3.12).
- the time delay between visual drive and postural response decreases, when the eigenfrequency of sway control is lower than the driving frequency (see eq. 3.7).
- the amplitude of postural sway decreases (see eq. 3.6).
- the expansion rate of the visual surround on the optic array of the observer decreases (see eq. 3.8).

2.2 Methods

2.2.1 Experimental set-up

Red/green stereograms were generated by a SUN4/260 CXP workstation and were projected onto a translucent screen of dimensions 2.5 by 2 m by a Barco Graphics 400 video projector (red phosphor p56, green phosphor p53). The green stimuli were barely visible through the red filter (Kodak Wratten nr 25), transmission being less than 2%, while the red stimuli were invisible through the green filter (Kodak Wratten nr 58), transmission being less than 0.5%. The screen was homogeneously white without any visible texture. The subject wore a pair of goggles that contained the filters and limited the field of view to approximately 120 deg wide by 100 deg high. Due to the restriction of the viewing range, the edge of the screen was not visible to the subject.

The subject stood approximately 50 cm in front of the screen, wearing a flat-topped helmet on which six infra-red light emitting diodes (ireds) were mounted (Fig. 2–1). The positions of these markers were measured with two cameras of a Watsmart system (Northern Digital Inc) at a rate of 400 Hz. The two cameras were placed approximately 2 m behind and 1.5 m above the subject in order to have the best possible signal to noise ratio for detecting movement in a horizontal plane. The walls and ceiling of the room were covered with infra-red absorbing cloth. This same cloth also hung in front of the screen, leaving only a small, door-sized window for the subject to stand. With these precautions I never encountered difficulties with reflections of the infra-red light.

The 2D coordinates of the two cameras were converted real-time into 3D coordinates and sent to the SUN4. This computer was programmed to generate a new stereogram of the wall from the current viewpoint of each eye of the observer, using an algorithm to be explained below. With this set-up every frame (15

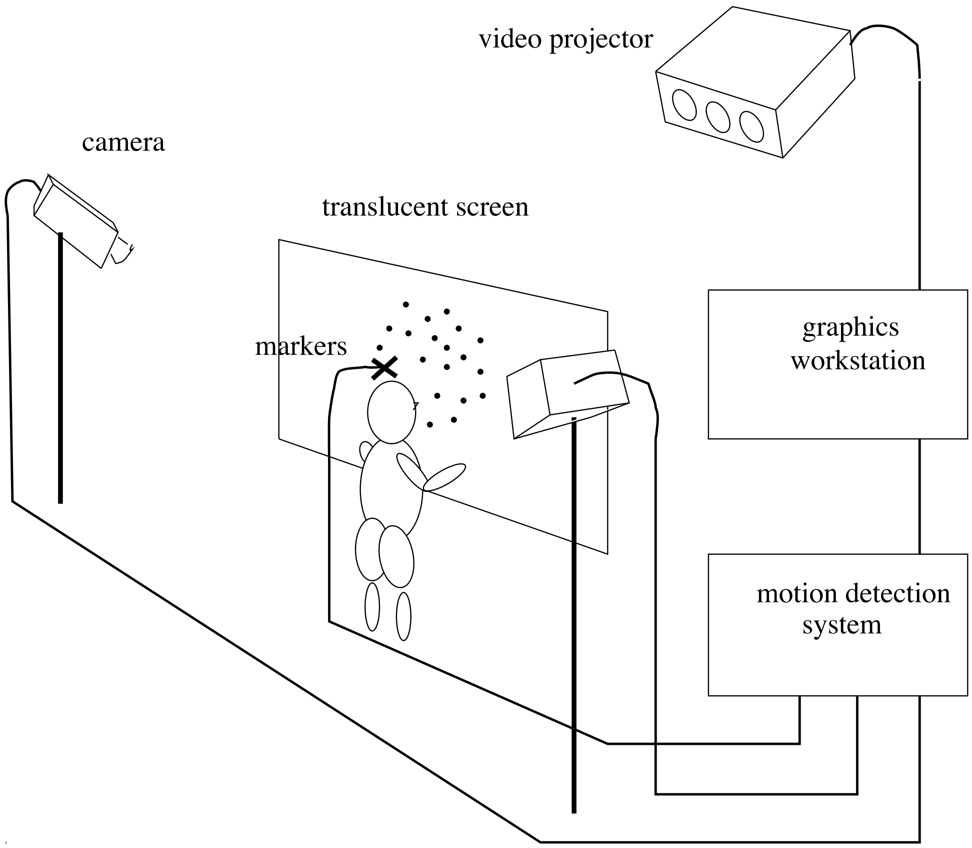


Figure 2–1 Experimental set-up.

ms) provided a new view of the simulated wall. The mean 3D position of the viewpoints, the position of the cyclopean eye, and the orientation of the head were stored together with the position of the stimulus for later analysis. These signals were sampled at a rate of 66 Hz. The delay in the feedback-loop between eye translation and position change of a pixel in the middle of the screen was measured using a turntable and found to be 43±3 ms. The small variability was probably caused by the fact that SunOS is not a real-time operating system. Because of the relatively slow postural movements, this delay was never noticed by the subjects.

The position of each eye in 3D space was calculated using a quaternion algorithm described by Horn [32]. Each session started with a calibration in which the subject faced the cameras and held two additional markers in front of the eyes. The position of all ireds in this configuration was sampled for 200 ms at 200 Hz and from these data the position of each eye relative to each marker of the helmet was calculated. Then the experiment would start with the subject facing the screen. The rotation of the helmet relative to its orientation in the calibration procedure was calculated using a real-time implementation of Horn's algorithm for a planar figure [32, Section 5]. Thus the position of each eye could be calculated during the experiment using the positions of the eyes with respect to the ireds on the helmet. It should be noted that I did not measure the orientation of the eyes relative to the head (rotations of the eyes in their sockets) but the position of each eye in 3D space. Assuming the accuracy of the Watsmart system to scale linearly with the calibrated volume, which in our case was a cube with sides of length 0.6 m, I estimate the accuracy for the position of one marker to be 3 mm [2]. From this I estimate the systematic error of eye position relative to the simulated wall to be of the order of 0.5 cm. The dynamic noise in the eye position was approximately white and had a standard deviation of 1 mm.

2.2.2 Stimuli

The stimuli simulated a fronto-parallel wall covered with 140 stereo dots, each with a size of 0.2 deg by 0.2 deg. The simulation was implemented by keeping a list of the dots in the memory of the graphics processor. The projection matrices (one for each eye) were recalculated every frame based on the most recent positions of each eye and on the sinusoidal translation of stimulus. The density of dots was uniform per solid angle as seen from the view position when the experiment started. Therefore, dot density could not provide a cue on the distance to the wall. The dots lay in an annulus between 10 deg and 60 deg visual eccentricity and thus had a density of 0.056%. The hole in the middle of the stimulus was made to suppress the visibility of aliasing effects, which are most visible in the foveal region. The spatial resolution of the system was 1152 by 900 pixels and the frame rate was 66 Hz.

The wall was suggested to be at distances of 25, 50, 100 and 200 cm. The first is suggested to be in front of the screen, the second on the screen and the last two behind the screen. Distance was suggested both by stereo vision as well as by simulation of the geometrically correct displacement of the image on the screen, using feedback of the eye position of the observer. Subjects reported that stereo vision strongly enhanced the perceptual quality of the display and made viewing for longer times more comfortable. However, it is not essential for our experiment, since in a pilot experiment to compare trials with and without stereo vision I found no significant differences. Subjects might also use accommodation as a distance cue but this cue has been shown to be very weak [25].

The wall was sinusoidally driven with an amplitude of 4 cm and a frequency of 0.2 Hz. The amplitude was chosen so as to give clear responses at a wall-distance of 200 cm. However, I did not want the amplitude to be so large as to be very clearly noticeable by the subjects. The amplitude of 4 cm is a compromise between these two contradictory demands. The frequency was chosen so as to lead to clear in-phase behaviour. Each of the four distances was measured both with and without perturbations, giving a total of eight conditions. The perturbed trials had three perturbations of 180 deg each always at the point of maximum velocity (so velocity was reversed and there was no discontinuity in position, cf. Fig. 2–2) and occurring at random moments in time. Each condition was repeated four times and the order of trials was random. The experiment was conducted in two sessions of approximately 1 hour each and always started with a trial without any visual stimulus for dark adaptation of the subject. In all there were 34 trials for each subject. Each trial lasted for 140 s and the first 20 s were for adaptation and were not stored.

2.2.3 Data analysis

Data analysis was done in two ways: a linear time invariant analysis, which is quite common in physiology [47] and a dynamic approach, as described in e.g. [65]. The data for both types of analysis were the sinusoidal motion of the wall and the response of the subject in fore/after direction sampled at 66 Hz for 2 minutes.

Linear systems theory

In the linear time invariant analysis I calculated the Fourier transform of the drive and the response of the subject. From these I obtained the spectra of the magnitude squared coherence (MSC), the phase and the gain. The MSC is a measure for the strength of locking of the subject to the movement of the wall if the system is linear, or of the relative contribution of the linear part of the system if the system has nonlinear components [11]. All spectra were only evaluated at the driving frequency because there were never any clear peaks at other frequencies. All spectra were calculated with a Welch procedure [48] in order to obtain unbiased estimates of the spectra. For the unperturbed trials I used 15 overlapping segments each 15 s long and a factor 7 zero padding. The MSC depends strongly, the phase and gain very weakly on the number of segments [11]. I only use the variation of the MSC with distance and this is independent of the number of segments. For the perturbed trials I excluded two cycles after a perturbation to assure stationarity (all relaxation times are below 5 s) and for each of the remaining four parts I used 7 segments and a factor 3 zero padding to get approximately the same segment length as in the unperturbed case. I always scaled the Fourier transformed signals in such a way that Parseval's theorem would hold, i.e. that the power in the original signal and in the Fourier transformed signal is the same.

Fourier techniques were also used to determine the amplitude of the expansion rate of the stimulus on the optic array of the subject (the inverse of time-to-contact). I first calculated a time series of the retinal eccentricity of the edge of the stimulus θ at a distance $Z(t)$ by $\tan\theta = X_0/Z(t)$, with X_0 the distance between the centre and edge of the stimulus (physical size). From this I extracted a time series of expansion rate by dividing the time derivative of θ by θ. In the terminology of Tresilian [75] this is the inverse of dilatation tau, denoted by $\tau_l^{(1)}$. This time series of expansion rate, being constructed from a difference of two sinusoidal signals (viz. the motion of the wall and of the subject) is itself a sinusoidal signal at the same frequency as the drive. As a characterisation of the time series of expansion rate I calculated its amplitude by Fourier transforming it and taking the height of the peak in the spectrum, which always occurred at the driving frequency, as the amplitude of the expansion rate. This amplitude differs only by a scaling factor from the r_e of eqs. 3.8 and 3.9.

Time series of relative phase

The dynamic analysis involved calculating a discrete time series of relative phase (phase of response minus phase of stimulus [36]). This discrete estimate of relative phase has the advantage over continuous estimates (using the phase angle in the phase plane) that the signals need not be sinusoidal. From the input and response data the significant extrema both of position and velocity traces were picked using a peakpicker. Before peakpicking, the data were smoothed using a gaussian window with a standard deviation of 0.25 s. The criterion for significance of an extremum was a fraction of the range (i.e. the difference between maximum and minimum) in a segment of one cycle before and one cycle after the extremum. An extremum in position was accepted as significant when it differed more than 40% from the neighbouring extrema. For an extremum in velocity the criterion was 70%. I chose these percentages different because velocity tended to be somewhat more peaked than position. The results depend weakly on the precise values of these percentages.

From the 8 time series of extrema (maxima and minima, both position and velocity of both input and response) relative phase was calculated with the drive as reference and the response as target. This was done by matching each extremum in the reference signal to all extrema in the target signal of the same type and within half a cycle before and half a cycle after the extremum[2]. A relative phase value was calculated by taking the time difference between two extrema of the same type and dividing this by the time difference between two extrema in the reference signal. To this phase value I added 2π times the number of wraps i.e. the difference in number of cycles between reference and target. When there is no extremum in the target signal the number of wraps is decreased by 1. When there is more than one extremum in the target signal the number of wraps is increased by 1 for each extremum except the first. As the time value of this relative phase in the time series I used the time of the target extremum. The result of these manipulations are 4 time series of relative phase: maxima and minima of position and of velocity. Because these four time series were not very different (see results section) I combined[3] these time series in one overall time series of relative phase (see Fig. 2–2 and 2–3). As I have four samples per cycle the mean sampling frequency is 0.8 Hz.

Measures calculated from the time series of relative phase

For all trials I calculated mean phase and angular deviation from the combined time series of relative phase, using circular statistics [3]. These were calculated through:

$$r \exp(i\bar{\phi}) = \sum_n \exp(i\phi_n),$$

with n running over all phase values. The mean phase is $\bar{\phi}$ and the angular deviation, s, is given by $s = \sqrt{2(1-r)}$. Mean phase is a measure of the time delay between input and response and angular deviation is a measure of the stability of the response. For the perturbed trials I excluded the data points in the two cycles after the perturbation, because mean phase and angular deviation are measures for stationary behaviour.

The relaxation time, the time it takes the system to regain its in-phase behaviour after a perturbation, was estimated as follows: I defined a band around

---

[2] Because of this procedure, reference and drive signals are not completely symmetric. Preferably the reference signal is the slower signal.

[3] The combination of time series of relative phase is a somewhat thorny issue because the different time series do not necessarily have the same number of phase values. For instance, a small fast bump might lead to a pair of extrema in velocity but not in position. In such a case I assumed I missed a pair of extrema in position.

the mean of relative phase of the unperturbed part of 1.5 times its angular deviation. After a perturbation the phase generally leaves this band. I calculated reentry time as the time between the perturbation and the time of reentry into the band. The reentry time could also be used as a measure of stability but I found it to be more variable within a trial than the relaxation time. I calculated relaxation time by fitting an exponential to the points outside the band and the first four points inside the band (cf. Fig. 2–2). The fit was implemented by linear regression on the logarithm of the absolute phase values. As a rough quality measure of the fit I took the explained variance of the fit divided by the angular deviation. If the relative phase did not leave the band or if the quality measure was lower than 2, I excluded the perturbation. The relaxation time of a trial was the average of the non-excluded relaxation times in a trial. The relaxation time depends somewhat on the parameters used but the trend with distance does not depend on them. It should be noted that this procedure probably somewhat underestimates the larger relaxation times: these generally occur for lower stability, which means that the band is wider so the system is bound to return faster inside the band.

2.2.4 Subjects

Four subjects with normal or corrected to normal vision and normal stereo vision were tested in all conditions. Three of the subjects were familiar with the purpose of the experiment. The main findings were confirmed by a fifth subject who was naive as to the purpose of the experiment. This subject was not tested in all conditions of the experiment. Subjects were instructed to look at the centre of the stimulus and to stand relaxed; they stood on a firm stable support in normal Romberg posture.

2.3 Results

Results were obtained for four subjects whose order of presentation will be: GS, MG, SG, CK. Unless otherwise noted, all statistical tests are one-way ANOVAs with distance as the independent variable at a significance level of 5%. For post hoc analysis (pairwise comparison) I used the Newman-Keuls procedure [29], also at a significance level of 5%. In order to keep the burden down I averaged over subjects, which is allowed because the assumption of sphericity was never violated [29] .

Generally the subjects responded to the sinusoidally moving wall with an almost sinusoidal postural response in fore/after direction. They moved in-phase

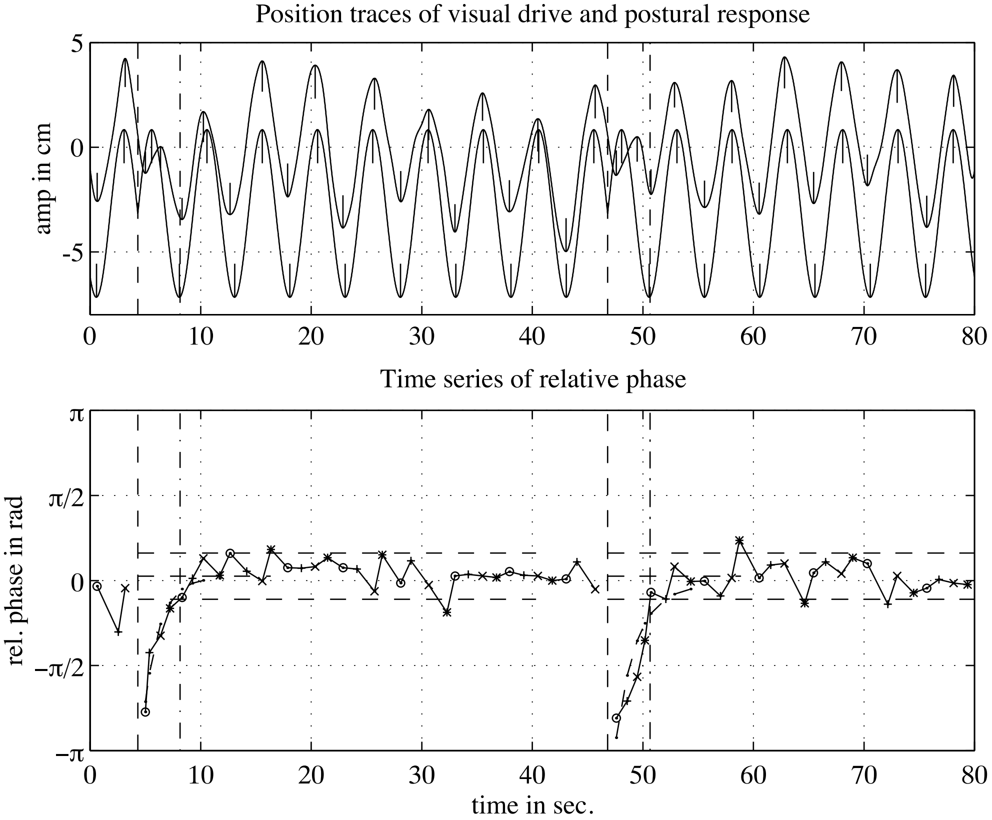


Figure 2–2 Example of a data record for a small mean distance (25 cm) between subject and wall. Upper panel: position traces of visual drive (shifted down for visibility) and postural response. The position of the short solid vertical lines denote the extrema as picked by the peakpicker, their length denotes the criterion used in picking. The dashed vertical line denotes the time of perturbation and the dash-dot line the reentry time. Positive means movement towards the wall (standing more on your toes), negative means away from the wall. Lower panel: derived time series of relative phase (solid line) with exponential fit of the relaxation (dashed line). Symbols denote the different types of phase values: x denotes peaks position, o denotes valleys position, + denotes peaks velocity and * denotes valleys velocity. Also given are the time of the perturbation (dashed vertical line) and the time where the relative phase enters the band around the mean phase (dash-dot vertical line). The two outer dashed horizontal lines after a perturbation are plus and minus 1.5 times the angular deviation around the mean. The middle dashed horizontal line denotes the zero level of the relaxation.

with the wall and with a relatively fixed amplitude of 4 cm, the same as the drive. They did not consciously perceive visual motion except for the condition where the wall was nearest (25 cm) and at the perturbations. In figure 2–2 I have plotted position traces of visual motion and postural response and the derived time series of relative phase for a small distance between subject and wall. The subject is clearly phase locked to the visual motion. The temporal stability is high as exemplified by the fast relaxations and the small variability. In figure 2–3 I have plotted the results for a trial with a large distance between subject and wall. The subject is still phase locked but not so stable as in the previous example. Also notice that the subject is somewhat delayed in his response relative to the visual motion.

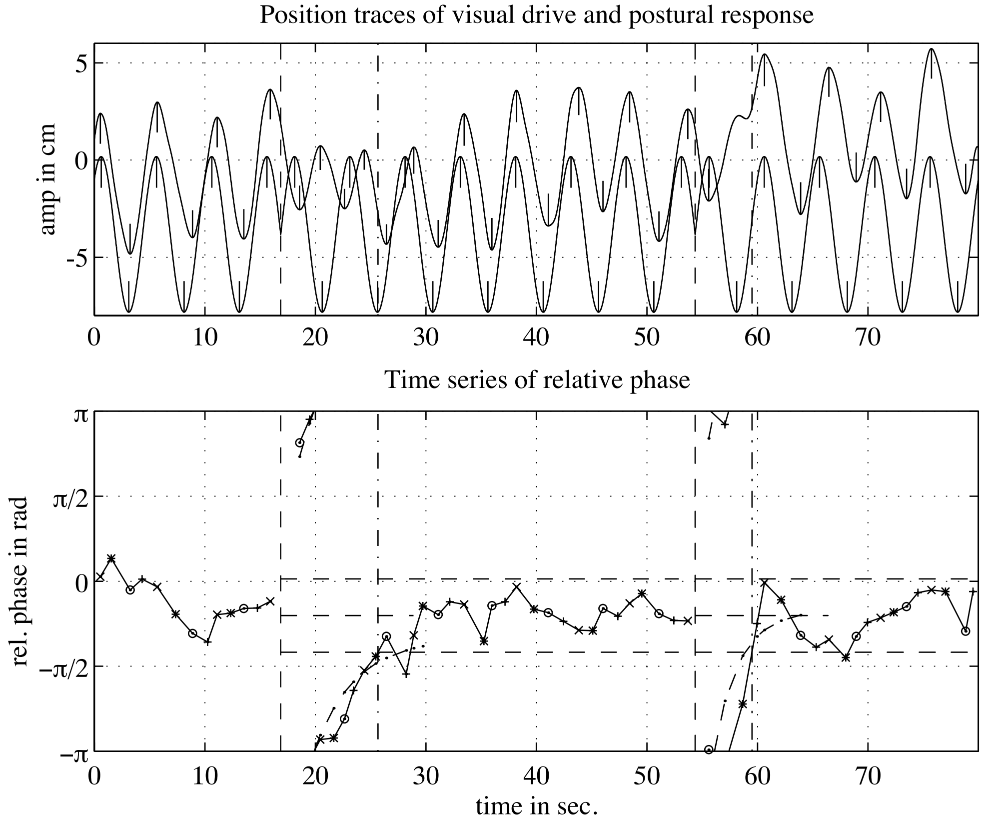


Figure 2–3 Example of a data record for a large mean distance (200 cm) between subject and wall. For a detailed legend see the previous figure.

Those parts of the trials where the response did not show a clear phase locked pattern were excluded from the analysis. I excluded these data because I focused on local stability measures in this experiment. In the next chapter I will consider both local and global stability. Also, the proportion of data excluded was relatively small: for the four subjects it amounted to 7%, 22%, 0% and 2%, respectively. Only for subject MG a considerable fraction was excluded: we will see below that this subject shows behaviour that is sometimes different from the others, possibly because I excluded a lot of the unstable trials at a distance of 200 cm. Most of these exclusions occurred in the first few trials after the start of the experiment or after the midway break and for a suggested distance of 100 or 200 cm. I interpret the fact that the trials with non-oscillatory parts occurred primarily subsequent to a break as a sign of long term adaptation. The detailed study of such adaptation effects is, however, beyond the scope of this contribution. In relation to the Schöner-model we note that a decreased phase locked response is expected at larger distances: stability is so low that phase locked behaviour sometimes does not occur.

As a descriptive measure for our data I calculated the RMS of the position and velocity signals of subject movement in fore/after direction. Neither of these changed significantly with distance for any of the subjects. Averaged over all subjects the RMS of the position signal at a distance of 200 cm decreased by 13% relative to the RMS at 25 cm. For the velocity the decrease was only 2%. The other signals (left/right and up/down translation of the eye and orientation of the head) showed little variation. The left/right translation was by far the largest of the other signals with a RMS of approximately 8 mm.

The mean and angular deviation of the four time series of relative phase were generally significantly different as revealed by two-way ANOVAs with distance and type (minima or maxima of position or velocity) as independent factors. For an example of this difference see Fig. 2–2 lower panel, where the estimates of relative phase after the perturbation based on velocity are more variable than the estimates based on position. The differences caused by type were not consistent across subjects except that the angular deviation of the relative phase calculated from the velocity signals was higher. This is to be expected since velocity is derived from the position by differentiation, a noise enhancing procedure. The ANOVAs revealed no interaction between type and distance and as I am interested in the effect of distance I combined the four time series. Further, after a perturbation the combined time series of relative phase generally is nicely approximated by an exponential (Fig. 2–2 lower panel) indicating that it is reasonable to combine the individual time series.

All statistical parameters calculated from the stationary part of the perturbed trials and from the unperturbed trials were never significantly different, so the perturbations did not change the stationary behaviour. Therefore I only show the results of the perturbed trials.

2.3.1 Stability

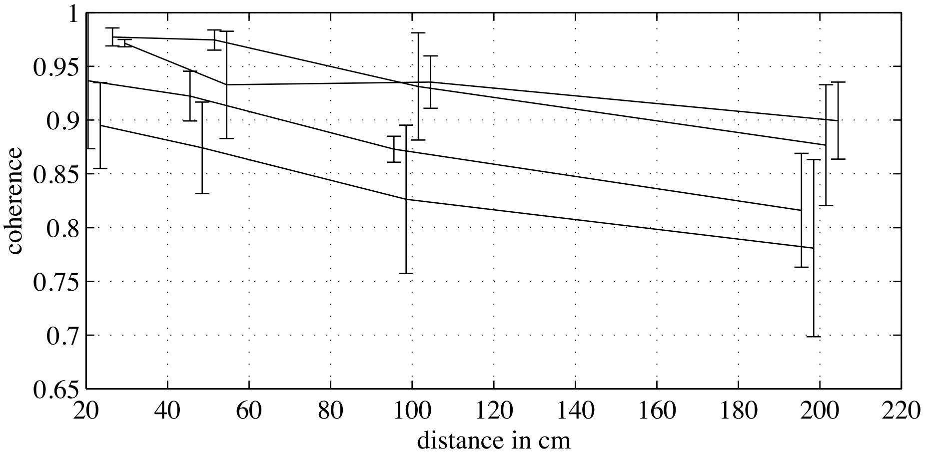


Figure 2–4 The magnitude squared coherence as a function of distance. The data points for each subject are slightly displaced horizontally for clarity. The ordering of the subjects from left to right is: GS, MG, CG, CK and will be the same in subsequent plots. The vertical bars denote the standard deviation.

In Figures 2–4 to 2–6 three different measures for the temporal stability are plotted. All three indicate a decrease of stability with increasing distance. The MSC (Fig. 2–4) decreases with increasing distance. This decrease is significant for subjects GS and SG. The mean decrease over all subjects between 25 cm and 200 cm is approximately 10%. Post hoc analysis revealed all pairs of conditions to be significantly different except the 25–50 cm pair.

The increase in angular deviation of relative phase (Fig. 2–5) is significant for three subjects, but not for subject SG. Also note that the stability across subjects as reflected in the angular deviation shows almost exactly the same pattern as reflected in the MSC: subjects SG and CK are strongly phase locked to the stimulus, GS is in the middle and subject MG shows the weakest phase locking. This is also reflected in the correlation between MSC and angular deviation: -0.75, -0.84, -0.78 and -0.50, respectively. The mean increase over all subjects in angular deviation from 25 cm to 200 cm is approximately 45%. Post hoc analysis showed only the 200 cm condition to be significantly different from all others.

The relaxation after a perturbation generally shows a clear exponential decay in phase. Figures 2–2 and 2–3 show two data records and the derived time series of

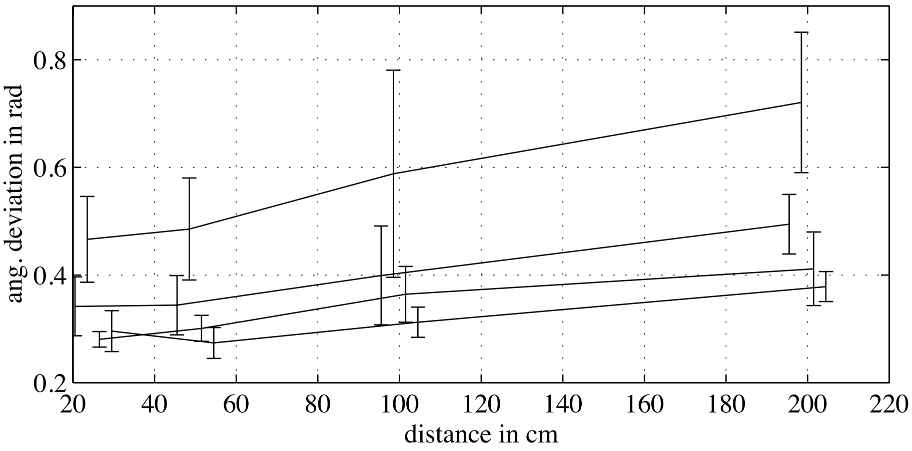


Figure 2–5 The angular deviation of relative phase as a function of distance. For details see legend of Fig. 2–4.

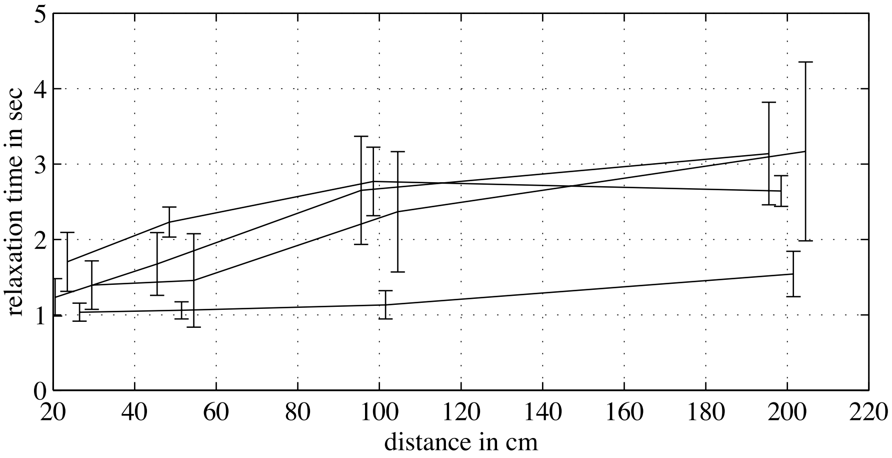


Figure 2–6 The relaxation time as a function of distance. For details see legend of Fig. 2–4.

relative phase. Figure 2–2 is from a trial with a small distance (25 cm) between subject and wall and shows a large stability as reflected in the fast relaxation after the perturbation and in the small angular deviation. Figure 2–3 is from a trial with a large distance (200 cm) between subject and wall and shows a small stability as reflected in the slow relaxation after the perturbation and in the large angular deviation. Those cases that did not show this exponential decay were excluded. For the four subjects the percentage of perturbations excluded was 17%, 44%, 6% and 6%, respectively. Roughly half of these exclusions were caused by non-oscillation as discussed before. The remaining exclusions almost always occurred for a distance of 100 and 200 cm, where stability is lowest. The exclusion was usually caused by the fact that it took the subjects a very long time (larger than 15 s) to return to in-phase behaviour. During this period the phase would drift considerably, causing our exponential fit to be a very bad approximation. The increase in relaxation time (Fig. 2–6) is significant for three subjects, not for MG. The mean increase over all subjects from 25 cm to 200 cm is approximately a factor 2. A separate ANOVA to test for an effect of the order of the perturbation within a trial revealed no effect of order on relaxation time. This is consistent with the fact that I never found any significant difference in the statistics between perturbed and unperturbed trials. Post hoc analysis revealed all pairs of conditions to be significantly different except the 25–50 cm pair.

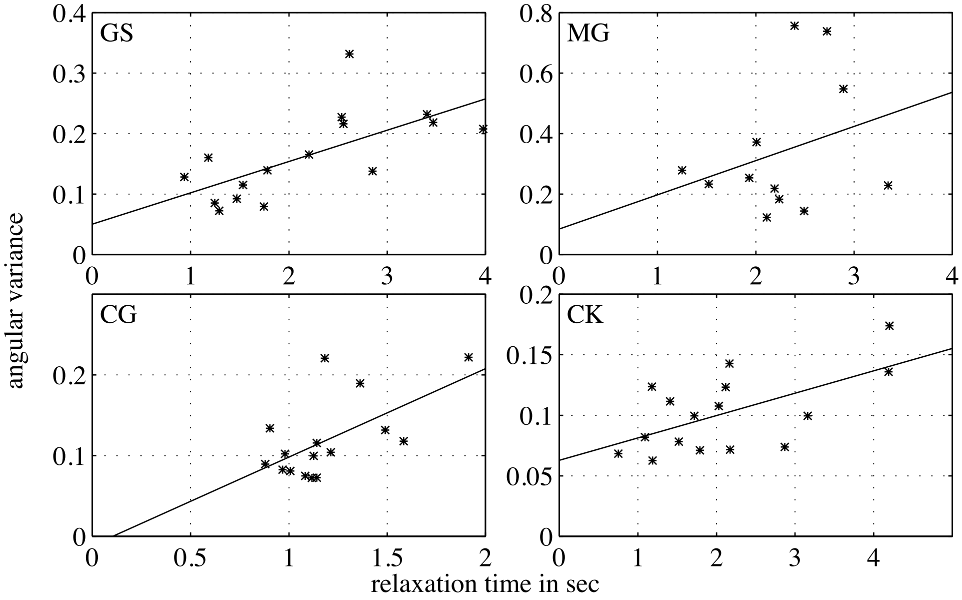


Figure 2–7 Scatter plot of relaxation time and angular variance of relative phase. The four panels are for four different subjects. The solid lines denote the best fit from a linear regression.

The Schöner-theory predicts a covariance between angular variance and relaxation time. Figure 2–7 gives a graphical impression of this covariance. The correlation coefficients for the four subjects are 0.67, 0.29, 0.61 and 0.62, respectively. Only the coefficient of subject MG is not significantly different from zero. This is probably caused by the small number of included relaxations at a distance of 2 m (2 out of 12). The temporal stability of this subject at this distance is so low that he often failed to return to a stable in-phase pattern after a perturbation within a few cycles. This of course biases the correlation, because I only take the fast relaxations[4].

2.3.2 Delay

In Figure 2–8 and 2–9 two different measures for the time delay between visual drive and postural response are plotted. In Figure 2–8 the delay as calculated from the phase spectrum is plotted, whereas in Figure 2–9 the mean phase difference as calculated from the time series of relative phase is plotted.

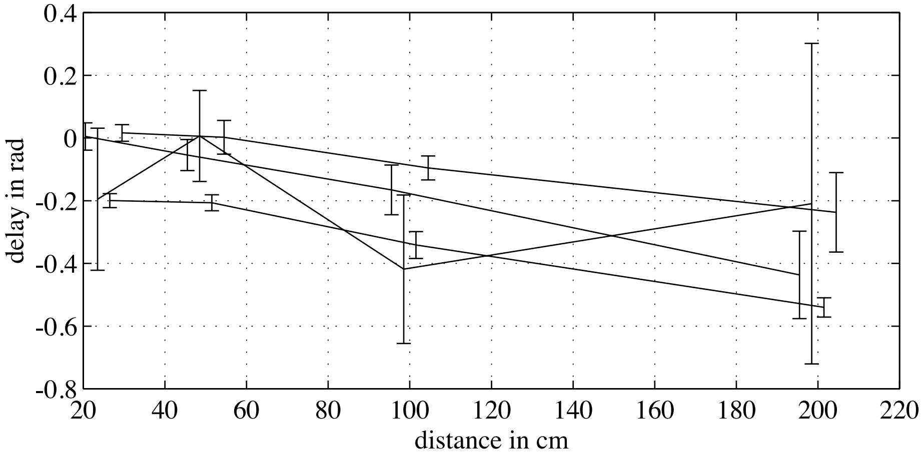


Figure 2–8 The delay calculated by Fourier analysis as a function of distance. For details see legend of Fig. 2–4.

Both indicate an decrease of delay with increasing distance. There is a significant effect of distance on both measures for three subjects, not for MG. The correlations between delay as calculated from the phase spectrum and as calculated from the time series of relative phase are large: 0.93, 0.97, 0.99 and 0.98,

---

[4] In section 3.3.2 I use a more advanced procedure to calculate the relaxation time. Using this procedure and *not* excluding any data, I find the following correlations: 0.76, 0.80, 0.83 and 0.71, respectively. I did not have to exclude any data because the algorithm to calculate the relaxation time of section 3.3.2 is much more stable than the algorithm I employed here. Also, the increase of relaxation time with distance is much stronger with the new algorithm and is significant for all subjects.

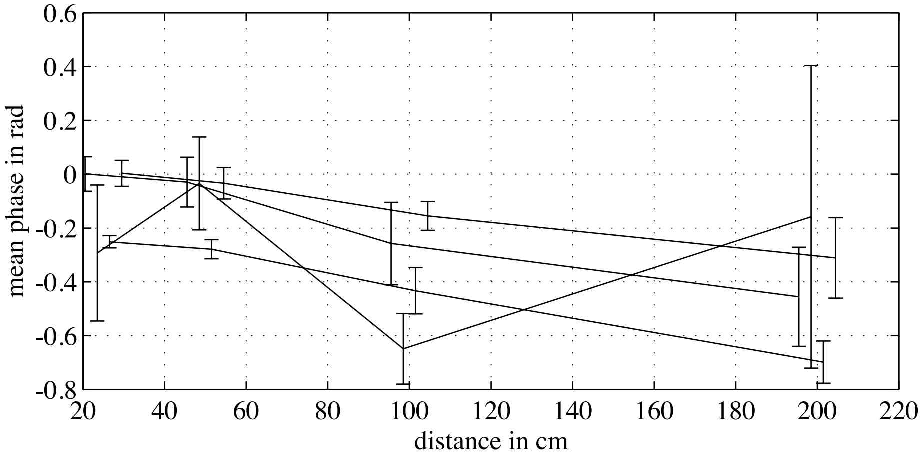


Figure 2–9 The delay calculated from the mean of the relative phase time series as a function of distance. For details see legend of Fig. 2–4.

respectively. This indicates that the procedure I used to calculate relative phase, leads to sound results. Post hoc analysis showed all pairs of conditions to be significantly different except the 25–50 cm pair and the 100–200 cm pair.

2.3.3 Gain and optic flow

In Figure 2–10 the gain is plotted as a function of distance. It shows a slight decrease which is significant only for subject SG. Further the gain is near 1, which means that the amplitude of the response of the subjects is the same as the amplitude of the drive. Especially for the smaller distances the gain is even larger than 1, indicating overcompensation for the perceived ego motion. Post hoc analysis showed only the 200 cm condition to be significantly different from all others.

Figure 2–11 shows the amplitude of the expansion rate of the visual surround on the optic array of the subject. This expansion rate is calculated using the sampled movements of the subjects relative to the movements of the wall and therefore depends on the distance to the wall, and on the gain and the delay of the response. For reference I also include a curve which gives the expansion rate if the subject would not move. The expansion rate decreases significantly for all subjects. Note that while the gain decreases by approximately 20% with increasing distance, the expansion rate decreases by approximately 70%. Post hoc analysis revealed all conditions to be significantly different from one another.

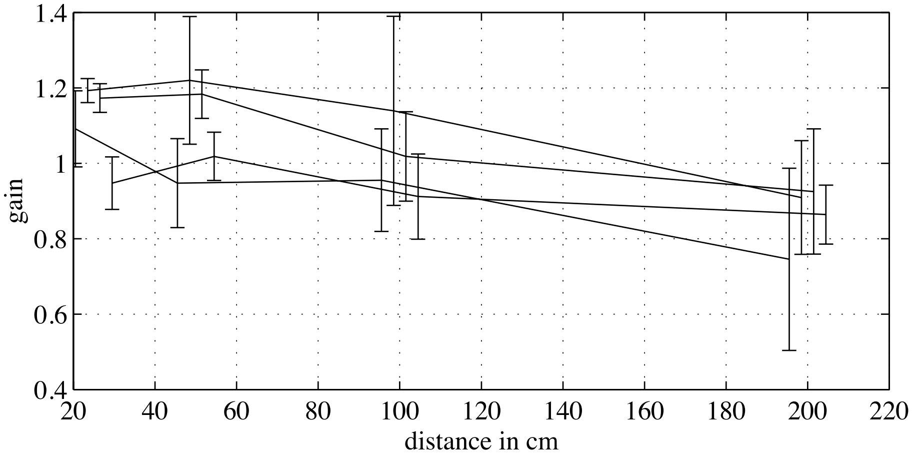


Figure 2–10 The gain as a function of distance. For details see legend of Fig. 2–4.

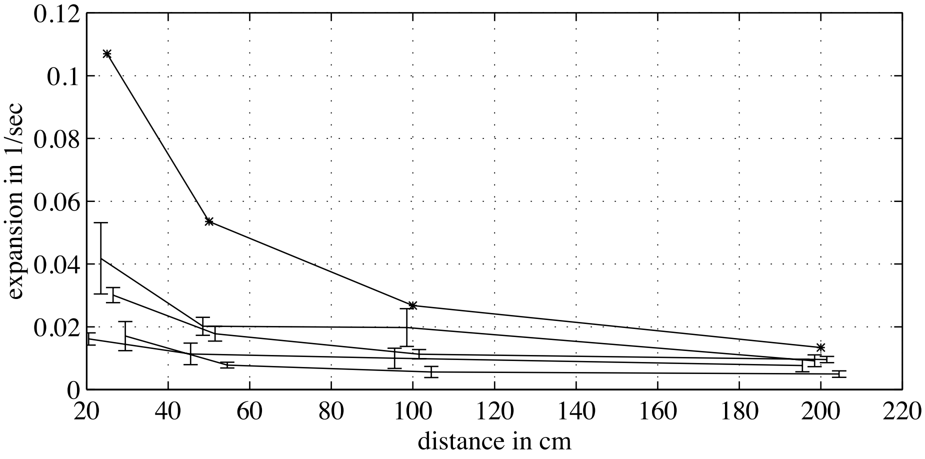


Figure 2–11 The expansion rate of the visual surround on the optic array of the observer as a function of distance. The upper curve gives the expansion rate if the subject would not move. For details see legend of Fig. 2–4.

2.4 Discussion

I reported results of experiments on visually induced postural sway in which information from the measured position of the eyes is fed back to the display computer such as to generate scenes with realistic geometry and consistent depth.

I focused on the temporal relationship between stimulus movement and postural response. To characterise this relationship I measured the relative phase between sway movement and the sinusoidal motion of the visual scene. This provides me with a time series characterizing the temporal evolution of this relationship. I manipulated visual distance because this parameter is predicted to affect the temporal stability of the action-perception cycle.

I found an enhanced variability of the timing of sway relative to the sinusoidal motion of the visual surround as the distance to the wall increased. Further, I found a larger relaxation time after a phase perturbation of the sinusoidal motion of the visual surround at larger distances. The correlation between the two measures was about 0.6. This shows that processes underlying recovery from perturbations and the processes controlling fluctuations of the unperturbed postural control system are governed by the same temporal stability. In the model of Schöner there is an exact relation between angular variance and relaxation time. Because the systematic change in temporal stability induced by our distance manipulation is not very large compared to the fluctuations in the angular deviation and relaxation time (see the large error bars in Fig. 2–5 and Fig. 2–6), we cannot, on statistical grounds, expect a large correlation between these two stability measures. Beyond statistics, I have no firm explanation for this low, although significant, correlation. In a separate analysis I ruled out one possible explanation by finding no significant effect when I tested whether the phase value just prior to the perturbation might influence the relaxation time. An explanation that remains possible is the fact, already mentioned in the Method section, that the procedure to calculate the relaxation time may well underestimate relaxation time for higher values of the angular deviation. It should be noted that our analysis of relaxation presumes the existence of a stable attractor (in-phase behaviour). When this attractor is very weak, as in the case of the excluded perturbations, one might expect the relaxation to be strongly influenced by noise. Thus the fact that I had to exclude some of the relaxations when stability is low, is qualitatively in agreement with the model. Generally, such exclusions are conservative with respect to the hypothesised decrease in stability

The decrease of stability is also reflected in the Magnitude Squared Coherence (MSC), although somewhat less sensitively. The observed decrease of temporal stability as visual distance increases is consistent with the predictions of the model. In the Schöner-model, this decrease is caused by the decrease of the amplitude of visual expansion rate (also experimentally observed, see Fig. 2–11), which leads to a reduced effective coupling strength (the c' in eq. 3.4) of posture to the time structure of visual motion.

Second, I showed a decrease of the delay between sway and visual motion as visual distance is increased. This observation is likewise consistent with the Schöner-model. In the model, this effect leads to the hypothesis that the eigenfrequency of the sway control system is lower than the driving frequency for all subjects. Eigenfrequencies of 0.2 to 0.15 Hz have been reported before by van Asten et al. [76]. Surprising is the small range of eigenfrequencies that is compatible with the data: eigenfrequencies larger than 0.2 Hz lead to a system that has a phase lead relative to the drive and eigenfrequencies below 0.15 lead to a steep decrease in delay. Only the small range of 0.16 to 0.19 Hz leads to reasonable results. So despite the considerable difference in biomechanics between the subjects[5], the eigenfrequency is strongly constrained in this experiment.

Third, there was no significant change of the amplitude of postural sway as visual distance was varied. Sway amplitude always closely matched the amplitude of the visual motion. I have found this result before [18]. The slight, non-significant decrease of the gain is not by far as strong as predicted by the Schöner-model and therefore is quantitatively in contradiction with it. From an ecological viewpoint this result makes sense: subjects correct their posture by matching the amplitude of their egomotion with the visual motion, irrespective of the distance to the visual surround.

Finally, I found that the expansion rate of the visual surround on the optic array of the observer was not constant. This contradicts the hypothesis that the system controls posture purely by minimizing the expansion rate of the visual surround [55]. For if this were the case, the expansion rate would always be at threshold and would not depend on distance as I found. Instead, this result indicates that the relation between visual motion and postural control must be viewed dynamically: the postural control system tends to minimise retinal slip, because the observed slip is smaller than when the subject does not move. However, this tendency also depends on the distance to the stimulus.

To examine the relationship of experimental results and the Schöner-model more quantitatively, I have estimated the order of magnitude of the model parameters. Such estimates can be based on the observed dependence of relaxation time, mean and angular deviation of relative phase on distance (cf. eqs. 3.7, 3.11 and 3.12). Proceeding in this manner, I found: $\alpha = 0.5$ Hz, $\omega_0 = 2\pi\, 0.18$ rad/s, $c_{env} = 50$ cm/s. In Fig. 2–12 I have plotted the resulting dependencies of the various observables on distance. Clearly, amplitude is predicted to decrease much more strongly than observed, while all other relationships are captured quite well.

It is a general property of driven linear systems that the response amplitude decreases as the coupling to the driving force decreases. If we assume that the

---

[5] Two normal-sized dutchmen and two short germans.

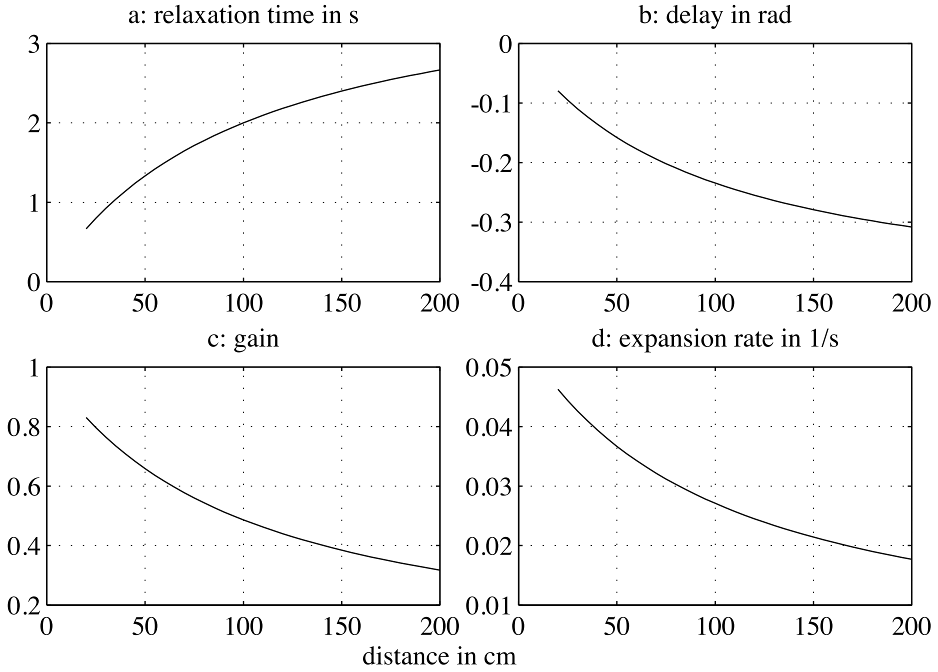


Figure 2–12 Predictions of the Schöner-model as a function of distance to the wall for (a) relaxation time (b) delay (c) gain (d) expansion rate.

visual expansion rate is the driving force, then relative timing stability decreases concomitantly. Therefore, the fact that this linkage of amplitude and stability is not observed in the present data hints that the theoretical picture of posture in a visual environment as a passive linear system driven by the expansion rate, is not adequate. Instead, the central nervous system might actively generate movements (corresponding, mathematically, to nonlinear dynamics possessing limit cycle attractors) which match the visual motion in amplitude and frequency. This active movement may then be coupled dynamically to the visual information, consistent with the successful model predictions for the timing aspects of the data.

Chapter 3

Visual control of posture: influence of frequency

abstract

In this experiment I varied the frequency of an oscillatory visual display and analysed the temporal relationship between visual motion and induced postural sway in terms of mean relative phase and its temporal stability. I found that subjects maintain sizeable sway amplitudes even as temporal coherence with the display is lost. Postural sway responses tended to phase lead (for frequencies below 0.2 Hz) or phase lag (above 0.3 Hz). However, I also observed at a fixed frequency highly variable phase relationships in which a preferred range of phase lags is prevalent, but phase jumps occur that return the system into the preferred range after phase has begun drifting out of the preferred regime. By comparing the results quantitatively with a dynamical model (the sine-circle map) I show that this effect can be understood as a form of relative coordination and arises through an instability of the dynamics of the action-perception cycle. Because such instabilities cannot arise in passively driven systems I conclude that postural sway in this situation is actively generated as rhythmic movement which is coupled dynamically to the visual motion.

This chapter is an expanded version of the paper "Frequency dependence of the action-perception cycle for postural control in a moving visual environment: relative phase dynamics" by Tjeerd Dijkstra, Gregor Schöner, Martin Giese and Stan Gielen, Biological Cybernetics 71, 489–501 (1994).

3.1 Introduction

The stabilisation of posture has been intensely studied over the last decades not only because of the clinical relevance of its understanding, but also because this basic behaviour can serve as a model system for multisensory integration. A prominent approach has been to look at the contribution of various sensory systems to the reduction of sway amplitude and to the temporal structure of sway. In particular, the role of visual information has thus been analysed and it was found that visual information stabilises posture under normal conditions [17], but may destabilise posture if brought into conflict with the stationary environment as sensed by other sensory channels [41, 43]. By varying the spatio-temporal structure of visual information the response properties of the postural system have been identified [76, 77].

In the previous experiment, where the distance of a moving visual scene from the observer was varied, I confirmed the predictions from the Schöner-model. However, important deviations from the mathematical form of the model (linear driven oscillator) were observed which I interpreted as indicative of an active contribution of the postural control system in response to visual motion. In this chapter I report experimental and modelling work that aimed to look afresh at the problem of how visual information is integrated into posture behaviour. I went back to the classical paradigm of characterising the response properties of the postural system to visual stimulation by varying the frequency of sinusoidal visual motion. Previous studies that also varied the frequency of visual motion [43, 72, 76] have concluded that the postural control system can be characterised as a second order system, passively driven by the visual motion. Linear systems theory was critically evaluated by Talbott and coworkers studying postural stability in dogs [72, 67, 73]. From a series of experiments in which dogs stand on a moveable table and are exposed to a moving visual environment these authors concluded that the influence of vision on posture is strongly task dependent, a conclusion in line with our present concerns. However, while gain was shown to be strongly nonlinear in response to both table motion and visual motion, no clear nonlinearities in response to visual motion alone were found. Also, they never found any significant response at frequencies other than those used for input, supporting their conclusion that the postural control system is essentially linear. Our aim is to challenge the view in these studies, that the postural system is passively driven by visual motion. Another way to put this is to ask if the influence of visual motion on posture relies on the existence of a coherent temporal relationship between postural sway and visual motion. If posture is essentially a passive control system driven by visual motion (or other sensory channels) then postural sway must reflect the time structure of sensory input whenever such input affects posture. If, on the other hand, postural sway is actively generated based on perceived conditions then input may affect posture even as temporal coherence is lost. Specifically, in the active case, periodic visual motion may induce periodic sway of adequate amplitude even if sway and visual motion are not phase locked.

To answer this question I have developed techniques which enable me to analyse the system in a regime where coherence between visual information and postural sway is lost. These techniques rely on the extraction of a relative phase time series as a measure of the temporal relation between postural sway and visual motion. These time series are analysed with respect to stability and dynamic properties through return maps and histogram techniques. Furthermore, a concrete model (stochastic sine-circle map) is fitted to the relative phase data in various ways. The stochastic sine-circle map has a dynamics which is rich enough to capture both phase locked (linear) and non-locked (nonlinear) behaviour. I will show that I can reliably fit the stochastic sine-circle map to the relative phase data. In addition, more common frequency domain methods are applied as well.

I exposed standing subjects to a simulated fronto-parallel wall, which was moved in foreward-backward direction with different frequencies (0.05–0.5 Hz). I covaried the amplitude of the movement with frequency in order to keep retinal velocity constant. I kept the amplitudes small (and thereby visual motion roughly at detection threshold) because I believe these amplitudes to be relevant for posture. This should be compared to Lestienne et al. [43] and van Asten et al. [76] who generally employed much larger amplitudes, that are more relevant for walking or running. I found that in an intermediate frequency regime the sway displays the typical phase locked characteristics observed earlier. However, outside this regime coherence sometimes breaks down and other types of coordination between postural sway and visual motion were observed. I observed both relative coordination, where there is a preferred phase relationship without stimulus and response being phase locked, and absence of any coordination. These phenomena can be explained by an instability of the underlying phase dynamics and its observation represents evidence for an effect of visual motion onto the temporal structure of postural sway in the absence of phase locking. No passively driven linear system can display relative coordination (because amplitude goes to zero when the system approaches the instability).

3.2 Dynamic models of the action-perception cycle

3.2.1 The Schöner-model

As a point of reference I review the dynamical model proposed by Schöner [66]. The model is based on the following assumptions: (1) The state of the postural control system can be described by the position, x, of the eye measured in foreward-backward direction. (2) Without vision (eyes closed) the posture control system generates a stable fixed point of this variable, here at $x = 0$ by choice of coordinates. The dynamics of x without vision, *the intrinsic dynamics*, is that of a second order linear system. (3) Visual information couples additively into this dynamics through the expansion rate, $e(x, t)$, of the visual surround. The significance of these assumptions is perhaps best brought out by considering some alternatives. In the discussion I will come back to some of these alternatives. The purpose here is to give the reader a feeling for the meaning these assumptions.

As an alternative to the first assumption one might suppose that the state is completely described by the position x and the eigenfrequency ω_0. The eigenfrequency would then not be a constant as in the Schöner-model. Adding the eigenfrequency to the intrinsic dynamics means that one assumes that the system knows its current eigenfrequency and that the dynamics of this variable is on the same time scale as the dynamics of the x (if it were on a different time scale on could regard the dynamics of x and ω_0 as decoupled). As an alternative to the second assumption one might consider the dynamics to have the character of a limit cycle. This means that the posture control system generates spontaneous sway of fixed frequency and amplitude even without vision (but note that because of the noise this fixed frequency and amplitude might be smeared). As an alternative to the third assumption one might suppose that the expansion rate couples into the dynamics in a more complex way e.g. the expansion rate might also influence the damping in such a way that damping becomes larger when the expansion rate is large. This would have the effect of returning the subject faster to an in-phase pattern after a perturbation. In general, the Schöner-model constitutes a minimal model that is consistent with the experimental facts.

Returning to the Schöner-model, we have mathematically:

$$\ddot{x} + \alpha\,\dot{x} + \omega_0^2\,x - \sqrt{Q_x}\,\xi_t = -c_{env}\,e(x,t), \tag{3.1}$$

with parameters, damping coefficient α, eigenfrequency ω_0 and noise strength, Q_x, where ξ_t is gaussian white noise (zero mean, unit variance). The left-hand side of this equation models the intrinsic dynamics as a linear damped harmonic oscillator. The noise models the random fluctuations of the posture system. In modelling the noise as additive gaussian white noise the following assumptions have been made: (1) many independent noise sources, (2) the noise sources are correlated over short time scales compared to the deterministic dynamics and (3) the noise sources are not zero at particular values of the dynamical variables. The expansion rate is given by:

$$e(x,t) = \frac{\dot{x} - \dot{D}(t)}{x - D(t)}, \tag{3.2}$$

with $D(t)$ the movement of the visual surround. The expansion rate gives the rate of expansion of the visual surround on the optic array of the subject. It depends both on the velocity of the subject and on the velocity of the surround. It couples with strength c_{env} such as to stabilise posture. For a sinusoidally moving surround, moving at a mean distance D_0 with an amplitude D_r and frequency ω_D, we have:

$$D(t) = D_0 + D_r \sin(\omega_D t). \tag{3.3}$$

Putting these equations together, we obtain a linear driven harmonic oscillator:

$$\ddot{x} + \tilde{\alpha}\,\dot{x} + \omega_0^2\,x - \sqrt{Q_x}\,\xi_t = c'\ \cos(\omega_D t), \tag{3.4}$$

with the effective damping $\tilde{\alpha} = \alpha + c_{env}/D_0$ and the effective driving amplitude $c' = c_{env}\,\omega_D\,D_r/D_0$. Note that vision makes two contributions to the dynamics: it increases the effective damping (and thus increases stability, see eqs. 3.11 and 3.12 below) and it drives the posture system with frequency ω_D. The increase in damping depends on the coupling strength and the visual distance; the effective amplitude depends also on amplitude and frequency of the visual drive. Because the system is linear we find that the postural response, the asymptotic solution ($t \gg \tilde{\alpha}$) of the deterministic part:

$$x(t) = r_0 \sin(\omega_D t + \phi_0), \tag{3.5}$$

is harmonic with the same frequency as the visual motion. Amplitude and relative phase are given by:

$$r_0 = \frac{c_{env}\,\omega_D\,D_r/D_0}{\sqrt{(\omega_0^2 - \omega_D^2)^2 + (\tilde{\alpha}\omega_D)^2}}, \tag{3.6}$$

$$\tan\phi_0 = \frac{\omega_0^2 - \omega_D^2}{\tilde{\alpha}\,\omega_D}. \tag{3.7}$$

The dependence of the amplitude r_0 and relative phase ϕ_0 on the driving frequency ω_D is well-known from systems theory. The dependence on visual distance D_0 is somewhat unorthodox and is plotted in fig. 2–12. We can use these results to calculate the amplitude of the expansion rate. Substituting eqs. 3.6, 3.7, 3.5 and 3.3 in 3.2, we find that the expansion rate is harmonic with the driving frequency. Its amplitude is given by:

$$r_e = \omega_D \, \frac{D_r}{D_0} \sqrt{1 - \frac{(\tilde{\alpha} - \alpha)^2 \omega_D^2}{(\omega_0^2 - \omega_D^2)^2 + (\tilde{\alpha}\omega_D)^2}}. \tag{3.8}$$

The dependence of r_e on visual distance is plotted in fig. 2–12. When the eigenfrequency is close the driving frequency $|\omega_0^2 - \omega_D^2| \ll \tilde{\alpha}\omega_D$, this can be approximated as[1]:

$$r_e = \omega_D \, \frac{D_r}{D_0} \, \frac{\alpha}{\tilde{\alpha}}. \tag{3.9}$$

The dynamics of the action-perception cycle can be studied by looking at solutions of the form $x(t) = r(t) \sin(\omega_D t + \phi(t))$ so that the relative phase $\phi(t)$ expresses the temporal relationship between sway and visual motion. A dynamics of relative phase can be derived by transforming eq. 3.1 into polar coordinates and applying the averaging method [27]:

$$\dot{\phi} = \frac{\omega_0^2 - \omega_D^2}{2\omega_D} - \frac{c_{env} D_r}{2 r_0 D_0} \sin\phi + \sqrt{Q_\phi}\, \xi_t. \tag{3.10}$$

For the averaging method to apply one has to assume that the changes in relative phase and in amplitude ($\dot{\phi}$ and $\dot{r}$) are slow compared to the frequency. This is a reasonable assumption for weakly perturbed oscillators. The noise in the relative phase dynamics is modeled in an ad-hoc fashion by additive gaussian white noise of strength Q_ϕ. The stationary relative phase of eq. 3.7 is a stable fixed point of the relative phase dynamics.

From the relative phase dynamics two measures of temporal stability can be derived. The first is obtained from linear stability analysis around the fixed point (see next section how this can be done). From this we obtain the *relaxation time*:

$$\tau_{\mathrm{rel}} = \frac{2}{\tilde{\alpha}}. \tag{3.11}$$

The relaxation time measures how fast the system returns to the stable state after a perturbation. The relaxation time contains two contributions, one from

---

[1] This is a good approximation for the experiment in chapter 2.

the intrinsic dynamics α and one from the visual surround c_{env}/D_0. Note that a decrease of mean distance D_0 decreases the relaxation time, whereas the frequency, ω_D, and amplitude, D_r, of the drive have no influence. A second measure of temporal stability is obtained by considering the fluctuations in relative phase around the fixed point. For the standard deviation of these fluctuations one obtains:

$$SD_\phi = \sqrt{Q_\phi/\tilde{\alpha}}. \tag{3.12}$$

Because of the circular nature of relative phase, I will quantify these fluctuations by the *angular deviation*, which is the circular equivalent of the standard deviation. When the fluctuations are small the two are equal. The dependence of the standard deviation of relative phase on the parameters of the dynamics is similar as for the relaxation time. This means that these two measures of temporal stability should covary when stability is manipulated (e.g. via the visual distance; see chapter 2).

If the driving frequency, ω_D, is significantly detuned from the eigenfrequency, ω_0, the fixed point could become unstable. This instability is, however, never reached in the linear model because the amplitude, r_0, vanishes simultaneously. It is easy to show that:

$$\frac{(\omega_0^2 - \omega_D^2)/(2\omega_D)}{(c_{env}D_r)/(2r_0D_0)} < 1.$$

Therefore, solutions with finite amplitude but unstable relative phase do not exist in this linear model (but they do exist in the sine-circle map introduced below). In light of the outcome of the experiment I note that a very similar equation for the relative phase as eq. 3.10 can be derived by averaging a driven nonlinear oscillator model [27, Sect. 4.2]. In this case the instability of the fixed point for relative phase can be reached by detuning the driving frequency from the eigenfrequency, because the amplitude of the oscillator does not necessarily go to zero when the stability limit for relative phase is reached.

3.2.2 The stochastic sine-circle map

For quantitative comparison to the experimental data I use a generalisation of the dynamics of relative phase eq. 3.10 which covers both the case in which such dynamics arise from a passively driven system as well as the case in which such dynamics arise from a driven active (nonlinear) oscillator. Because relative phase is estimated in this paper from experimental data at discrete points in time (e.g. at peaks or valleys of position), I use a discrete time dynamical model or map. The simplest dynamical system compatible with the circular nature of relative phase is the sine-circle map [6]:

$$\phi_{n+1} = \phi_n + a + b\sin(\phi_n - \phi_0) + \sqrt{Q}\xi_n, \qquad (3.13)$$

where $\phi_n \in S^1$ is the relative phase at time n and a, b and ϕ_0 are parameters to be explained below. For comparison to noisy data I have included a stochastic component through independent gaussian random forces ξ_n (zero mean, unit variance) acting with strength Q.

Maps of this type arise through Poincaré sections in models of driven nonlinear oscillators [6]. In these cases the parameter a arises as the difference in frequency between the oscillator and the forcing function and b reflects the coupling strength of the forcing function and the oscillator. I employ the model in a phenomenological spirit by mapping the observed relative phase series onto solutions of this dynamical system. In this case the parameters arise from numerical fits to experimental data (see section 3.3). The meaning of the parameters is best interpreted in terms of how the fixed points and their stability depend on the parameters. To discuss this I first consider the deterministic dynamics by setting $Q = 0$. The parameter ϕ_0 can be absorbed into a shift of coordinates and therefore does not affect the stability of the dynamics. In section 3.4 I show that ϕ_0 can be approximated as zero. Therefore I will set $\phi_0 = 0$ for the purposes of this discussion.

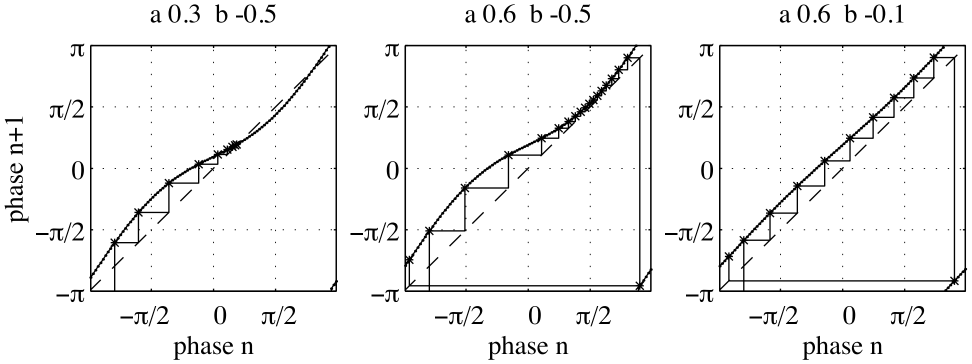


Figure 3–1 Examples of return maps of the sine-circle map for several parameter values. Note that phase is a periodic variable i.e. phases moving off at the top reappear at the bottom and phases moving off at the left reappear at the right.

The sine-circle map can be visualised by the return map, a plot of ϕ_{n+1} versus ϕ_n. In figure 3–1 I show examples for different values of the parameters a and

b. Note that when $|a| < |b|$ the system has two fixed points, only one of which is stable (which one depends on the sign of b). The stable fixed point is the one for which the derivative of the right-hand side of eq. 3.13 is smaller than 1 in absolute value. More precisely, for $|a| < |b|$ the two fixed points are:

$$\phi_f^1 = -\arcsin(a/b), \tag{3.14}$$

$$\phi_f^2 = \pi + \arcsin(a/b). \tag{3.15}$$

Stability of the fixed points is calculated by linearising the map in the neighbourhood of the fixed points. Take $\phi_n = \phi_f^1 + \delta_n$ and $\phi_{n+1} = \phi_f^1 + \delta_{n+1}$ with δ_n and δ_{n+1} small perturbations. By substitution in (3.13) and expanding the sine function up to second order, we obtain:

$$\delta_{n+1} = (1 + b\sqrt{1-(a/b)^2})\, \delta_n.$$

For $b < 0$ we find for an initial perturbation δ_0:

$$\delta_n = \delta_0(1 - \sqrt{b^2 - a^2})^n, \tag{3.16}$$

which converges to zero if $|1-\lambda| < 1$ in which case ϕ_f^1 is stable. Here I have defined the *local stability parameter* by $\lambda \equiv \sqrt{b^2 - a^2}$. Note that λ is only defined when $|a| < |b|$, i.e. when there exists a fixed point. The other fixed point is unstable with $\delta_n = \delta_0(1+\lambda)^n$ diverging. For $b > 0$ the roles of the fixed points ϕ_f^1 and ϕ_f^2 are reversed. Intuitively, large b or in terms of underlying oscillators, strong coupling, implies large λ inducing stable phase locking. Stability is similarly enhanced by small a, i.e. good match of driving frequency and eigenfrequency. Equation 3.16 describes the decay of perturbations of relative phase in units of the discrete time. Later I will compare relaxation processes across frequencies. Because the discrete time units are determined by the cycle time, this implies that the size of the relevant time units for measuring the relaxation process varies with frequency. To make comparisons across frequency I rescale the local stability as $\tilde{\lambda} = \lambda/\delta t$ where δt is the relevant time unit (quarter of a cycle). This rescaling converts the relaxation process to units of absolute time. For small δt, eq. 3.16 may be approximated as:

$$\delta_n = \delta_0 \exp(-T/\tau_{\mathrm{rel}}), \tag{3.17}$$

with $\tau_{\mathrm{rel}} = 1/\tilde{\lambda}$ the relaxation time and $T = n\,\delta t$ the time at which $n/4$ cycles have run past. Theoretically, the relaxation time of eq. 3.11 and of eq. 3.17 could be different. For simplicity, I will identify the two definitions.

At $|a| = |b|$ the model exhibits a tangent bifurcation: the two fixed points collide at phase $\pm\pi$. When $|a| > |b|$ no fixed points exist. Note that when $|a|$ is slightly larger than $|b|$ (see figure 3–1 middle panel) a narrow tunnel exists where the return map comes close to the diagonal. In this tunnel the relative phase takes small steps, whereas it takes large steps outside the tunnel. This means that the oscillator is not phase locked to the forcing function but is still influenced by it so that the oscillator lingers for longer periods of time near a relative phase of $\pm\pi/2$. When $|a|$ is much larger than $|b|$ (see figure 3–1 right panel) the phase takes constant steps without any significant slowing down. This means that the oscillator is moving independently of the forcing function. We may define a *global stability parameter*, γ, which indicates how close the system is to one of the bifurcations: $\gamma \equiv a/|b|$. For $|\gamma| < 1$ one stable and one unstable fixed point exists, for $|\gamma| > 1$ no fixed point exists. This parameter can be used to characterise the dynamics overall also where the local stability parameter is not defined.

The model of eq. 3.13 has a very rich dynamics, not all of which are relevant to our system. The model allows for non 1-1 locking which I never observed. Further when $\lambda > 2$ the stable fixed point becomes unstable again. Finally when $|b| > 1$ the map becomes non-invertible and chaos may occur [33]. I have neither observed these two phenomena nor did I obtain reconstructions of return maps corresponding to these parameter regimes, so I will not discuss these cases here.

The discrete time dynamics of eq. 3.13 can be related to continuous time dynamics of relative phase as discussed in the previous section. Clearly, in view of the richness of the discrete dynamics much exceeding that of a first order differential equation, this relationship is not exact. However, in the limit of the time unit, δt, of the discrete dynamics becoming infinitesimally small, the difference quotient $(\phi_{n+1} - \phi_n)/\delta t$ approximates $\dot{\phi}$. In this limit, eq. 3.13 becomes:

$$\dot{\phi} = \tilde{a} + \tilde{b}\sin(\phi(t) - \phi_0) + \sqrt{\tilde{Q}}\xi_t,$$

where the parameters have been rescaled according to $\tilde{a} = a/\delta t$, $\tilde{b} = b/\delta t$, $\tilde{Q} = Q/\delta t$ and ξ_t is gaussian white noise of unit variance. In my experiments, relative phase is obtained at 4 discrete points in each cycle, so δt equals a quarter cycle time. Even though this is not infinitesimal, I shall employ these rescaled parameter values to compare relative phase dynamics across frequencies. Comparing this equation to eq. 3.10, we can make the identification:

$$\tilde{a} = \frac{\omega_0^2 - \omega_D^2}{2\omega_D}, \tag{3.18}$$

$$\tilde{b} \;=\; \frac{c_{env} D_r}{2 r_0 D_0}. \tag{3.19}$$

Adding noise to the sine-circle map does not change the dynamics very much (unpublished observations; for low noise levels see [87]). When there is a stable fixed point the noise causes the phase values to scatter around the fixed point. The width of the distribution of phase values, when the noise is not too large, can be characterised by the standard deviation of relative phase and is given by [66]:

$$SD_\phi = \sqrt{Q/\lambda}. \tag{3.20}$$

Further, the bifurcation is softened by the noise: for $|b|$ slightly larger than $|a|$ the fixed point has low stability and due to the noise the system can escape easily from the attractor and wrap around. The stationary distribution cannot be calculated analytically. It can be solved numerically though by discretisation of the transition matrix and by calculation of the left and right eigenvectors with eigenvalue 1 [59].

3.2.3 Coordination patterns from the sine-circle map

Depending on the global stability parameter γ, I can model three types of coordination: absolute coordination, relative coordination and no coordination. This distinction was first introduced by von Holst for the observed patterns of coordination between fins of fish [82]. The situation of two oscillators that are phase locked was termed *absolute coordination* by von Holst. The situation when two oscillators have no fixed phase relationship, i.e both oscillate with different frequencies, was termed *uncoordinated*. Finally, the situation when two oscillators have no fixed phase relationship but have a preferred phase relationship was termed *relative coordination*. To get an intuitive feeling for relative coordination von Holst gives a nice example: “When a father goes out for a walk with his six- to eight-year-old son, one can often observe the following: The boy would like to keep pace with his father, but this does not work for long. After a number of coincident steps, he gradually looses the tempo, and in order to fall into step once more he makes one or two rapid additional steps. By doing so, he again falls in with his father’s pace and the game is all ready to start again”

I can illustrate these concepts in figures 3–4 to 3–6, where I have drawn figures based on actual data, all for a high driving frequency. These data will be discussed more extensively in section 3.4. The upper panels show the trajectories of the visual stimulus and the postural response and the middle panels show the traces of relative phase. The lower left panels show the return map as a smooth curve in a somewhat different format (this is handy for fitting the parameters[2].):

$$\phi_{n+1} - \phi_n = a + b \sin(\phi_n).$$

The lower right panels display the histogram of relative phase. In figure 3–4 we observe absolute coordination: we have $|a| < |b|$, the return map has a fixed point and the time series of relative phase scatters around this fixed point (cf. fig. 3–1 right panel). Because this trial is for a high driving frequency the subject is not very tightly locked and misses cycles. For a case where the subject is tightly locked see figure 3–3. In figure 3–5 middle panel we observe relative coordination in the middle part between the two perturbations. We see periods with slow dynamics interspersed with shorter periods of fast wrapping. From a fit of the return map using only the phase values of this middle section I obtain $b = -0.54$ slightly larger in absolute value than $a = 0.53$ showing that the return map has a very weak fixed point and that the system is close to the bifurcation (cf. fig. 3–1 middle panel). Finally, in figure 3–6 we observe no coordination: we have $|a| > |b|$ and the return map has no fixed point (cf. fig. 3–1 right panel). Further, there is continuous wrapping in the time series of relative phase.

3.3 Methods

A detailed description of the experimental set-up is given in section 2.2.1.

3.3.1 Stimuli

The stimuli simulated a fronto-parallel wall covered with 140 stereo dots, each with a size of 0.2 deg by 0.2 deg. The density of dots was uniform per solid angle as seen from the position when the experiment started. The dots lay in an annulus between 10 deg and 45 deg visual eccentricity. The hole in the middle of the stimulus was made to suppress the visibility of aliasing effects, which were most visible in the foveal region. The spatial resolution of the system was 1152 by 900 pixels and the frame rate was 66 Hz.

The wall was suggested at a mean distance of 50 cm, which was equal to the physical distance between observer and screen. Distance was suggested both by stereo vision as well as by simulation of the appropriate displacement of the image on the screen, using feedback of the eye position of the observer. The wall was sinusoidally driven with frequencies of 0.05 Hz, 0.1 Hz, 0.2 Hz, 0.3 Hz, 0.4 Hz and 0.5 Hz. I covaried the amplitude with frequency so as to keep the

---

[2] Fixed points are now determined by $a + b \sin(\phi_n) = 0$

mean absolute velocity of the wall constant. In this way I kept the perceptual strength roughly constant: all stimuli were near the threshold where stimulus motion could be consciously perceived. I took amplitudes of 10 cm, 5 cm, 2.5 cm, 1.67 cm, 1.25 cm and 1 cm respectively. All conditions except for 0.05 Hz had two perturbations of 180 deg at the point of maximum velocity (so there was no discontinuity in position, cf. fig. 3–3, upper panel). The perturbations occurred at random moments in time, but were separated in time by at least 35 seconds, thus ensuring that there were enough unperturbed cycles to estimate the relaxation time. Because of the small number of cycles in the 0.05 Hz condition I had only one perturbation for this condition.

Each condition was repeated 6 times and the order of trials was balanced in a latin square design. With this design every condition followed every other condition precisely once. The experiment was conducted in two sessions of approximately 1 hour each and always started with two trials at 0.2 Hz for stimulus adaptation of the subject. In all there were 40 trials for each subject. Each trial lasted for 140 s and the first 20 s were used for adaptation and were not stored.

3.3.2 Data analysis

Data analysis was done in two ways: a linear time invariant analysis and a dynamic approach based on the time series of relative phase. Both are described in section 2.2.3. There were only minor differences. First, I smoothed the data differently, viz I used a gaussian window with a standard deviation of 0.18 s (for the 0.05 Hz condition this was 0.64 s and for 0.1 Hz this was 0.36 s). Second, I took the percentages to determine whether an extremum is significant or not lower, because the amplitude was much more variable within a trial than in the previous experiment. I took 20% for the position extrema and 50% for the velocity extrema.

Measures calculated from the time series of relative phase

For all trials I determined mean phase and angular deviation from the combined time series of relative phase, using circular statistics [3]. These were calculated as in section 2.2.3. It should be noted that angular deviation is biased: it tends to be underestimated for small samples [3, p 46]. In the experiment the number of phase values is dependent on the mean frequency of the response and can vary by as much as a factor 10. There are complicated correction procedures for this bias (which assume specific distributions), but I will present the angular deviation as calculated above and not make strong conclusions from it alone. The advantage of the angular deviation over most of the stability measures presented below is that it can be calculated for all trials, no matter how unstable. I excluded the phase values in a period of 8 s after the perturbation, because mean phase and angular deviation are measures for stationary behaviour.

As a further measure of the stability I calculated the winding number defined as $W = (\phi_N - \phi_0)/(2\pi N)$, with N the number of phase values. The winding number measures the mean wrapping of the response relative to the drive. Negative winding number means that the response is slower than the drive (this occurs for high driving frequencies) and positive means that the drive is slower than the response (this occurs for low driving frequencies). A winding number near zero means that the response is phase locked to the drive.

Relaxation time

The relaxation time, the time it takes the system to regain its phase locked behaviour after a perturbation, was estimated as follows: I defined a band around the local mean of relative phase of 1.5 times the local angular deviation. Local mean phase and angular deviation were calculated from a segment starting 8 s after the perturbation and with a length of four cycles. After a perturbation the phase generally leaves this band. I calculated reentry time as the time between the perturbation and the time of reentry into the band. From this I calculated the relaxation time by fitting an exponential to the points outside the band and the first six points inside the band (cf. fig. 3–3, middle panel). The fit was implemented by a nonlinear Levenberg-Marquardt routine [57]. The points inside the band were given a larger weight because they were less prone to noise. I did not use linear regression on the logarithm of the phase values as in chapter 2 because the phase values are not necessarily all at the same side of the local mean phase. If the relative phase did not leave the band or if there occurred many wrappings near the perturbation (cf. fig. 3–6, middle panel), I excluded the perturbation. The relaxation time depends somewhat on the parameters used but the trend with frequency does not depend on them. It should be noted that this procedure probably underestimates the larger relaxation times because these occur for lower stability, which means that the band is wider and hence that the system is bound to return faster inside the band. I calculated the relaxation time of a trial by averaging the relaxation times of the two perturbations within a trial.

Fitting of the return map and histogram

I estimated directly the parameters a, b and Q of the stochastic sine-circle map. I did this both by fitting the return map and by fitting the histogram of relative phase. In theory these two fits should contain the same information, but in practice they do not necessarily contain the same information. First, the consistency between the two fits gives information on how accurate the stochastic sine-circle map models the posture system. For an example where the consistency is high, see Fig. 3–4 lower left panel and for an example where the consistency is low, see Fig. 3–6 lower left panel. Second, depending on the stability and the noise strength, the accuracy of the parameters estimated with the fits can be very different. For high stability and low noise the variability is low. This means that the histogram has a sharp peak, which is easily fitted, but that the data points in the return map are concentrated around the fixed point, making the fit statistically less reliable (see Fig. 3–3 lower panels). When the variability is high, due to high noise or low stability, the histogram is very flat and hard to fit. However, the data points in the return map are spread out over the whole interval $(-\pi, \pi)$ which makes the fit statistically more reliable (see Fig. 3–5 lower panels).

I fitted the return map by linear least squares of $y = a + bx$ with $y = \phi_{n+1} - \phi_n$ and $x = \sin \phi_n$ where ϕ_n is the n-th relative phase value. The unexplained variance of the fit gives an estimate of the noise strength Q. I included the phase values directly after a perturbation in the fit because these provide information about the transient relaxation towards the attractor. Further, I excluded 2.5% of the phase values from the return map fits. This was done because some of the phase values were outliers. Some of these outliers in the return map were caused by small (but fast) bumps in the response data, leading to an extra pair of extrema in the velocity and thus to a wrap in the time series of relative phase. Especially for the phase locked trials, where all phases scatter around the attractor, these outliers changed the fitted parameters considerably. Some other outliers were caused by transients such as shown, for example, in figure 3–4, middle panel, where the subject lingers around phase -2.5 radians after the first perturbation. Excluding the outliers from the return map fit also led to better agreement of the parameters from the fit of the return map and of the histogram. As a measure of significance of the fit I used its F-value. I excluded the results of a fit when its significance exceeded the level of 10%.

The histogram was constructed from the time series by grouping the phase values in bins. I calculated the stationary distribution of the stochastic sine-circle map [59] and used a nonlinear least squares fit procedure (Levenberg-Marquardt) to fit the parameters of the stationary distribution to the experimentally observed histogram. The starting values of these fits were the parameters from the return map fit. When I found an attractor ($|a| < |b|$) with the return map fit, I fitted a and Q, otherwise I fitted a and b. Fitting all three parameters led to unstable results, because the histogram has only two conspicuous features: a mean and a width. The number of bins was adapted to the number of phase values and the angular deviation of the phase values. The results of the fit depend weakly on the precise number of bins. As a measure of significance of the fit I took the χ^2-value which I used to test whether the histogram differs significantly from a uniform distribution. I excluded the results of a fit when its significance was more than 1%.

For comparison of the parameters a, b and Q with their counterparts in the continuous model eq. 3.10 I rescaled the parameters with the sampling frequency of the time series of relative phase (four times the mean frequency of response).

3.3.3 Subjects

Six subjects with normal or corrected to normal vision were tested in all conditions. Three of the subjects were familiar with the purpose of the experiment. It should be noted that some of the results (especially relative coordination) were not expected at the time of the experiment. Four subjects had participated in the previous experiment. Subjects were instructed to look at the centre of the stimulus and to stand relaxed. Subjects stood on a firm stable support in normal Romberg posture.

3.4 Results

Generally the subjects responded to the sinusoidally moving wall with a sinusoidal postural response in fore/after direction. They displayed a rich set of dynamical behaviour: tightly phase locked (absolute coordination) in the middle frequency range[3], both absolute coordination and no coordination at the low frequencies and all three coordination patterns at the high frequencies. I used a rough classification based on the global stability parameter γ to get an idea of the frequency of occurrence of the various coordination patterns. When $|\gamma| < 0.75$, I classified a trial as absolutely coordinated, when $|\gamma|$ was between 0.75 and 1.25, I classified a trial as relatively coordinated and when $|\gamma| > 1.25$, I classified it as uncoordinated. Proceeding in this way, I found approximately 81% of the trials to be absolutely coordinated, 11% to be relatively coordinated and 8% to

---

[3] This was the only coordination pattern I found in the experiment of the previous chapter.

be uncoordinated. To give the reader an idea of the rich dynamical behaviour, I display and discuss a few data sets.

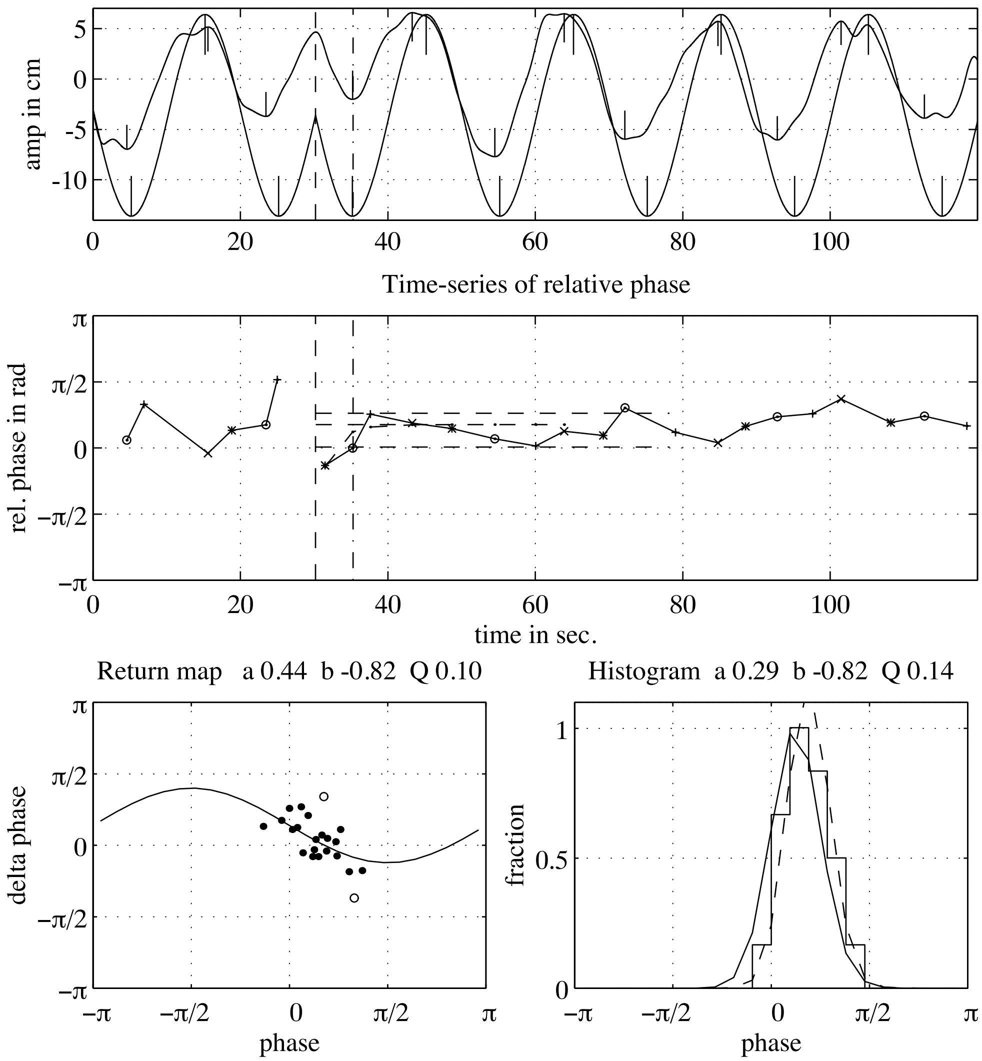
3.4.1 Examples of dynamic behaviour

In figure 3–2 I show an example of absolute coordination at a low driving frequency. In the upper panel I show the trajectories of the stimulus and the response of the subject. The subject is advanced relative to the stimulus and is phase locked. In the middle panel I show the corresponding time series of relative phase. The relative phase is generally positive, indicating that the subject is advanced, and scatters somewhat around the mean value. The return map is shown in the lower left panel and clearly shows the attractor to be stable as indicated by the negative slope at the fixed point. This is again corroborated in the lower right figure, where I show the histogram of relative phase indicating a clear peak centered on the mean phase. Note that the predicted distribution of relative phase as derived from the parameters of the fit of the return map (dashed line) is close to the optimal fit of the histogram (solid line), just as in the following figures.

In figure 3–3 I show an example of absolute coordination at a mid-range frequency. From the return map and the histogram one can see that the system has a clear attractor. There is a tight phase locking with a relatively small amount of noise. Note that the amplitude is relatively constant which should be compared to the large variability in amplitude in figures 3–4 to 3–6. Further, there is a very slow drift of equilibrium position on a time scale of 30 s. This slow drift is probably related to comfortable standing.

In figure 3–4 I show an example of absolute coordination at a high driving frequency. The subject is phase locked to the visual motion most of the time, but misses a few cycles, mostly at points where he increases his amplitude (around time 40 and 95). From the return map and the histogram one can see that the system still has an attractor but that it is not very strong. As I discussed in the theory section, the noise tends to blur the bifurcation and thus the distinction between the coordination patterns. There are two reasons why I would classify this coordination pattern as absolute coordination. First, amplitude clearly goes up as coordination is lost, contrary to the next figure. Second, the return map has a fixed point with $|\gamma| < 0.75$ even when I leave the data of the first and the last part of the time series of relative phase (which are clearly phase locked) out of the fit.

In figure 3–5 I show an example of relative coordination at a high driving frequency. In the middle panel in the stretch between the two perturbations


Time-series of relative phase

Return map a 0.44 b -0.82 Q 0.10

Histogram a 0.29 b -0.82 Q 0.14

Figure 3–2 Example of a data record at low driving frequency (0.05 Hz). Upper panel: position traces of visual drive (shifted down for visibility) and postural response. Middle panel: derived time series of relative phase (solid line) with exponential fit of the relaxation (dashed line). See fig. 2–2 for details. Lower left panel: derived difference return map with scatterplot of the data (open dots are excluded from the fit) and the curve of the best fit. The parameters of the fit are indicated. Lower left panel: derived histogram of relative phase (staircase plot) with the best fit (solid line) and the stationary distribution as calculated with the parameters of difference return map (dashed line).

one clearly sees slow dynamics at a phase of roughly $-\pi/2$ and fast dynamics at

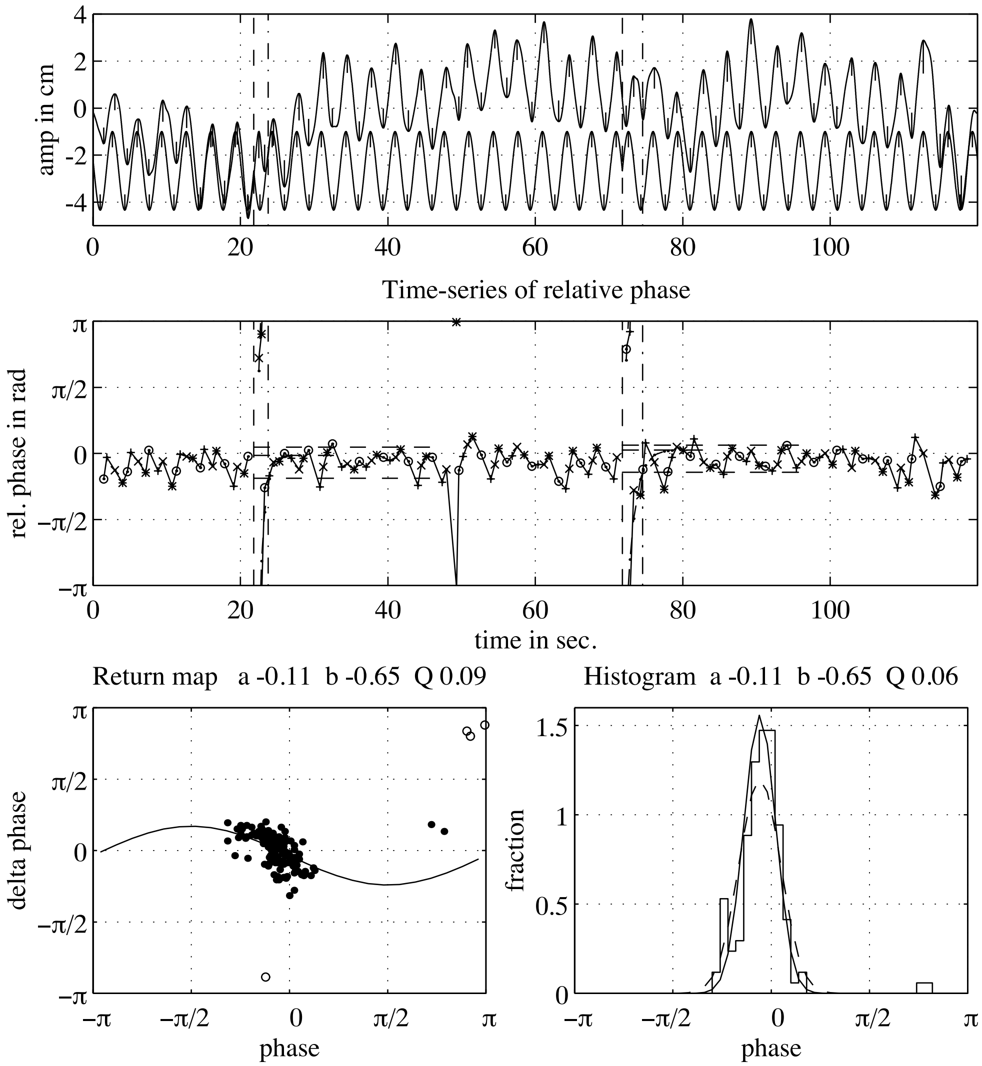


Figure 3–3 Example of a data record at medium driving frequency (0.2 Hz). For details see legend of fig. 3–2.

the other phases (cf. fig 3–1 middle panel). The trajectories clearly show that postural sway is still oscillatory in spite of the loss of strict phase locking. This is an indication that sway motion is actively generated and not passively driven by visual motion. Because the stretches before the first perturbation and after the second are phase locked one does not see the relative coordination clearly

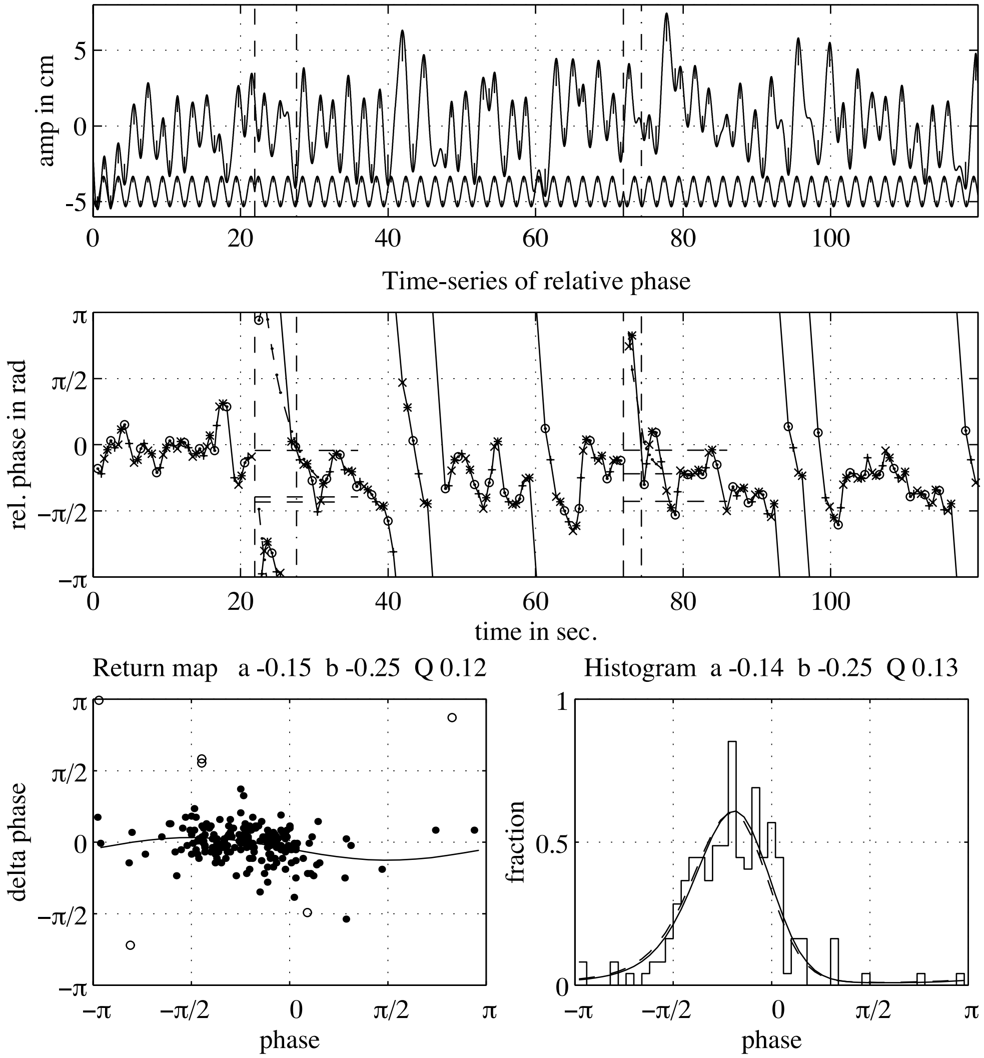


Figure 3–4 Example of a data record at high driving frequency (0.5 Hz). This trial shows absolute coordination. For details see legend of fig. 3–2.

reflected in the return map and the histogram. However, fitting the return map and the histogram only for the phase values between the two perturbations I found $a = 0.53$ and $b = 0.54$. It is interesting to note that the perturbations seem to cause a switch from one type of dynamics to another. Also interesting is the finding that the two relaxation times are very different: the first is 4.0 s and the second 1.4 s. This is consistent with the finding that the stretch after the first

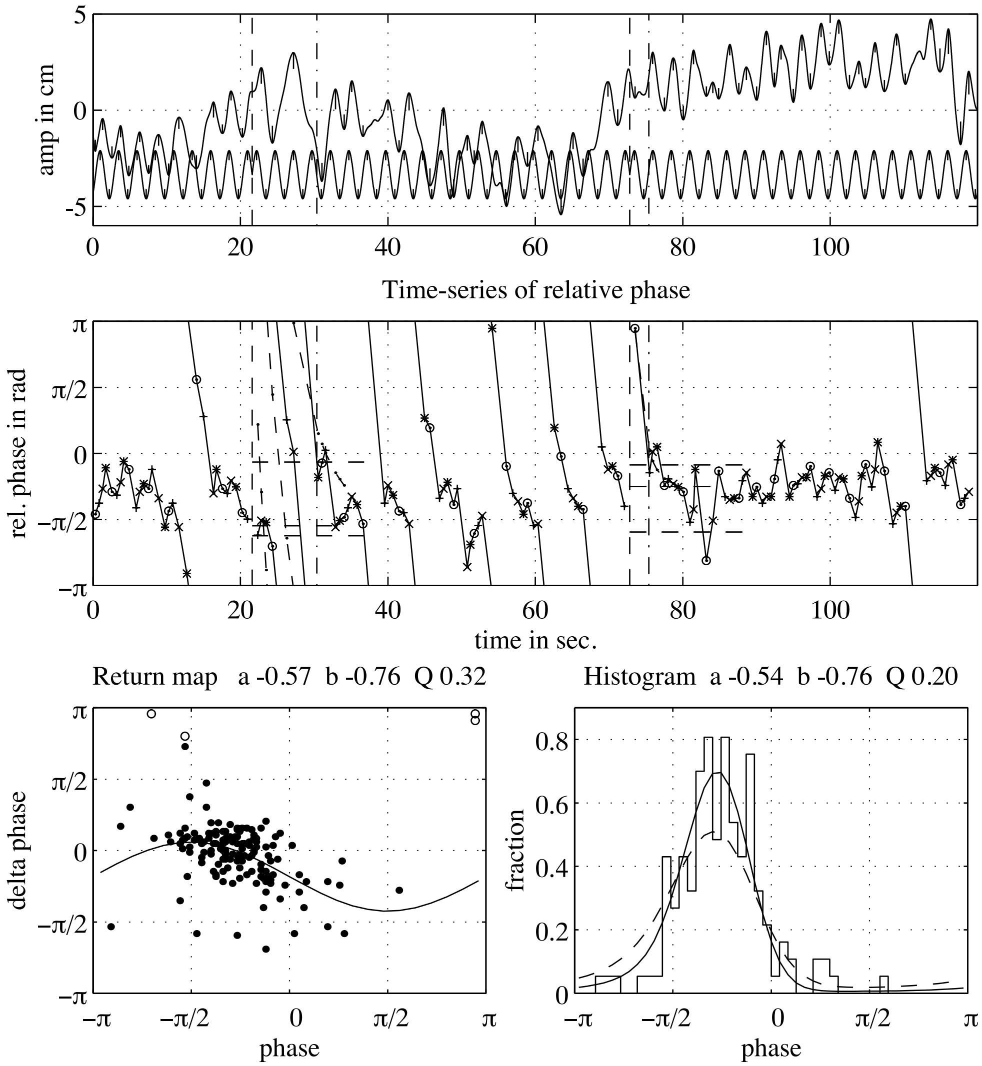


Figure 3–5 Example of a data record at high driving frequency (0.4 Hz). This trial shows relative coordination. For details see legend of fig. 3–2.

perturbation has a low local stability, whereas the stretch after the second has a high local stability.

In figure 3–6 I show an example of absence of coordination at a high driving frequency. In the middle panel one can see that the subject continuously wraps in phase indicating that he oscillates at his own frequency, different from the

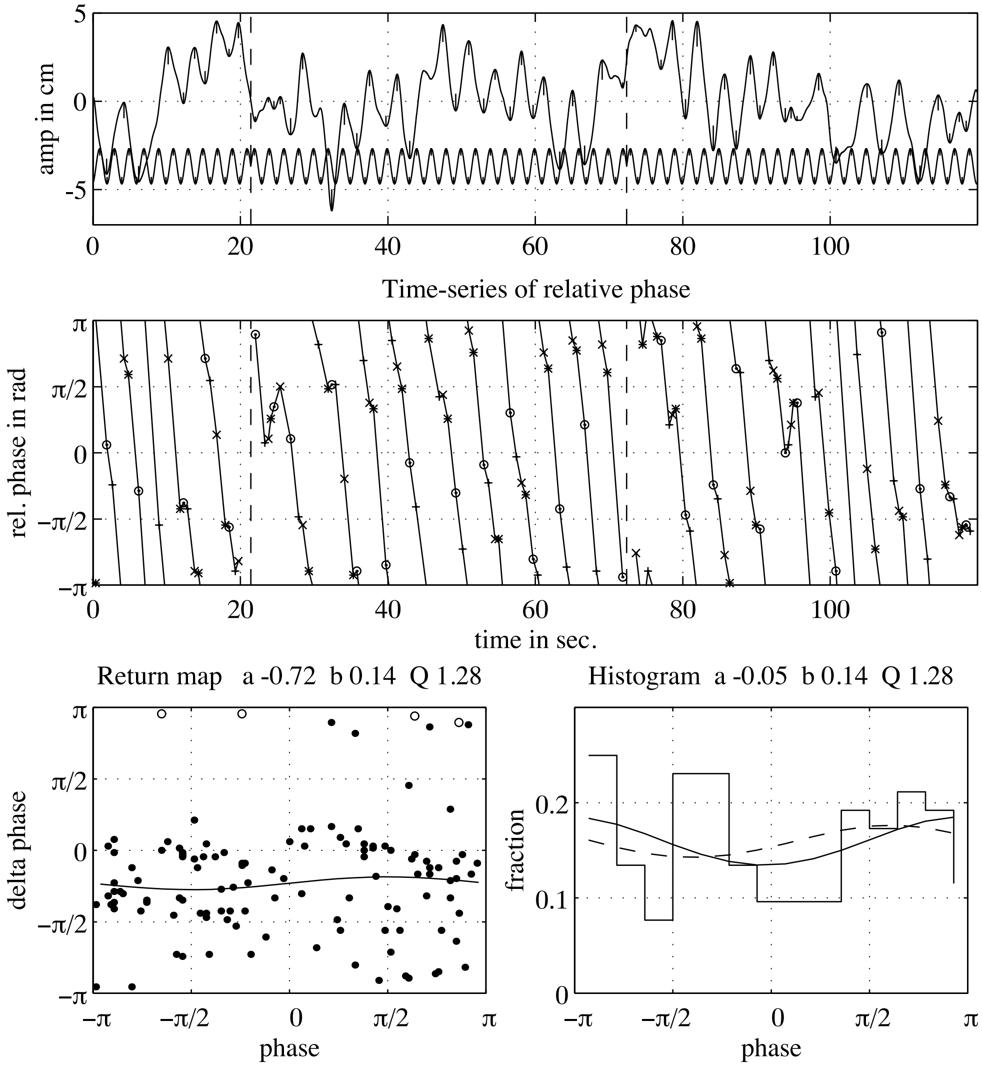


Figure 3–6 Example of a data record at high driving frequency (0.5 Hz). This trial shows no coordination. For details see legend of fig. 3–2.

driving frequency. Like relative coordination this form of uncoordinated sway is indicative of active generation of sway motion. The relative phase is very variable as can be seen from the return map, where the data points fill the lower part (with delta phase negative). The noise strength Q is very large. Note that the histogram is not a very useful tool in this case because it does not differ significantly from a uniform distribution.

3.4.2 Mean relative phase, stability and mean response frequency

In this section I present the main findings. I tested all effects with a one-way ANOVA with independent factor frequency for each subject separately at a significance level of 5%. In general most effects of frequency presented here were significant.

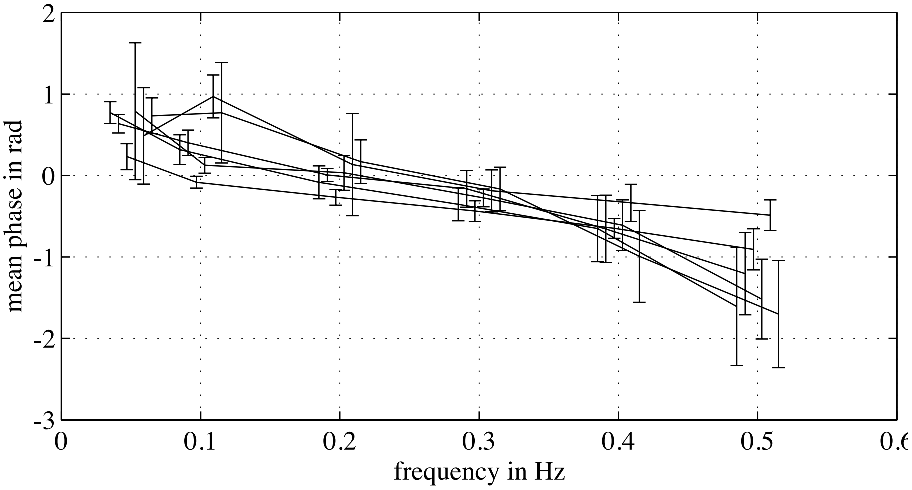


Figure 3–7 Mean phase as a function of driving frequency for all subjects. The curves for different subjects are horizontally displaced for better viewing. The errorbars denote standard deviation.

In figure 3–7 I have plotted the mean phase difference between visual drive and postural response as calculated from the time series of relative phase. All subjects had a significant frequency effect for mean phase. From the figure we observe that subjects are advanced relative to the stimulus at frequencies below 0.2 Hz, are almost in-phase near 0.2 Hz and are delayed at high frequencies. This frequency dependence is predicted from eq. 3.7, in terms of which the mean eigenfrequency of the postural system is in all subjects approximately 0.2 Hz. Also note that there is a tendency that the errorbars are larger for the low and high frequencies. This enhanced variability across trials reflects the reduced temporal stability at these frequencies (see below). The mean phase as calculated from the time series of relative phase correlates significantly with the mean delay as calculated from the spectrum: the correlations for each subject are in the range 0.62 to 0.82, mean 0.76. The mean phase at a driving frequency of 0.2 Hz for the four subjects who also participated in the previous experiment matches very closely the results in that study for equivalent conditions (distance of 50 cm): Subject CG has a mean phase near 0.27 rad in both experiments, the other subjects have a mean phase close to zero in both experiments.

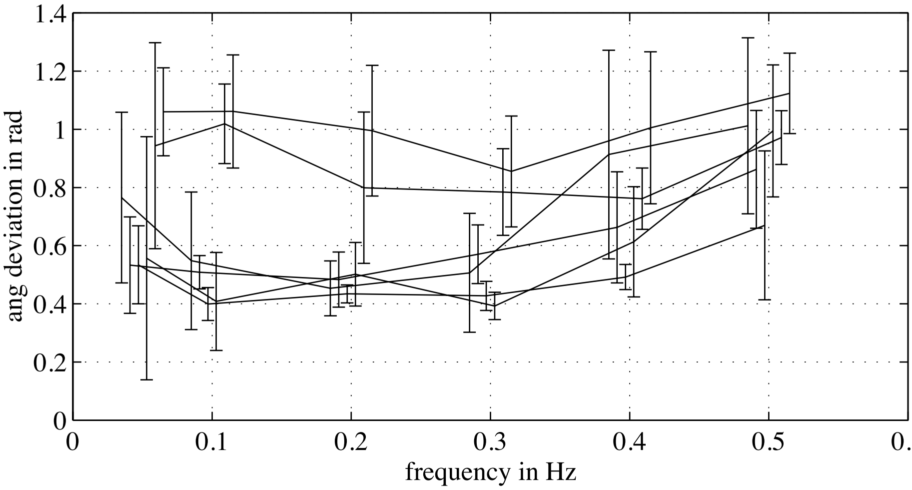


Figure 3–8 Angular deviation of relative phase as a function of driving frequency for all subjects. The curves for different subjects are horizontally displaced for better viewing. The errorbars denote standard deviation.

In figure 3–8 I have plotted the angular deviation. This measure for the temporal stability has a significant frequency effect for four of the six subjects. The figure shows the stability to be higher (variability lower) in the range 0.2–0.3 Hz with a decrease in stability towards the low and high frequencies. The two subjects lacking significance can be seen to have higher variability throughout, which also increases towards the higher and lower frequencies. The linear dynamical model does not predict any effect of frequency on stability (cf. eq. 3.12) so that the observed effect is evidence for nonlinearities. Because of the bias of the angular deviation (see Methods) I cannot directly compare the stability of the low and high frequencies. The angular deviation correlates significantly with the Magnitude Squared Coherence (MSC) calculated from the spectrum: the correlations for each subject are in the range -0.66 to -0.87, mean -0.78. The correlations are negative because small angular deviation and high MSC both signify high stability.

In figure 3–9 I have plotted the relaxation time. There was no significant effect of frequency on relaxation time for all but one of the subjects. Excluding the relaxations at the low frequency of 0.05 Hz, where the sampling of the phase

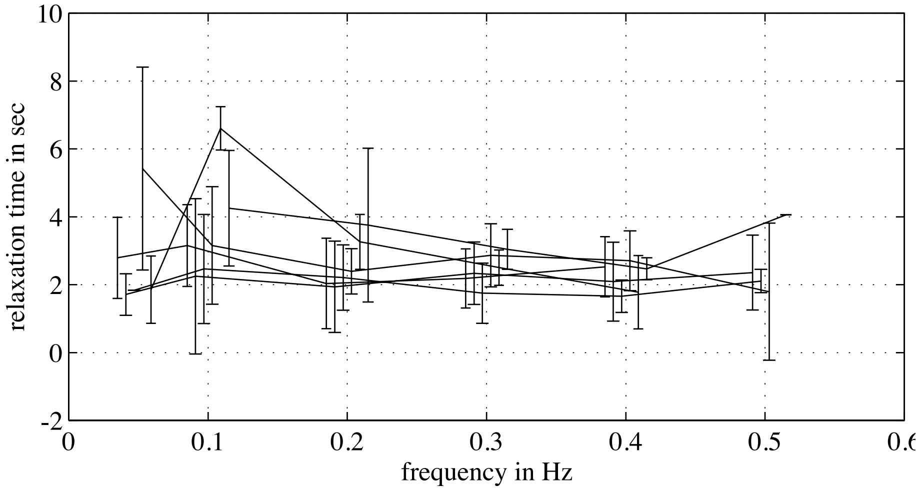


Figure 3–9 Relaxation time as a function of driving frequency for all subjects. The curves for different subjects are horizontally displaced for better viewing. The errorbars denote standard deviation.

is problematic, the relaxation time is at a constant level between 2 and 3 s, except for some outliers. For the low and high frequencies many perturbations had to be excluded because the subject showed no clear locking after the perturbation. Subjects did not differ in relaxation time but did differ in the number of included relaxations (ranging from 97% to 45%, mean 72%), especially at the low and high frequencies. These exclusions may lead to an underestimation of relaxation time at the low and high frequencies and therefore obscure an effect of frequency on relaxation time.

In the previous experiment I reported a significant correlation between angular variance and relaxation time. Here I found correlations ranging from 0.05 to 0.62, mean 0.37 (see fig. 3–10 for a graphical impression of this covariance). Half of these were significant at a level of 5% and all of them were lower than in the previous experiment for the subjects that also took part in that experiment. Of course, we cannot expect a strong correlation because the frequency manipulation does not induce a significant effect in the relaxation time and because the angular deviation is biased.

In figure 3–11 I have plotted the difference between mean frequency of the response and the driving frequency. The mean response frequency was calculated from the time series of relative phase. All subjects display a significant effect in this variable. For frequencies in the range 0.1–0.2 Hz we note that the mean frequency of the response equals the frequency of the input, except for two subjects.

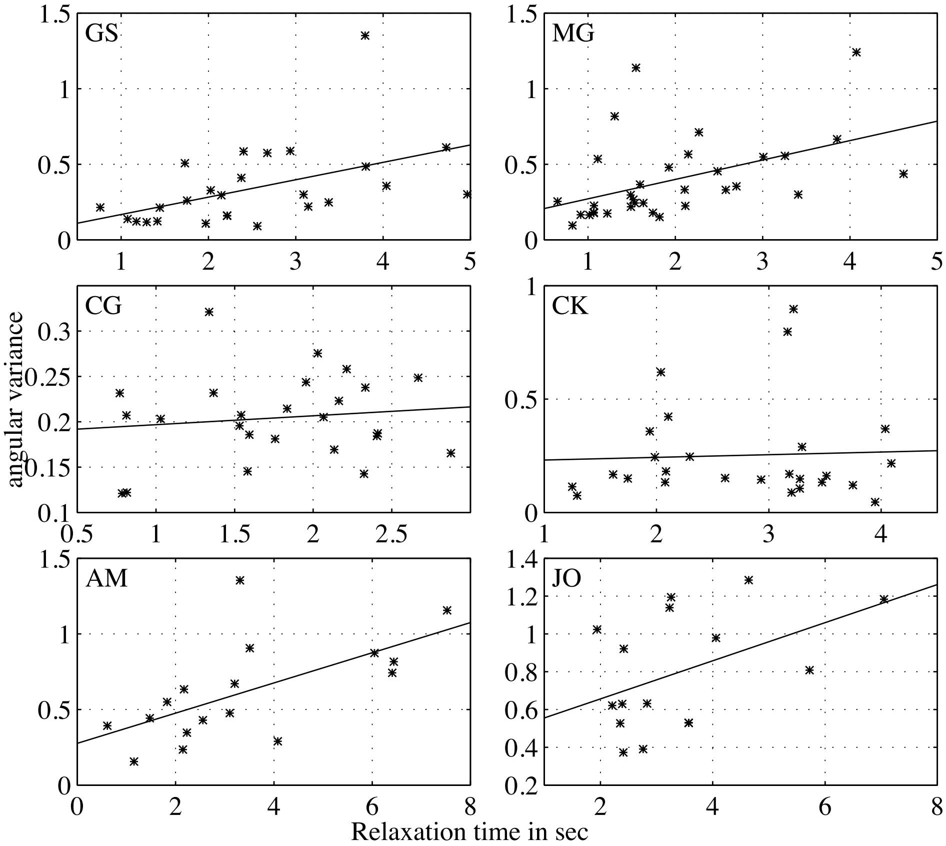


Figure 3–10 Scatterplot of the correlation between the relaxation time and the angular variance. The six panels are for six different subjects. The solid lines denote the best fit of a linear regression.

These subjects also showed a higher angular deviation at these frequencies. For other frequencies, especially at 0.5 Hz, there are marked differences. Subjects have a tendency to respond with a higher frequency for the low driving frequency of 0.05 Hz and to respond with a lower frequency for the high driving frequencies. This type of behaviour cannot be exhibited by a purely linear system because for a linear system the frequency of the response always equals the frequency of the input. Also note that the variability across trials as denoted by the size of the errorbars is higher at the extreme frequencies, especially 0.5 Hz, than at the middle frequencies.

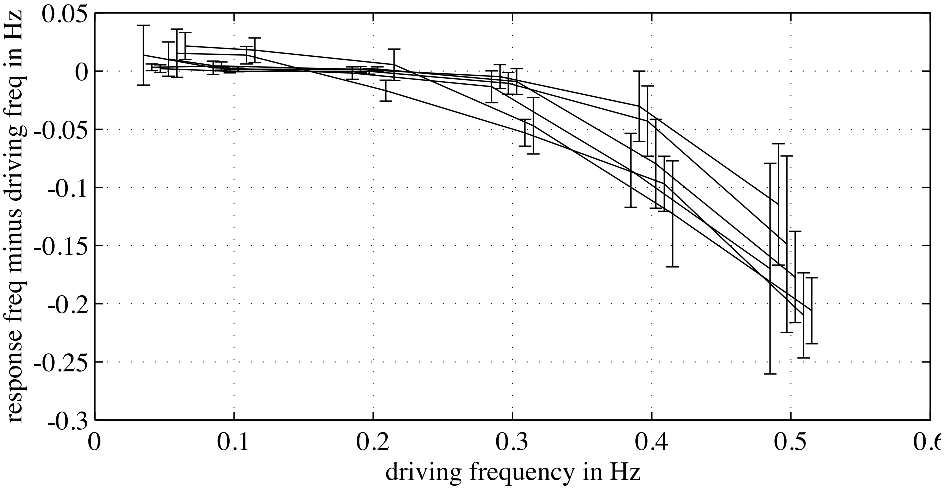


Figure 3–11 Difference between mean frequency of response and driving frequency as a function of driving frequency for all subjects. The curves for different subjects are horizontally displaced for better viewing. The errorbars denote standard deviation.

3.4.3 Relative phase dynamics

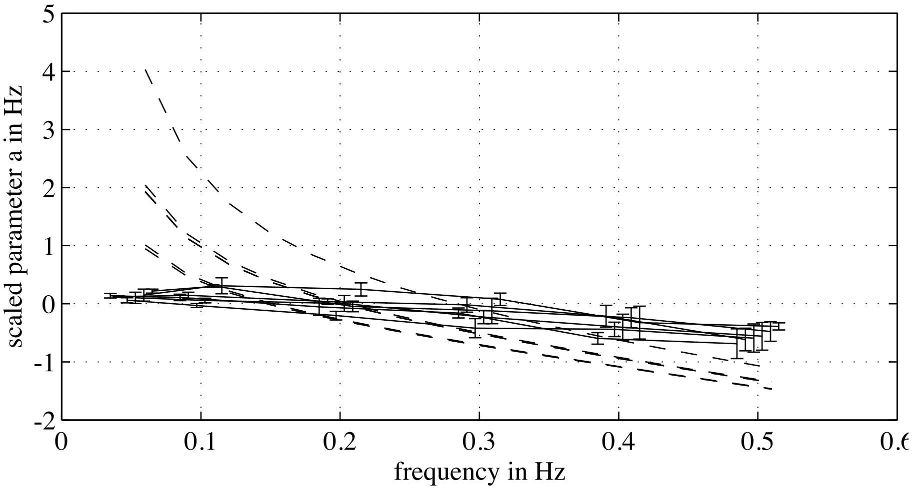


Figure 3–12 Parameter $\tilde{a}$ of return map fit as a function of driving frequency for all subjects. The curves for different subjects are horizontally displaced for better viewing. The errorbars denote standard deviation.

In figure 3–12 I have plotted the scaled model parameter $\tilde{a}$ as calculated by fitting the return map to the experimental return map estimates. All subjects have a significant frequency effect for this parameter. In light of the interpretation of $\tilde{a}$ as the difference between driving frequency and eigenfrequency (see section 3.2) the curves of fig. 3–12 are consistent with an eigenfrequency around 0.2 Hz. In fig. 3–12 I have also plotted (broken line) least squares fits using eq. 3.18. The fits resulted in an eigenfrequency (mean over all subjects) of 0.2 Hz, individual eigenfrequencies ranging from 0.15 to 0.28 Hz. However, the bad quality of the fit indicates deviations from the linear model which can be interpreted in terms of adaptation of eigenfrequency[4].

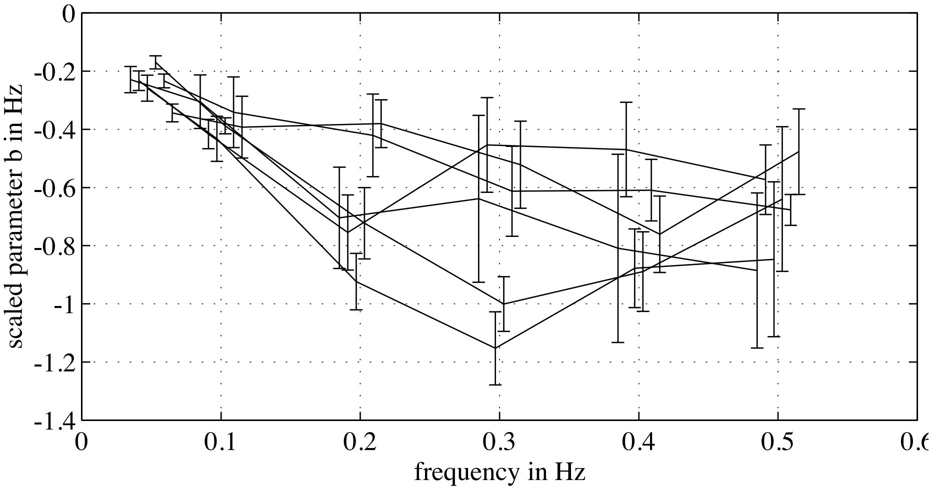


Figure 3–13 Parameter $\tilde{b}$ of return map fit as a function of driving frequency for all subjects. The curves for different subjects are horizontally displaced for better viewing. The errorbars denote standard deviation.

In figure 3–13 I have plotted the scaled model parameter $\tilde{b}$. All subjects except one have a significant frequency effect for this parameter. Because visual distance D_0 and the gain r_0/D_r are relatively constant, we can interpret $\tilde{b}$ as an indicator of the coupling strength between stimulus and response (see eq. 3.19). The results indicate that the coupling strength is low in absolute value for 0.05 Hz and 0.1 Hz and higher for the range 0.2–0.5 Hz. There is a considerable variation between subjects for this measure.

---

[4] This point will be elaborated in future work.

3.4.4 Consistency of fit procedures

I have calculated many characteristics of the data. Some of the characteristics can be calculated in different ways. An important point then is to show that the various ways of calculating a particular characteristic lead to the same values.

First, I checked that ϕ_0 in eq. 3.13 can be set to zero: The proportional reduction in variance brought about by adding ϕ_0 to the fits of the return map was larger than 5% only in exceptional cases. Adding ϕ_0 to the fit of the histogram was useless because only two parameters could be fitted to it reliably, as I argued in section 3.3. Further, as I will show below, the fits with ϕ_0 set to zero, generally reproduced the other observables nicely. This agreement in parameter estimates between different procedures was less when I included ϕ_0 in the fit.

Second, I showed that the results of the fit of the return map and the histogram were in agreement. Note that for the histogram I fix one of the parameters to the value from the return map fit (see Methods): when I find $|a| < |b|$ in the return map fit, I fit a and Q from the histogram, otherwise I fit a and b. For the parameter a, the agreement was excellent, with correlations in the range 0.87–0.99. For the parameter b, the agreement was even better, with the correlations in the range 0.9–0.99. For the parameter Q I find lower correlations in the range 0.69–0.96. This is probably because the return map fit overestimates the noise when the system is close to the bifurcation. Especially the phase values off the attractor cause this high noise estimate.

Having obtained a and b from the fits I can calculate the mean phase with eqs. 3.14 and 3.15, provided that $|a| < |b|$ (i.e. that there exists a fixed point). The correlations between this predicted mean phase and the measured mean phase for each subject are in the range 0.84–0.99.

Likewise, based on a, b and Q from the fits I can determine the angular deviation with eq. 3.20, provided that $|a| < |b|$. The correlation between predicted and measured angular deviation for each subject are in the range 0.82–0.94.

A non-trivial result is that the inverse of the relaxation time correlates significantly with the local stability measure $\tilde{\lambda}$. We cannot expect these correlations to be very large because the frequency manipulation has no significant influence on relaxation time (see fig. 3–9). I find correlations ranging from 0.17 to 0.65, mean 0.38. Half of these correlations are significantly different from zero at a significance level of 5%. For a graphical impression of this correlation see fig. 3–14, where I have made scatterplots of the inverse of relaxation time versus $\tilde{\lambda}$ for all subjects. The connection between $\tilde{\lambda}$, obtained on the basis of fitting the return map and hence incorporating only temporally contiguous phase values, and relaxation time, obtained on the decay of relative phase over multiple time units

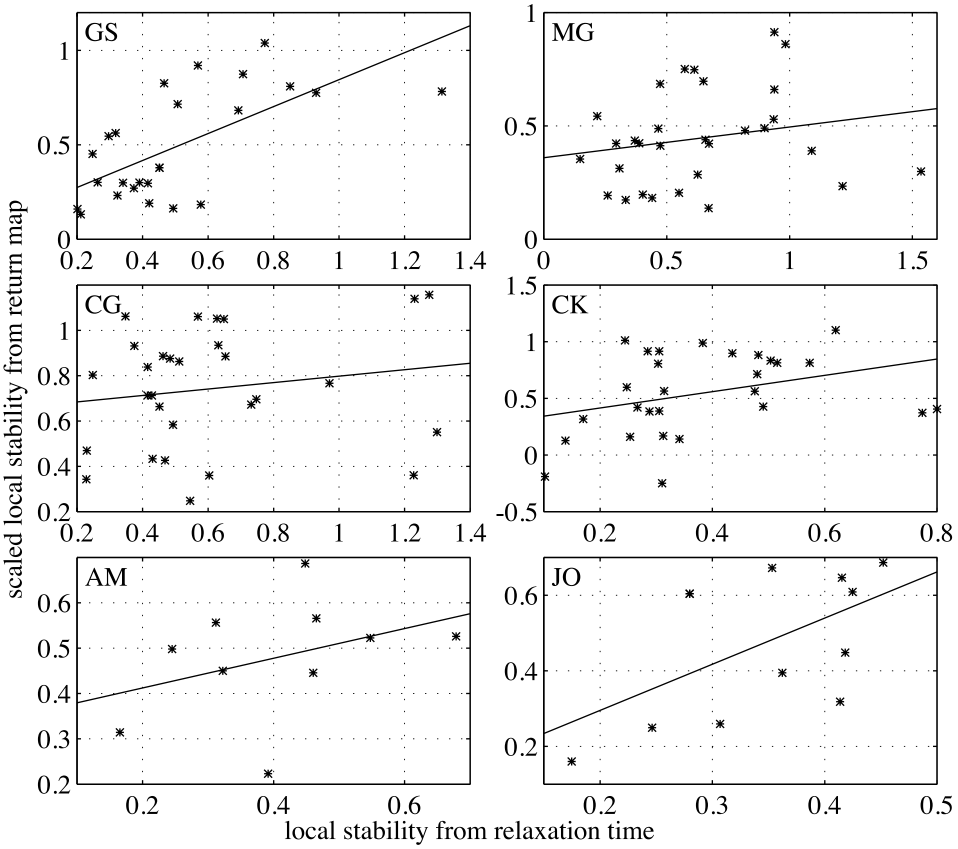


Figure 3–14 Scatterplot of the correlation between inverse of relaxation time and the local stability measure $\tilde{\lambda}$ as calculated from the return map fit. The six panels are for six different subjects. The solid lines denote the best fit of a linear regression.

shows that the fitted dynamics correctly characterises the dynamic properties of the system.

A final result is that the winding number correlates significantly with the global stability measure γ. I find correlations ranging from 0.58 to 0.92, mean 0.8. All of these correlations are significantly different from zero at a significance level of 5%. For a graphical impression of this correlation see fig. 3–15, where I have made scatterplots of the winding number versus γ for all subjects. The high correlation is partly caused by the fact that the sign of a correlates highly with the sign of the winding number. One can also see from the plots that the

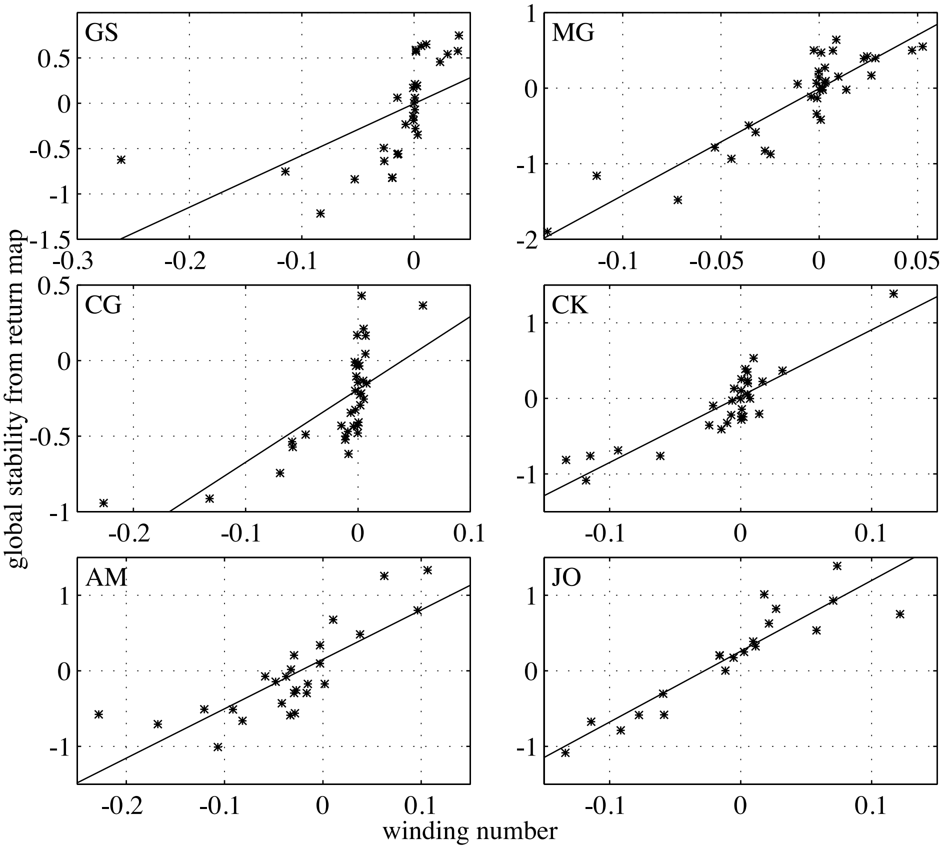


Figure 3–15 Scatterplot of the correlation between winding number and the global stability measure γ as calculated from the return map fit. The six panels are for six different subjects. The solid lines denote the best fit of a linear regression.

winding number is usually zero for a range of global stability values around zero (especially clear for subjects CG and CK). The relation between winding number and the parameters of the sine-circle map is not known for the noise levels we observe[5]. From simulations I have the impression that the observed correlations are in the range of what can be expected.

---

[5] It is an interesting unsolved problem to calculate the winding number of the stochastic sine-circle map.

3.5 Discussion

I reported the results of experiments on visually induced postural sway. My focus was on the coordination patterns between stimulus movement and postural response. To characterise this relationship I measured the relative phase between sway movement and the sinusoidal movement of the visual scene. This provided me with a time series characterising the temporal evolution of this relationship. I manipulated the frequency of the stimulus, keeping constant the mean absolute velocity. This allowed me to explore the limits of the Schöner-model. In general, I found the linear model, where the postural control system is modeled as a passive system, to be a bad approximation of the data. Instead, I obtained evidence for active generation of postural sway at three levels of analysis: (1) the phenomenology of observed coordination patterns, (2) the frequency dependence of measures of temporal order, and (3) the estimated underlying dynamics of relative timing, using the stochastic sine-circle map.

In relation to the observed patterns of coordination of sway and visual motion, I found a rich variety of patterns, which can be described using the vocabulary of von Holst [82]. For the middle frequencies (0.2–0.3 Hz) we observed almost exclusively *absolute coordination*, i.e. stimulus movement and postural response were phase locked. This type of coordination was also observed for low (0.05–0.1 Hz) and high (0.4–0.5 Hz) frequencies. I found *relative coordination*, where the stimulus clearly influences the response but the influence is not sufficiently strong to establish phase locking, occasionally at the higher frequencies. During relative coordination, the system remains locked at relative phase values $\pm\pi/2$ for periods of time with little systematic change (slow dynamics). Occasionally, locking is lost and phase wrapping is observed (fast dynamics), until the system is again caught into a locked state near $\pm\pi/2$. Finally, we also observed *absence of coordination*, during which there is no observable influence of the stimulus on the response in terms of relative timing (but in terms of movement amplitude and frequency). This pattern was occasionally observed at the low and high frequencies. Both relative coordination and absence of coordination cannot arise in a passively driven system. Such systems always respond with the driving frequency and thus are phase locked. Observing these patterns is therefore evidence that the postural control system is actively generating postural sway, a nonlinear process in dynamic terms. The existence of relative coordination indicates that such actively generated sway is dynamically coupled to the time structure of the stimulus. Thus preferential relative timing may arise, but when the system cannot follow the stimulus time structure it wraps rapidly and then tries to lock again.

The dependence of mean phase difference between stimulus and response on frequency was similar to that reported by Lestienne et al. [43] and van Asten et al. [76]. In contrast to these studies much smaller amplitudes of visual motion (relevant for posture in a resting visual world) were used in the present study (the amplitudes employed by Lestienne et al. and by van Asten et al. might be relevant for tasks such as walking.) Talbott's [72] study on standing dogs used more comparable amplitudes. These results have conventionally been interpreted as supportive of the notion that the posture control system can be described by a second order linear (and hence, passive) system. Our results reveal that this might not be a valid conclusion. Mathematically, driven nonlinear oscillators may behave similar to a driven linear system in terms of the mean relative phase.

Hints at nonlinear behaviour are obtained from the dependence of the angular deviation of relative phase on frequency. Its increase towards higher and lower frequency points at a decrease of relative timing stability which is not predicted by the Schöner-model. Relaxation time, the time it takes the system to establish phase locking after a perturbation, did not change with frequency. This would hint at constant stability of relative timing independent of frequency. However, this result as well as the weak correlation of the two stability measures, relaxation time and angular variance of relative phase, could be due to underestimation of relaxation time for states of lower stability. Corroborating evidence for the decrease of stability at higher and lower frequencies comes from the increase in the number of nonlocked trials at these frequencies.

I investigated a dynamical model of the relative phase time series based on the sine-circle map. This model contains all observed coordination patterns as qualitatively different solutions: absolute coordination in terms of a stable fixed point, relative coordination when the fixed point attractor is close to a tangent bifurcation and absence of coordination sufficiently far beyond the bifurcation. Our attempts to directly determine the dynamics of relative phase underlying the action-perception coupling were highly successful. I was able to reliably fit the parameters (a, b, and Q) of the dynamical model both on the basis of the return map as well as on the basis of the relative phase histogram with convergent results. Moreover, I was able to predict further measurables from the estimated dynamics. I found, for instance, that the inverse of relaxation time could be predicted from the local stability measure $\tilde{\lambda}$ as obtained from a fit of the return map.

At the level of the estimated model parameters, the fact that these parameters indicate a bifurcation at increasing frequency is evidence for the active, nonlinear nature of the postural control system. As mentioned earlier, linear relative phase dynamics cannot display such bifurcations. Additional evidence comes from the details of the frequency dependence of the model parameters. The parameter a which can be interpreted in terms of the difference between the eigenfrequency of the postural system and the driving frequency, shows roughly the same behaviour as the mean phase. Eigenfrequency is higher than driving frequency at low driving frequencies and reversely at high driving frequencies. However, if the frequency difference predicted by the linear model is fitted to the model parameter a as obtained at each frequency, a bad fit (around a mean eigenfrequency of 0.2 Hz) is obtained. This suggests that the eigenfrequency is not a constant, but that it is adapted to the driving frequency. The parameter b which can be interpreted as the coupling strength between visual motion and postural response shows an increase in absolute value with frequency.

An interesting open question is why subjects display such a great variety of dynamical behaviours at the extreme frequencies. Some subjects show all three types of coordination for different repetitions of the same frequency condition. I tried to relate the behaviour in a trial to the driving frequency of the preceding trial (testing for some kind of hysteresis) but never found an effect. Presumably, if there is such a form of adaptation at all, it has already taken place in the first 20 s of each trial which was not recorded. It would be an interesting experiment to switch frequency within a trial and see how subjects adapt. In view of the theoretical results it is important to keep in mind, that near the bifurcation, small changes in the parameters of the underlying dynamics (which might occur spontaneously even at constant frequency in the form of parametric fluctuations) lead to large and even qualitative differences in the resulting solutions. In this respect it is remarkable, that even as the types of solutions vary, the estimated model parameters change continuously (cf. Figures 3–12 and 3–13).

In summary, from my results I conclude that temporally structured postural sway is actively generated. The largely periodic temporal structure of postural sway reflects perceived or adapted-to parameters of visual motion. Coupling to visual motion is used to generate phase locked postural sway that minimises motion relative to the visual world. Phase locking to visual motion is not, however, a prerequisite to postural sway with a significant amplitude.

Chapter 4

Perception of 3D shape from ego- and object-motion

abstract

I investigated whether ego-motion information (knowing where and how fast you are moving) is used directly in the perception of 3D shape. I compared the performance of curvature detection in large field and small field views of 3D spheres and planes in three conditions. (1) Observer movement, where the observer moved parallel to the screen and the object was simulated by feedback of the eye position. (2) Object translation, where the observer was stationary and the object moved the same relative to the observer as in the previous condition. (3) Object rotation, where the observer was also stationary and the object rotated in depth generating the same image on the optic array as in the observer movement condition. I found performance for detecting the presence of curvature (disregarding the sign) to be the same in the observer movement and object rotation conditions indicating that ego-motion information is not used directly in shape perception. Subjects performed well in the object translation condition with large field stimulation, but were at chance level with small field stimulation. This indicates that ego-motion information may be used to stabilise the image of the object on the retina. I found a depth ambiguity for small field presentation of rotating objects, when subjects reported the sign of curvature at random. As no such ambiguity was found for the other conditions, I concluded that ego-motion information helps to disambiguate the sign of curvature.

This chapter is an expanded version of the paper "Perception of 3D shape from ego- and

4.1 Introduction

A new field of research in the computer graphics and computer vision communities is devoted to active vision i.e vision by an actively moving observer [1]. Within computer vision active vision is seen as a means for a robot to extract 3D information from the environment by using ego-motion information from nonvisual sources in the evaluation of visual information. However, it is unclear whether and to what extent human observers use nonvisual information in this direct way. So far only a few psycho-physical studies have been performed in this field and very few comparisons between the perceptual effects of active and passive vision have been made.

The relationship between active movements and 3D shape perception has been pioneered by Rogers and Graham [61] who simulated a corrugated surface on an oscilloscope screen. The spatial pattern consisted of a Julesz-pattern. The motion of the dots in the Julesz-pattern was linked to the movements of the observer (subject movement): some horizontal lines of dots moved with the observer and some other lines moved in opposite direction, creating a compelling view of a surface with vertical corrugations. They also did the experiment with movement of the oscilloscope and linked the motion of the dots to the oscilloscope movement (object translation). They found that the perceived depth of a surface is about 15% higher when the motion parallax is generated by active movements of the observer rather than by movement of the stimulus presented to a stationary observer.

A good interpretation of the decreased performance in the object-translation condition is difficult for three reasons. First, head movements in the active condition were not stored. Therefore, the movement of the stimulus relative to the head might have been different in the subject-movement and object-translation conditions. Second, it is known [23] that the fixation of a point of the stimulus is better in the subject-movement condition than in the object-translation condition. Also, the otolitho-ocular reflex might contribute to a better retinal image stabilisation during ego-motion [10]. This leads to two possible explanations for the finding that perceived curvature is higher for subject-movement than for object-translation. One explanation might be that subjects make use of proprioceptive information in the subject-movement condition, which is unavailable in the object-translation condition, in the evaluation of the afferent visual information. An alternative explanation, advanced by Cornilleau-Pérès and Droulez [14], might be that there is more retinal slip in the object-translation condition, which

---

object-motion: comparison between small and large field stimuli" by Tjeerd Dijkstra, Valérie Cornilleau-Pérès, Stan Gielen and Jacques Droulez, Vision Research 35, 453–462 (1995).

makes curvature detection harder. Thus it is of interest to add a condition where the retinal slip is reduced by combining the object translation with a rotation around the observer's eye, mimicing a perfect retinal stabilisation through eye movements. The resulting object movement is a rotation in depth. Third, there was no fixation point in the stimuli of Rogers and Graham. This is of importance since it has been demonstrated recently that performance in detection of the sign of curvature is dependent on fixation [28].

Therefore, Cornilleau-Pérès and Droulez [14] have compared the sensitivity for the detection of curvature of a moving surface for the conditions of subject-movement and object-movement. They constructed the experiment so that the relative movement between the object and the observer was identical in all conditions. They tested three experimental conditions: 1) in the subject-movement condition the observer moved his head sinusoidally in left/right direction, viewing a stationary 3D object, 2) in the object-translation condition the subject's head was fixed and the object translated sinusoidally by the same amount in left/right direction, 3) in the object-rotation condition the subject's head was fixed and the object rotated in depth. This last condition was obtained from the previous one by adding a rotation of the object around the observer's eye to the motion of the object. This extra motion is devoid of any depth information for the observer. The resulting movement of the object is a rotation in depth around the fixation point. The movement involves the same relative movement between observer and object as in the two other conditions, but no eye movements are needed to stabilise the image of the object. The stimuli, with a diameter of 8 deg, were either planar or convex with a fixed curvature and the observer's task was a forced choice between planar and convex. The results showed that curvature sensitivity is much higher in the subject-movement condition than in the object-translation condition, and that the object-rotation condition yielded the best performance. To explain their findings, the authors invoked the global image motion which results from different oculomotor behaviour in the three conditions, as the main factor which determines the performance. This explanation is based on the fact that global image motion is known to impair the visual sensitivity to differential velocity [51], and that the detection of surface curvature from motion is likely to be mediated by the processing of the spatial variations in image velocity [12].

The optic flow field plays a double functional role in visual perception: it provides an observer with exteroceptive information about the structure (distance, slant and curvature) and motion (velocity and rotation-rate) of objects in the environment as well as with proprioceptive information about the movements of the observer in the environment (velocity and rotation rate). From theoretical studies on optic flow processing it is known that the parameters of the relative motion between observer and object cannot be separated from the recovery of the structure of the object (see the next chapter). Hence, the visual system could take advantage of self-motion to improve its ability to solve the problem of structure from motion.

The first goal of the present experiment is to investigate whether proprioceptive ego-movement information (knowing where and how fast you are moving) is used directly in the perception of shape. Self-motion is processed both from non-visual information such as efference copies and vestibular signals and from visual information. Although different variables interact in the perception of self-motion, the size of the stimulus is one of the major factors which influences vection or the control of stance (for a review see [83]). In particular, when a lamellar flow field due to the frontal translation of a plane is presented in central vision, Stoffregen [69] finds that compensatory body sway is very small for a stimulus width of 20 deg, and increased much as this width reaches 40 deg. Similarly, Post [56] shows a large reduction of circular vection when the stimulus size was 30 deg wide, rather than full-field. Therefore, the small stimuli (8 deg diameter) used by Cornilleau-Pérès and Droulez were poor in terms of visual information about self-motion. In order to create a stronger impression of self-motion I extended the experiment of Cornilleau-Pérès and Droulez to large field stimuli (90 deg visual angle).

Since I found that proprioceptive information is not used in a direct way in the perception of shape, it is natural to ask whether the ego-movement information is not used at all. A possible use for this information may be to assist in fixating a certain point on the object. This is important because detection of differential image motion is known to deteriorate for common image motion [51]. Moving observers can use both the otolitho-ocular reflex and the optokinetic nystagmus to maintain a stable fixation. Stationary observers viewing a translating object can only use optokinetic nystagmus (OKN). Since the gain of OKN is known to depend on the field of view [80] (the gain is closer to 1 for a larger field of view) I have tested the performance for shape perception in the three conditions mentioned above for stimuli with a large diameter (90 deg) and small diameter (8 deg). If stimulus size improves fixation, a smaller difference in performance between different movement conditions is expected for large-field stimuli than for small-field stimuli. This leads to the second goal of the present experiment: is ego-motion information used to stabilise the image, thereby improving the detection of shape? If ego-motion information is used, then we expect performance for ego-motion and object-translation to be about equal for a large field of view (OKN-gain is close to 1) whereas we expect performance for ego-motion to be higher than for object-translation for a small field of view (OKN-gain is lower than 1).

In the object-rotation condition, I noticed an ambiguity between concave and convex spheres. This ambiguity was already reported by Hayashibe [28] and Rogers and Rogers [64]. Rogers and Rogers find that both perspective and non-visual information about self-motion contribute to raise this ambiguity. The third goal of this paper was thus to compare the efficiency of self-motion and perspective information in raising a depth ambiguity. Hence, instead of asking the subject to report only the presence of absence of surface curvature, I also required that he reported the sign of curvature.

4.2 Methods

Wide-field and small-field experiments were performed in different laboratories. I therefore start with a separate description of each of the set-ups.

4.2.1 Experimental set-up for large-field stimulation.

The set-up is described extensively in section 2.2.1. The only thing that was different is that the stimuli were viewed monocularly. Thus, the stimuli were green (phosphor p53) and had a luminance of 0.5 Cd/m^2.

4.2.2 Experimental set-up for small-field stimulation.

The stimuli were presented on the monitor of a Silicon Graphics workstation (resolution 1280x1024 pixels, frame rate 60 Hz). The stimuli were white (phosphor p22), had a luminance of 1.4 Cd/m^2 and were presented at a rate of 30 Hz (each frame is displayed twice).

The subject was sitting at a distance of 72 cm from the monitor with one eye covered. He had a light-weighted helmet on his head on top of which was fixed a mobile bar. The weight of the bar was sufficiently small so as not to hamper head movements. It was mobile in a pulley with very low friction, and could therefore translate along itself. The pulley could rotate around the vertical and horizontal axes passing through its centre. Three potentiometers delivered analog signals linearly or sinusoidally related to each of the translations of the head (up-down, left-right and backwards-forwards). These signals were converted to digital by a microcomputer and were then sent to the workstation through an RS232 bus at a rate of 9600 baud. The workstation was programmed to generate a video image of a 3D shape, viewed from the current position of the eye. The delay in the feedback-loop was 55 ms.

The microcomputer was used to calibrate the three head translation signals. Repeated calibrations performed on 105 points lying within a parallelepiped centred on the median subject's head position (30 cm in horizontal, 20 cm in vertical, 6 cm in depth) showed that the mean error on head position is 1.7 mm, with a maximum of less than 5 mm. A restricted calibration was performed prior to each experiment, in order to estimate the potentiometer offsets and gains that could vary in time.

4.2.3 Stimuli

Because of different technical constraints, the parameters of the large-field stimuli (hereafter LF) and small-field stimuli (hereafter SF) are not precisely the same. However, as shown in table 4–1, they are generally sufficiently similar so that the two experiments remain comparable.

field of view deg	subject	movement frequency Hz	peak-to-peak amplitude cm	SD p-to-p amplitude cm
90	TD	0.33	21.8	2.1
90	MG	0.33	23.1	2.4
90	PS	0.33	19.8	2.2
8	TD	0.33	26.4	3.3
8	VCP	0.33	22.9	1.3
8	OV	0.5	19.5	2.8

Table 4–1 Comparison of some experimental parameters and movement characteristics for large-field and small-field stimulation.

Stimuli were curved or flat surfaces covered with 300 (LF) or 400 (SF) random dots, each of diameter 0.2 deg (LF) or 0.02 (SF) deg. The distribution on the surface was such, that the density of dots was uniform per solid angle. This was done to minimise the possibility to use the local density of dots as a feature to estimate the curvature of the surface. The large-field stimulus covered a range between 2 and 45 degrees of visual eccentricity (field of view 90 deg), and had a fixation cross of 2 by 2 deg at the centre. The small-field stimulus covered a range between 0 and 4 degrees of visual eccentricity (field of view 8 deg), and had a bright fixation dot of diameter 0.05 degrees at the centre. The shape of the large-field stimulus was a section of a sphere which could have a curvature of -0.67 m^{-1}, -0.33 m^{-1}, -0.17 m^{-1}, 0 m^{-1}, 0.17 m^{-1}, 0.33 m^{-1} or 0.67 m^{-1}. The shape of the small-field stimulus was a section of a sphere which could have a curvature of -5 m^{-1}, -4 m^{-1}, -2.85 m^{-1}, 0 m^{-1}, 2.85 m^{-1}, 4 m^{-1} or 5 m^{-1}. Negative curvatures denote concave sphere segments, curvature 0 denotes a plane and positive curvatures denote convex sphere segments. It should be noted that the rim of the stimulus is a planar curve which has the same projection for all curvatures and hence cannot be used as an artifactual cue.

The stimuli were shown for 6 s in a dark room and on a dark background. The fixation point was the point in the centre of the simulated surface and was straight in front of the subject at the beginning of a trial. The distance from the eye to the fixation point was chosen randomly between 40 cm and 60 cm (LF) or between 75 cm and 85 cm (SF). This made it difficult for the subject to use the mean retinal velocity as a cue for the shape (see subsection about control experiments). At the start of each trial the tangent plane at the fixation point was fronto-parallel. Due to the head-movements the viewing distance and the orientation of the tangent plane changed in the course of a trial.

I compared thresholds of curvature detection in three conditions: a subject-movement condition, an object-translation condition and an object-rotation condition (see fig. 4–1). In the subject-movement condition subjects moved in left/right direction at a frequency of 0.33 Hz (LF, SF) or 0.5 Hz (SF) and with an amplitude of 10 cm. Pilot results and a control experiment on subject VCP have shown the effect of frequency to be very small. A metronome helped the subjects to maintain a constant frequency. The frequency and amplitude of movement were trained at the beginning of each session by giving the subject feedback about his movement. Subjects could readily perform this with a relative standard deviation in amplitude of movement of about 10% (Table 4–1). I stored a time series of the translation of the eye together with the positions of the random dots relative to the eye on disk. This information was used later in the two object-movement conditions to generate the same projections on the optic array.

In the object-translation condition the head of the subject was fixed using a chinrest and the stimulus translated with the translation of the head previously recorded in the subject-movement condition. Thus the subject had to make tracking eye movements in order to fixate the fixation point.

In the object-rotation condition the head was also fixed but the motion of the stimulus was a pure rotation in depth. This rotation was calculated from the previous translation by adding a simulated eye rotation so that the stimulus on the optic array of the subject is the same as in the subject-movement condition. The torsion component of rotation was set to zero: the torsion of the head was negligible and the torsion of the eye was not measured but is known to be small

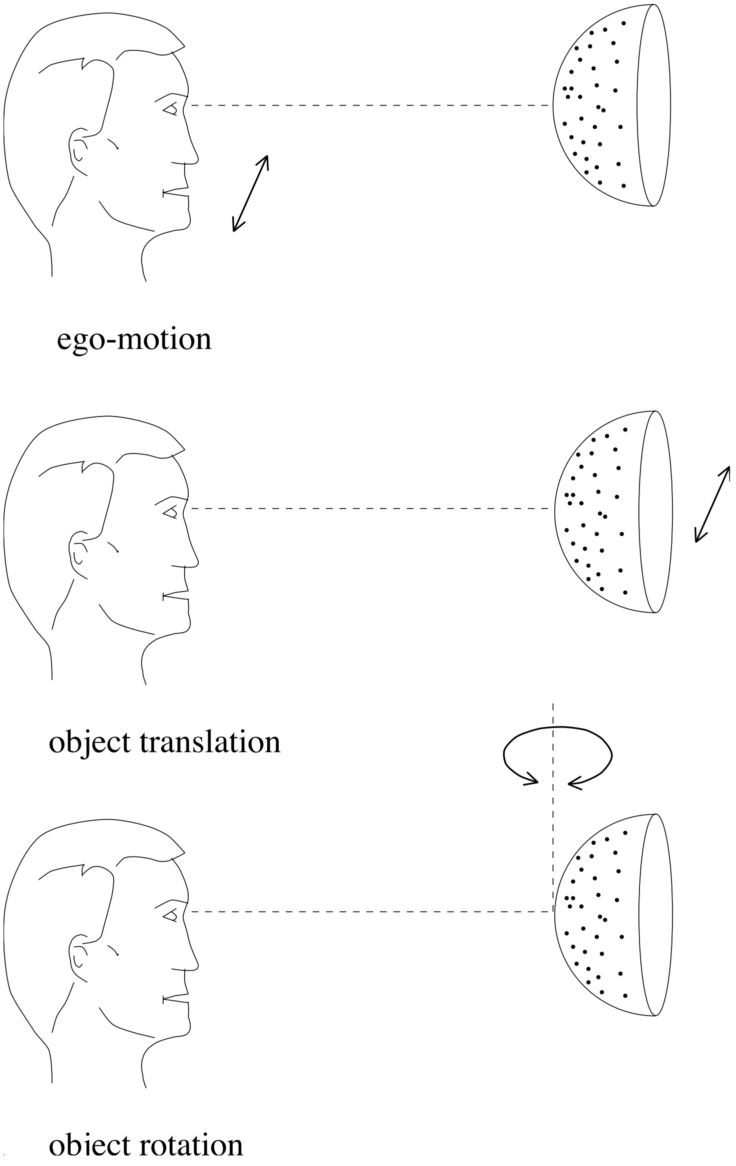


Figure 4–1 Schematic illustration of the different movement conditions.

for eye orientations of up to 10 deg [50].

4.2.4 Protocols

Shape detection

This protocol was used both for large-field and small-field stimulation. My main conclusions are based on the results of this protocol. The stimulus could be either a concave, planar or convex surface with equal probability. The subject's task was to detect curvature in a forced choice between concave, planar or convex. No feedback on performance was given.

Each experiment consisted of 5 sessions that lasted approximately 45 min each. Each session consisted of two subsessions. Each subsession consisted of 6 blocks in fixed order: for the left eye subject-movement, object-rotation, object-translation, then for the right eye subject-movement, object-translation and object-rotation. Each block consisted of 18 stimuli: 2 repetitions for each of the 6 curvatures and 6 repetitions for the plane in random order. So in all sessions together there were 40 repetitions per movement condition and per curvature, 20 for the left eye and 20 for the right eye. For the plane these numbers are three times as high.

Shape discrimination

This protocol was used only for large-field stimulation. I used this protocol to check the result of the detection protocol that the different movement conditions do not lead to a significant difference in performance. The subject was shown a pair of stimuli one of which was planar. The other was concave or convex with equal probability. The stimuli were shown one after the other. The subject's task was to discriminate which of the two stimuli seemed most convex. No feedback on performance was given. With this protocol I obtained a complete psychometric function (see figs. 4–4 and 4–5). The advantage of the discrimination protocol over the detection protocol is that the subject does not have to have an internal reference to separate the stimulus categories. The disadvantage is that the technique cannot cope with ambiguities as found for object rotation with small-field stimulation.

Each experiment consisted of 10 sessions that lasted approximately 30 min each. Each session consisted of 6 blocks in fixed order: for the left eye subject-movement, object-rotation, object-translation, then for the right eye subject-movement, object-translation and object-rotation. Each block consisted of 12 pairs of stimuli: 2 repetitions for each of the 6 curvatures. So in all sessions together there were 40 repetitions per movement condition and per curvature, 20 for the left eye and 20 for the right eye.

I analysed the data by fitting a normalised errorfunction to the data[1]. The normalised errorfunction is the integral of a gaussian distribution with mean μ and standard deviation σ. From this fit I obtained a measure for the point of subjective planarity μ and a measure for the discriminability σ [45]. For fitting, I used both the Levenberg-Marquardt algorithm [57] and the routine CNLR of SPSS for Windows, release 5.01 [53].

With the Levenberg-Marquardt algorithm I minimised the χ^2-measure. I estimated the standard deviation of each data point by assuming an underlying binomial distribution: when the subject responded "convex" a fraction p_i out of N_i trials (N_i is always 40 in this experiment), I estimated the standard deviation by:

$$\sigma_i = \sqrt{p_i(1 - p_i)/N_i}$$

I estimated the errors in the fitted parameters from the covariance matrix.

With routine CNLR I derived a maximum likelihood estimate of the unknown parameters. Estimates of the errors in the fitted parameters were obtained by the so-called bootstrap method [46]. We found the values of the fitted parameters to differ little with either fitting method. The estimates of the errors in the fitted parameters were usually higher with the bootstrap method than with the covariance matrix method, consistent with [46]. Therefore, I will use the results of routine CNLR.

4.2.5 Subjects

Three subjects participated for each field of view in the detection protocol, one of the authors (TD) was tested for both SF and LF stimulation. Three subjects were naive as to the purpose of the experiment (MG, PS and OV). Two subjects participated in the discrimination protocol, one of which was naive as to the purpose of the experiment (JW). All subjects had normal vision or corrected to normal vision wearing contact lenses.

4.3 Results

4.3.1 Shape detection

In figure 4–2 I show the results for large-field stimulation for two subjects, the results for the remaining subject are similar (see table 4–2). In general the stimuli with the largest curvature (both convex and concave) could be perceived with

---

[1] This work was done in collaboration with Peter Werkhoven.

almost 100% accuracy. For smaller curvatures, the probability of a correct response decreased gradually. The percentage of planar responses generally peaks at zero curvature and decreases when absolute curvature becomes higher. The general features of these curves are roughly the same for each movement condition. This is also the message from table 4–2 where I compare the performance in detection of absolute curvature for the different movement conditions. The percentage correct responses (PCR) for large-field stimulation does not differ significantly between the three movement-conditions. Subjects always perceived a rigid shape and found the object-translation condition more difficult than the other two, although their performance is not significantly worse.

field of view deg	sub	diff in PCR OR-SM	sign	diff in PCR SM-OT	sign	diff in PCR OT-ch	sign
90	TD	-1.9	ns	1.9	ns	43.6	$p<0.005$
90	PS	1.9	ns	-3.6	ns	41.7	$p<0.005$
90	MG	0.8	ns	1.1	ns	34.4	$p<0.005$
8	TD	8.6	$p<0.1$	13.9	$p<0.05$	-2.8	ns
8	OV	10.8	$p<0.01$	14.2	$p<0.01$	2.2	ns
8	VCP	5.6	$p<0.05$	27.2	$p<0.01$	4.7	$p<0.05$

Table 4–2 Comparison of the performance of absolute curvature detection (disregarding sign) expressed in percentage correct responses (PCR) of the three movement conditions (SM subject-movement, OT object-translation, OR object-rotation, ch chance level) for large-field and small-field stimulation. The PCR is averaged over all curvatures.

In figure 4–3 I show the results for small-field stimulation for two subjects, the results for the remaining subject are similar (see table 4–2). One subject (TD) was also tested for large-field stimulation (see Fig 4–2). The results for the three movement conditions are very different from one another. The curves for the subject-movement condition are qualitatively the same as for large-field stimulation, albeit that performance is somewhat lower for subject TD. The curves for the object-translation condition are close to chance level, which is very different from the large-field result. In this condition only the percentage of concave responses at a curvature of -5 m^{-1} for subject OV is clearly different from chance level. For object-rotation the percentage of planar responses shows the normal profile with a peak at zero curvature. The width of this curve, which is a measure for the performance of detection of absolute curvature, is smaller in the object-rotation condition than in the subject-movement condition. The other two curves

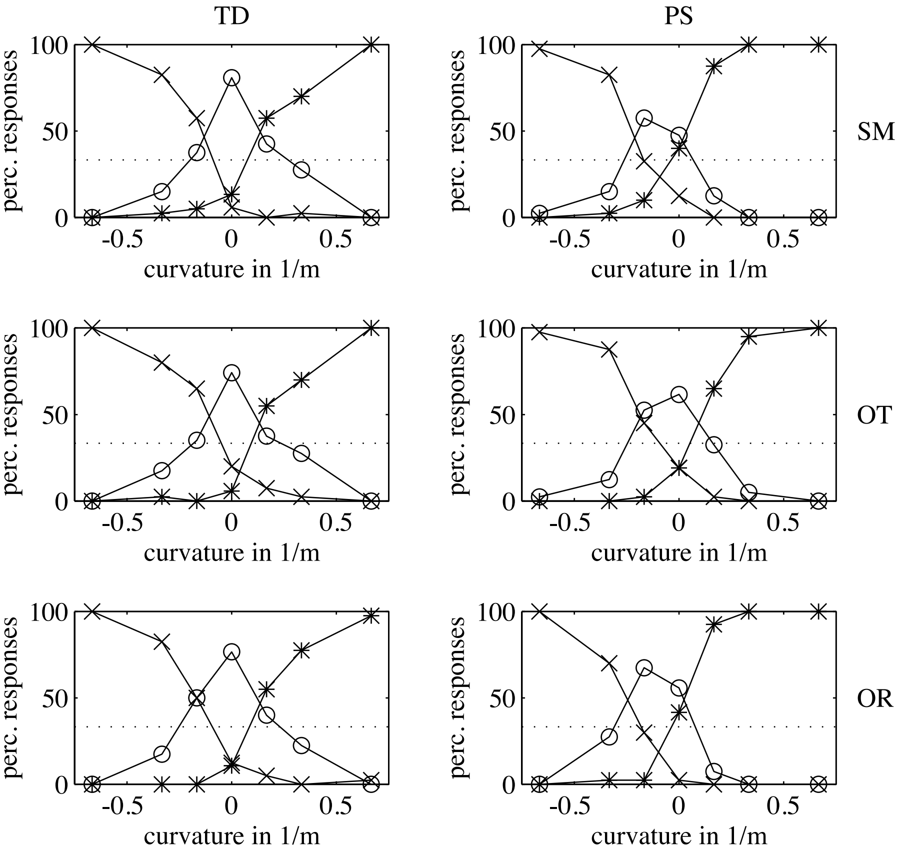


Figure 4–2 Response curves for subjects TD and PS for large-field stimulation in each of the three movement conditions. X: percentage of concave responses, O: percentage of planar responses, *: percentage of convex responses. Left column: subject TD, right column: subject PS. Upper panels: subject-movement (SM), middle panels: object-translation (OT), lower panels: object-rotation (OR). The horizontal dotted line indicates chance level.

did not converge to 100% for the extreme curvatures as for large-field stimulation. Since the number of false positives for planar surfaces is very low for the larger curvatures, this indicates that subjects had no problem in distinguishing a planar or curved surface but had difficulties in detecting the sign of curvature. Table 4–2 shows that performance in the detection of absolute curvature is best in the object-rotation condition, and that the subject-movement condition yields slightly worse performance than the object-rotation condition. The differences between the three conditions are significant to the level $p < 0.05$, and performance in the object-translation condition does not, in general, differ from chance to the level of $p < 0.1$.

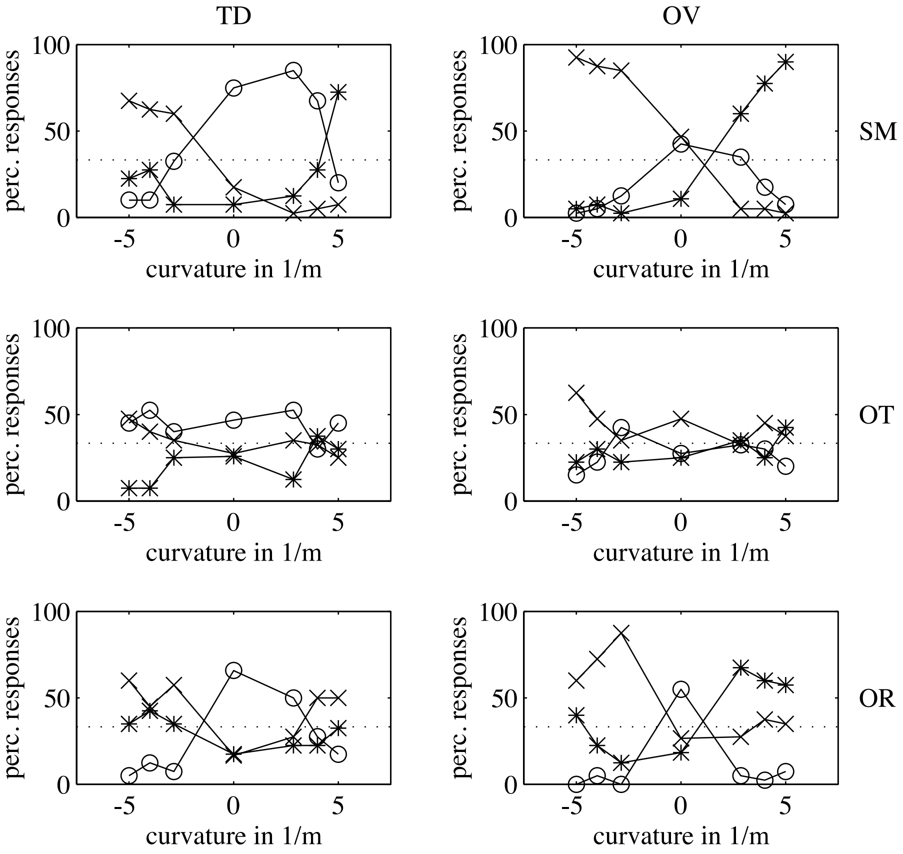


Figure 4–3 Response curves for subjects TD and OV for small-field stimulation in each of the three movement conditions. For details see the legend of fig. 4–2.

In order to compare the performance in large-field stimulation with the performance in small-field stimulation I calculated that the response curves of large-field stimulation need to be scaled in curvature by somewhat more than a factor 10 to be comparable with the response curves of small-field stimulation. For subject TD, the only subject tested in both set-up's, the factor is 17 for the percentage of planar responses in the object-rotation condition.

Subjects perceived the shape always rigid in the subject-movement condition with small-field stimulation. They found the object-translation condition with small-field stimulation to be very difficult, which is reflected in their performance being near chance level. Subjects sometimes perceived the shapes as being hyperbolic i.e. of negative gaussian curvature. In that case they reported the sign of the largest absolute curvature. Subjects frequently reported apparent deformations of the stimulus for the object-rotation condition with a small field of view. This is reflected in the large number of errors in the sign of the curvature for object-rotation. The percentages of curvature-inversions for this condition were: TD 40.0%, VCP 24.2% and OV 29.2%. From the response curves in the lower panels of figure 4–3 one can even see a more surprising fact: the percentage of reversals does not decrease with increasing absolute curvature[2]. Thus, on one hand performance for detection of absolute curvature becomes better with increasing curvature, because the percentage of planar responses becomes smaller for larger absolute curvatures. On the other hand the percentage of convex responses for concave stimuli is seen not to decrease with decreasing curvature. The same holds true for the percentage of concave responses for convex stimuli. This effect was also noticed by Braunstein and Tittle [8] who reported an increase in depth ambiguity as motion parallax and the amount of perceived depth increase.

In summary, the main results are that the performance for curvature detection in the three movement conditions is roughly constant for large-field stimuli but shows great differences for small-field stimuli. Small-field stimuli give a reasonably good performance for the subject-motion condition and a performance near chance level for the object-translation condition. Although the distinction between a planar or curved surface could be made very well in the object-rotation condition, ambiguity between convex and concave surfaces was present for the small-field stimuli.

4.3.2 Shape discrimination

In figures 4–4 and 4–5 I show the results for large-field stimulation for two subjects. First of all, it is clear that the errorfunction is a good approximation of the data. From this we can conclude that it is *curvature* itself that is extracted and not, say, the logarithm of curvature. This is also confirmed by another fit where we fitted a normalised errorfunction based on an unknown power of curvature. The fitted exponent of curvature did not differ significantly from 1.

In table 4–3 I show the resulting parameters and their estimated errors. In general, the parameter μ does not differ significantly from 0 (the only exception

---

[2] This is also the case for subject VCP.

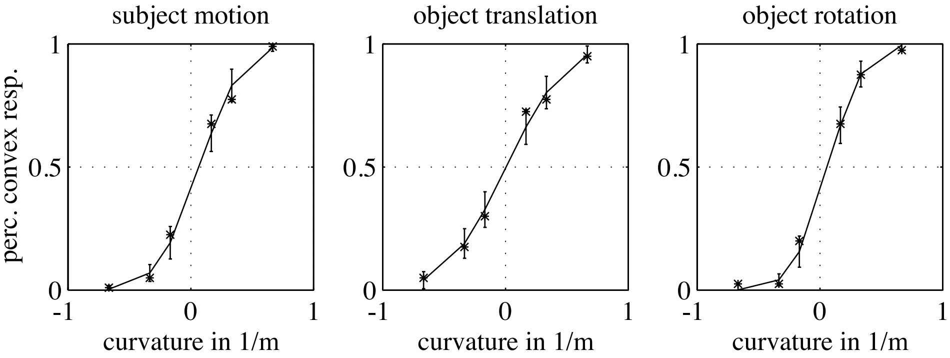


Figure 4–4 Response data (denoted by *) and fitted psychometric function for subject TD for large-field stimulation in each of the three movement conditions. The errorbars denote the estimated standard deviation.

sub	SM μ	SM σ	OT μ	OT σ	OR μ	OR σ
TD	0.07±0.02	0.29±0.02	0.01±0.02	0.39±0.02	0.08±0.04	0.30±0.03
JW	0.01±0.05	0.42±0.07	-0.06±0.03	0.50±0.05	-0.05±0.03	0.42±0.06

Table 4–3 The parameters of the psychometric functions (in m^{-1}) in the three movement conditions (SM subject-movement, OT object-translation, OR object-rotation).

being subject TD in the subject-movement condition). Because μ gives the point of subjective planarity we can conclude that subjects perceive planes as planes.

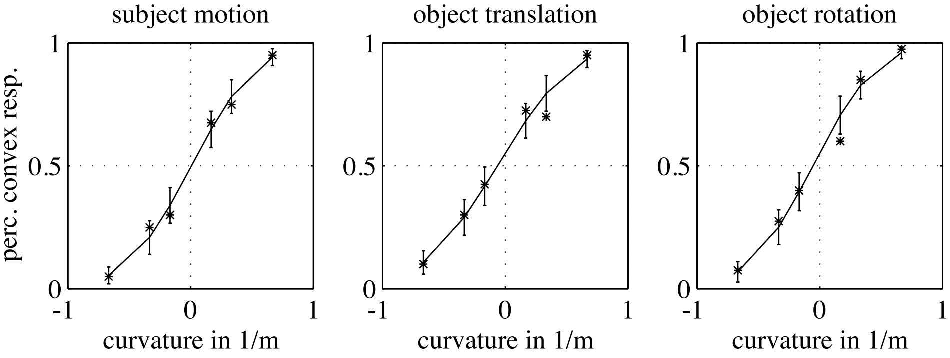


Figure 4–5 Response data (denoted by *) and fitted psychometric function for subject JW for large-field stimulation in each of the three movement conditions. The errorbars denote the estimated standard deviation.

For parameter σ, which gives the discriminability there is a difference between the movement conditions. For subject TD there is a significant difference between the object-translation condition on one hand and the subject-movement and object-rotation on the other. Although performance is also lowest in the object-translation condition for subject JW, the difference with the other conditions is not significant. I speculate that this small difference is caused by a slightly larger retinal slip in the object-translation condition (see discussion).

4.3.3 Control experiments

First I performed a control session with the set-up of the large-field stimulation but using the same field of view as in the small-field stimulation, using the monitor of the SUN4. In this session I took the curvatures precisely 10 times as high as in the large-field stimulation, thus covering a larger range of absolute curvatures than the small-field stimuli. This gave essentially the same results as I found for small-field stimulation. The only difference was a clear above chance level performance with no depth reversals for the larger curvatures in the object-translation condition. This shows that the differences in set-up are unimportant in a comparison of the results of large-field and small-field stimuli.

Our experiment is based on the assumption that the subjects really report on the 3D structure that they have perceived. The need for verifying that no other cue incidently related to surface curvature is used by the subjects, has recently been underlined in a series of papers [9, 68]. There are two main cues that could be used as artifactual cues by the subjects in experiments on the perception of surface curvature from motion: 1) the spatial variations of dot density that occur in the extreme positions reached by the object relative to the observer (recall that the dot density was uniform per solid angle for the median position), 2) the magnitude of the image velocity of the moving dots relative to the observer.

In the object-rotation condition the position of the dots relative to the eye is exactly the same as in the other two conditions, if a perfect stabilisation of the image occurred (if the subject kept fixating the fixation point). Therefore we can assume that if the subject based his responses on any of the above artifacts, then he could do it particularly well in the object-rotation condition for which image stabilisation is the easiest. So I restrict the discussion to this latter condition.

For large-field stimulation I did a control experiment in which subject TD was shown the extreme positions of the object for 3 s each. Otherwise the stimuli were equal to the large-field stimuli of the experiment. The responses did not significantly differ from chance level. I conclude that there is little evidence that the static variations in dot density are used as a cue for curvature.

In order to test whether the mean velocity magnitude of dots on the retina of the subject is used as an artifactual cue, I calculated it for each trial for large-field stimulation. The mean retinal flow i.e. the spatial and temporal mean of the retinal velocity is plotted in the upper panels of figure 4–6 for two subjects. As one can see there is a large overlap in mean retinal flow for each of the curvatures. There is still a noticeable trend of mean flow with curvature, especially the sphere with curvature 0.67 m^{-1} leads to a higher retinal flow. Note that the flow is relatively small for the concave spheres, which are on average closer to the subject and is relatively high for the convex spheres which are on average farther away. This is because retinal flow depends both on distance from the subject and on the 3D velocity. The latter is higher at a given eccentricity for a convex sphere than for a concave sphere. From the data in figure 4–6 I have calculated response curves based on the assumption that the subject used mean retinal flow for curvature detection. To that purpose I have subdivided the mean retinal flow from each trial in 8 categories with the smallest 45 mean flow values in the first category, etc. I counted the percentage of concave, planar and convex responses in each of these categories. In the lower panels of figure 4–6 I have plotted these percentages vs the mean of the 45 flows in each of the categories. Clearly, the results are very different from the results reported in figure 4–2. Although there is a tendency of the percentage of concave (convex) responses to decrease (increase) with mean retinal flow, the effects are much smaller than in figure 4–2. I also did the above analysis with the maximum retinal flow and found essentially the same. I conclude that there is little evidence that subjects used retinal flow as an artifactual cue for curvature detection.

For small-field stimulation Cornilleau-Péréès and Droulez [13, 14] performed control experiments which showed that the mean retinal velocity and the spatial variations of dot density in individual images were not used as artifactual cues in experiments involving the discrimination between convex spherical surfaces and planes. Another argument against the use of mean retinal velocity as a bias for the detection of curvature with small-field stimulation is provided by the high number of curvature inversions (lower panels of figure 4–3) in the object-rotation condition. Since the mean dot velocity is a monotonous function of sphere curvature, the use of this velocity as an artifactual cue should result in a better discrimination between convex and concave spheres, than between spheres and planes, which is the opposite of what is observed.

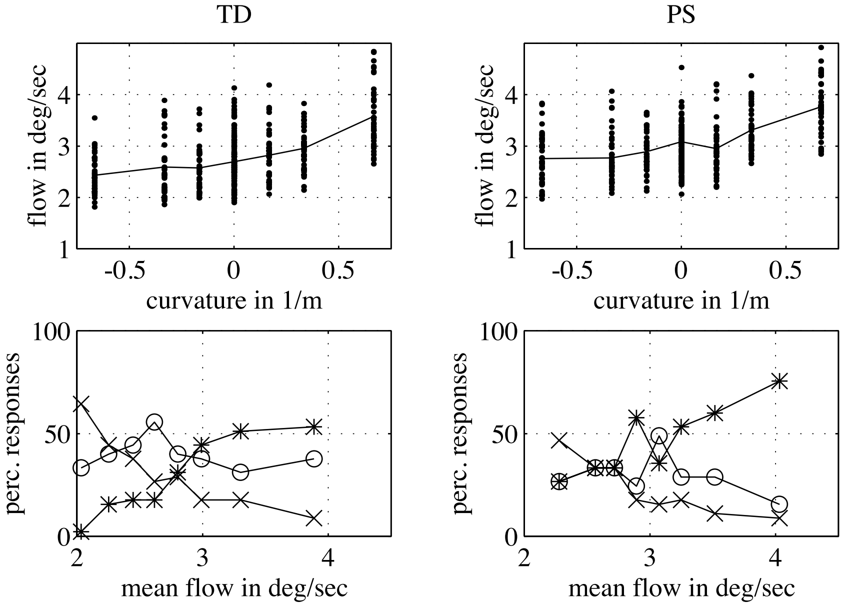


Figure 4–6 Results of calculation to show that subjects do not use mean retinal flow as an artifactual cue for curvature detection. X: percentage of concave responses, O: percentage of planar responses, *: percentage of convex responses. Left column: subject TD, right column: subject PS. Upper panels: mean retinal flow vs curvature for every trial. The solid line indicates the average mean flow for each curvature. Lower panels: response curve if the subject would use mean flow for detection.

4.4 Discussion

I have conducted experiments in which I tested the performance of detection of curvature of 3D objects with three different movement conditions and two different sizes of the field of view. All movement conditions resulted in the same information on the optic array of the observer. The difference between the conditions was the amount of extra-retinal information available. In the subject-movement condition the subject moved his head left/right and was shown a simulated segment of a sphere or plane. In the object-translation condition the head of the subject was stationary but the stimulus translated left/right. In the object-rotation condition the head of the subject was also stationary but the stimulus rotated in depth.

The first conclusion that emerges from our experiments is that the accuracy in discriminating between planar and non-planar surfaces is not increased by proprioceptive ego-motion information. This is borne out by the fact that the performance in the object-rotation condition, where the subject cannot use any proprioceptive ego-motion information, is the same (large-field stimulation) or better (small-field stimulation) than in the subject-movement condition, where the subject can use all available ego-motion information. For small-field stimulation this result was already reported before by Cornilleau-Pérès and Droulez [14]. Our results suggest that proprioceptive ego-motion information is not used directly in the perception of curvature of 3D shapes.

Second, our results suggest that proprioceptive ego-motion information is used in the stabilisation of the retinal image of the object, thereby enhancing the sensitivity to detection of curvature. Stabilisation of the retinal image is important because it affects the sensitivity to detection of differential motion [51] and this is thought to be used in the detection of 3D shape [12, 39]. Both our findings for small-field stimulation and for large-field stimulation corroborate the hypothesis that proprioceptive ego-motion information is used for the stabilisation of the retinal image. For small-field stimulation we found performance in the detection of absolute curvature to be highest in the object-rotation condition, we found intermediate performance in the subject-movement condition and the worst performance in the object-translation condition. It is known [24] that stabilisation of gaze is better with a stationary observer than with an observer rotating his head. With a small target of 0.2 deg Ferman et. al [24] found a mean retinal slip velocity of 0.4 deg/s, whereas they found a mean retinal slip velocity of 0.7 deg/s for subjects rotating their heads at 0.33 Hz. In another study, van den Berg and Collewijn [80, p 1216, table 1, discarding subject C.E.] found a retinal velocity of 1.25 deg/s for a small translating target of 0.15 deg. The results of these studies cannot be directly applied to our experiment because the stimulus characteristics are somewhat different from ours. In particular, we expect a larger retinal slip in the object-translation condition because the translation of the subject is not purely horizontal, but contains some vertical movement (this depends on how the subject moved his heads in the subject-movement condition). But we expect the *ordering* in the amount of retinal slip to hold in our experiment also. Thus, we expect that retinal slip is smallest in the object-rotation condition, intermediate in the subject-movement condition and largest in the object-translation condition. For large-field stimulation we found performance in all conditions to be roughly equal. I could not find any studies comparing retinal slip for a large field of view in our different movement conditions. For stimulus characteristics comparable to my large-field stimulation, van den Berg and Collewijn found a retinal slip velocity of 0.55 deg/s for a target moving together with a large field. This shows that retinal slip is probably quite small in our object-translation condition, compared to the retinal slip for small-field stimulation. The retinal slip is probably not much smaller in the other two movement conditions, which is consistent with our finding that performance is equal in the different movement conditions. A final point which might influence performance is the fact that subjects always had a part of the stimulus in central vision in large-field stimulation, whereas they had more trouble in fixating the stimulus for the object-translation condition with small-field stimulation. In this last case the stimulus was not always in central vision.

Further, I have evidence that ego-motion information is helpful in disambiguating the sign of curvature, at least when perspective information is not strong enough to disambiguate the curvature of the surfaces. Rogers and Rogers [64] found that both ego-motion information and perspective information can disambiguate the curvature sign. They used stimuli that were far above the threshold of curvature detection with a depth extent of 3 cm and a field of view of 17 deg. For large-field stimulation all movement conditions in our study led to unambiguous percepts of the sign of curvature. The depth range present in our stimuli, for a curvature of 0.4 m^{-1}, is 6.5 cm. For small-field stimuli, the subject-movement condition led to unambiguous percepts of the sign of curvature and the object-rotation condition to many reversals of the sign of curvature. Thus full proprioceptive ego-motion information was enough to disambiguate the sign of curvature whereas the small amount of perspective information (0.52 cm for a curvature of 4 m^{-1}) in the object-rotation condition was not enough. Whether the information from tracking eye movements alone is sufficient to disambiguate the sign of curvature, is not clear from our results but the control session (large-field set-up with small-field stimuli) seems suggestive that they are.

As far as object-rotation is concerned, the finding that many ambiguities are found for a small field of view but not for a large field of view needs to be explained. Assessing the respective roles of *perspective information* (the difference between images from concave and convex surfaces increases with the view angle in polar projection) and of the *field size* is difficult because these two variables covary in our experiment. However, two findings suggest it is mainly field size that influences the ambiguity. First, apparent deformations of the stimulus were frequently seen with small-field stimulation but never with large-field stimulation. These deformations indicate that the amount of perspective information with small-field stimulation was large enough to make the images of concave and convex surfaces visibly different. Second, the amount of perspective information increases with surface curvature and I found that the number of errors in detecting the sign of curvature sometimes increased with increasing curvature. Hence, I speculate that rigidity of the object plays a more important role for peripheral vision, where most motion is caused by ego-motion and thus objects are moving rigidly, than for central vision where we often observe nonrigid motion in natural situations.

The performance in curvature detection in the small-field object-rotation condition can be compared to the results of Norman and Lappin [52]. They found a percentage of correct responses of 96.5% for the discrimination between a plane and a concave sphere with a curvature of 4 m^{-1}. Their field of view was smaller than ours (2 deg vs 8 deg) whereas their amplitudes (35 deg vs 10 deg) and movement frequencies (1.2 Hz vs 0.33 Hz) were higher than ours. Further they gave the subjects feedback about their performance. Despite these differences I find a comparable performance in detection of absolute curvature of 93% (average over our 3 subjects).

Contrary to Rogers and Graham [61] I find that curvature detection is best in the object-rotation condition for small-field stimulation. Besides differences in stimulus geometry and field of view there are two important differences. First, the subject could see the layout of the experimental room in the set-up of Rogers and Graham, and this visual information about ego-motion could be responsible for the largest depth being perceived during self-motion. Second, the task of the subjects of Rogers and Graham was to estimate the amount of depth of the corrugations, whereas our subjects estimated the curvature.

Chapter 5

Extraction of 3D Shape from Optic Flow

abstract

I show how a scale invariant measure of three dimensional shape can be derived from the velocity field generated by a rigid curved surface patch under perspective projection. I use invariance under rotation of the image plane (the Lie group SO(2)) to decompose the second order velocity field in differential invariants. From a combination of these invariants I construct an approximation to the absolute value of Koenderink's shape index. I will show that the effect of these approximations on the shape index is small, especially under parallel projection. Furthermore, I provide an explanation for the psychophysical finding that umbilic shapes are more readily detected than parabolic or hyperbolic shapes. From the invariants I can also derive approximations of the principal directions, the curvedness, the slant and the tilt.

This chapter is an expanded version of the paper "Extraction of 3D Shape from Optic Flow: a Geometric Approach" by Tjeerd Dijkstra, Peter Snoeren and Stan Gielen, Journal of the Optical Society of America A 11, 2184-2196 (1994). Part of this chapter is from a contribution with the same authors and title in the IEEE Conference on Computer Vision and Pattern Recognition (pp. 135-140). Los Alamitos CA: IEEE Computer Society Press

5.1 Introduction

Almost two decades have passed since Koenderink and van Doorn's original article on the relation between the geometry of a surface and the induced optic flow was published [37]. In this article they relate the geometrical properties of rigid objects to invariants of the velocity field generated by the moving object. They were quite successful for the first order object properties (e.g. slant), but of the second order properties (e.g. curvature) they could only calculate the sign of the gaussian curvature. The research initiated by Koenderink and van Doorn has taken a more algebraic turn in the work of e.g. [44] and [84]. In a recent paper Koenderink and van Doorn [39] focused on the second order structure. They were able to fully describe the structure but only in terms of a somewhat unusual quantity viz. the projected indicatrix of Dupin. In this paper I will go back to the geometric approach, based on differential invariants, as in [37] and extend it to second order properties.

Central in my approach to the extraction of structure from motion is the use of *geometry*. Geometrical properties are those properties that are invariant when the coordinate system is changed. More specifically, we will study the invariance of the velocity field under the group of rotations of the plane (the Lie group SO(2)). Because the velocity field is a vector field its invariants are not necessarily scalars, but usually vectorlike quantities. Paradoxically, this means that the invariants of the velocity field can change when the coordinate system is rotated. A definition of invariance of vector fields is beyond the scope of this chapter but I will indicate how the invariants of a vector field are calculated in section 5.3. For a definition of invariance of non-scalar fields the reader is referred to [34] and the references therein. From the point of view of machine vision rotational invariance allows one to do the calculations independent of the orientation of the camera. Moreover, one can view rotational invariance as a convenient computational tool: by expressing the relations between unknowns and observables in invariants the equations become simpler. Group theory is often used for this purpose in mathematics and image processing. From the point of view of human vision the use of the group of rotations of the plane (corresponding to torsional eye movements or torsional movements of the object) is not so easily defended. It would be more natural to study the invariants of the velocity field under the full rotation group in three dimensions SO(3). Since this is computationally much more complex and since SO(2) is a subgroup of SO(3), one can view the current approach as a first step. Group theory has been introduced into psychology by Hoffman [31] but it has attracted few adherents [20]. Recently, it has been applied successfully in neurophysiology [26]. In the rest of this chapter I will use invariance to mean rotational invariance. The only exception is the next section when I introduce the shape index [40], which is a scale invariant descriptor of shape.

Smooth rigid objects have many differential geometric properties, which can be classified according to order. The zeroth order property is the distance of the object. First order properties can be described by the slant and the tilt. Note that distance, slant and tilt of an object will in general change when it moves relative to the observer. The second order properties can be described by e.g. the two principal curvatures and the direction of maximal normal curvature. My focus will be on these properties. Note that the principal curvatures are the lowest order properties which do not change when the object moves relative to the observer i.e. they are inherent to the object. The third order properties can be described by the gradients of the two principal curvatures. I will not deal with these although the method presented here could conceivably be used to deduce the third order properties from invariants of the velocity field too.

An important point to keep in mind is that metric information cannot be obtained from the velocity field, because the velocity field is derived from a projection. This means that one cannot obtain the complete 3D translation velocity of the observer relative to the object. Neither can one obtain the distance to the object nor its curvatures. Of course one can construct combinations of these properties that do not depend on metric information e.g. the ratio of the principal curvatures or the ratio of the 3D velocity and the distance. Since the principal curvatures are the lowest order properties that are inherent, the ratio of the principal curvatures (or any function thereof) is the only inherent property of the object that can be derived from the velocity field (up to second order).

My calculations are based on a number of assumptions to make the problem manageable. I assume the existence of a dense smooth vector field on a planar camera, which is generated by a smooth rigid surface patch by perspective projection. The assumption of the existence of a dense vector field is probably not necessary as algorithms have been proposed to extract first order differential invariants directly from the spatio-temporal luminance pattern [86, 19]. These algorithms could be extended to include the second order as well. The assumption of a planar camera (equipped with the natural metric) is accurate in machine vision but probably not for human vision when wide-field stimuli are employed, as in the previous chapter. The assumption of a rigid smooth surface patch is realistic for rigid surfaces away from their contour. Finally, perspective projection is the correct way of proceeding. The approximations I will introduce below are not necessary under parallel projection. The use of perspective projection will allow me to give an indication for the error I make in these approximations.

Finally, I want to point out the usefulness of the current approach to the problem of binocular vision. One can view the disparity field as the equivalent of the velocity field and use the same relations introduced below. Two of the differences between stereo and motion are the possible use of nonvisual information in stereo (e.g. vergence angle of the eyes) and the fact that the eyes are usually horizontally aligned making my principle of rotational invariance less relevant for stereo. Nevertheless, I am able to offer an explanation for some results of shape-from-stereo experiments.

5.2 Description of shape measures

In this section I will shortly introduce the shape measures to be used in the sequel: the shape index and the curvedness [40]. Further I will introduce some notation to describe a surface patch up to second order.

Many descriptors of shape for the purpose of vision and image processing have been used, but the descriptors that have been used most often are the principal curvatures κ_{max} and κ_{min} and the gaussian and mean curvature, K and H [5]. The principal curvatures are the maximum and minimum of the normal curvature κ_n. The normal curvature is obtained as the curvature of the curve one gets when one cuts the surface with a plane through the normal of the surface. The relation between the normal curvature in direction $\mathbf{e}_\alpha$ and the principal curvatures is given by Euler's formula [54]:

$$\kappa_n(\alpha) = \kappa_{max} \cos^2(\alpha - \alpha_0) + \kappa_{min} \sin^2(\alpha - \alpha_0), \tag{5.1}$$

with $\alpha_0 \in [0, \pi)$ denoting the direction of maximal curvature. The direction of minimal curvature is always orthogonal to the direction of maximal curvature.

An important thing to note about these descriptors for curvature is that they are scale dependent e.g. making a sphere twice as large will change the values of all these descriptors. This does not comply with our intuition that spheres of different radius have the same shape. Using this intuition and some other desirable properties, Koenderink and van Doorn proposed the following descriptors [40]:

$$\begin{aligned} S &= (2/\pi) \arctan(\frac{\kappa_{max} + \kappa_{min}}{\kappa_{max} - \kappa_{min}}), && (5.2) \\ C &= \sqrt{\frac{\kappa_{max}^2 + \kappa_{min}^2}{2}}, && (5.3) \end{aligned}$$

with the *shape index* S carrying all scale independent information about shape, and the *curvedness* C carrying all scale dependent information. Algebraically one can view the shape index and the curvedness as scaled polar coordinates in the κ_{max}, κ_{min} half-plane (see fig. 5–1). I take $\kappa_{max} \geq \kappa_{min}$ because interchanging κ_{max} and κ_{min} has the same effect as interchanging x and y (see eq. 5.1 with $\alpha_0 = 0$) and thus only changes the orientation of the surface, not its shape.

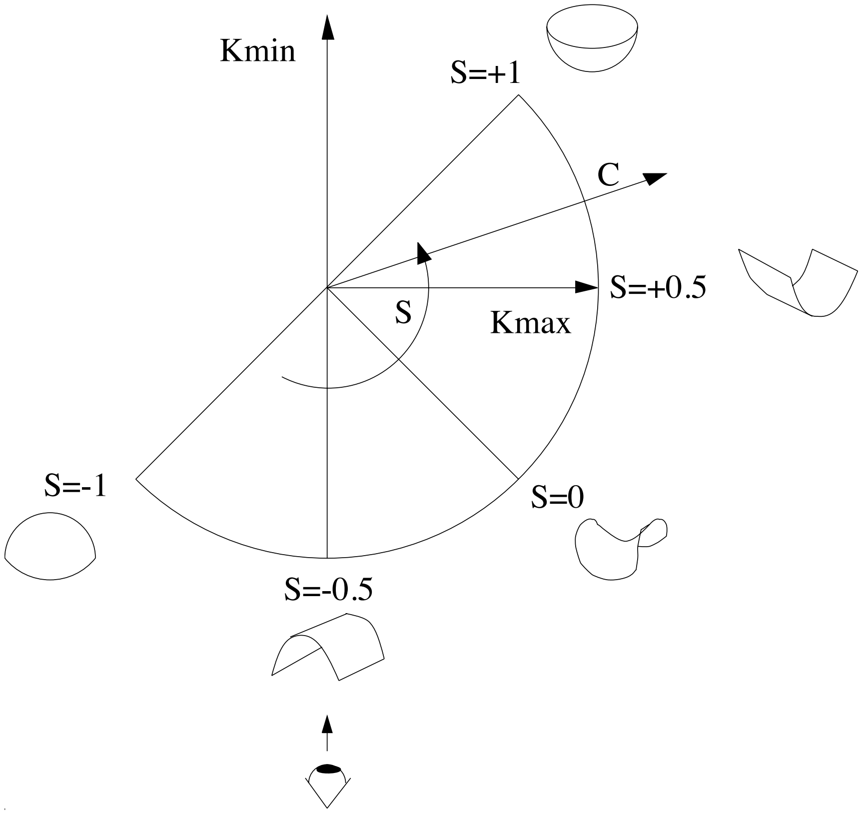


Figure 5–1 Objects in $\kappa_{max}, \kappa_{min}$ space. By a change of coordinate system in this space, objects can also be characterised by an angular coordinate, the shape index S, and a radial coordinate, the curvednes C. The shapes drawn to illustrate various values of S are to be viewed from below.

I parametrise a smooth surface patch using the range function $Z(X, Y)$ with $(X, Y, Z) \in \mathbf{R}^3$. I write the surface patch up to second order in a Taylor series as follows:

$$Z(X,Y) = Z_0 + Z_X\,X + Z_Y\,Y + 1/2\,Z_{XX}\,X^2 + Z_{XY}\,XY + 1/2\,Z_{YY}\,Y^2, \quad (5.4)$$

where Z_0 is the distance to the patch, the pair $(Z_X, Z_Y)^T \equiv \nabla Z$ is the range gradient and the remaining three parameters Z_{XX}, Z_{XY} and Z_{YY} denote the second order derivatives of the range function. The range gradient is related to the attitude of the surface patch by $\nabla Z = \tan\sigma(\cos\tau,\ \sin\tau)^T$ with $\sigma \in [0, \pi/2]$ denoting the slant and $\tau \in [0, 2\pi)$ denoting the tilt. The slant equals the angle between the normal of the surface and the line of sight (I take the direction of the normal always towards the viewer). For a fronto-parallel plane we have $\sigma = 0$, for a plane viewed edge on we have $\sigma = \pi/2$. The tilt equals the angle between the normal of the surface when projected on the image plane, and the X-axis.

The relation between normal curvature and the second order derivatives of the range function depends also on the first order derivatives. Taking a fiducial direction $\mathbf{e}_\alpha$ in the tangent plane (which has angle α with the X-axis) one can show [54, p 221]:

$$\kappa_n(\alpha) = \frac{1}{\sqrt{1 + Z_X^2 + Z_Y^2}}\ \frac{Z_{XX}\cos^2\alpha + 2Z_{XY}\cos\alpha\sin\alpha + Z_{YY}\sin^2\alpha}{(1 + Z_X^2)\cos^2\alpha + 2Z_X Z_Y\cos\alpha\sin\alpha + (1 + Z_Y^2)\sin^2\alpha}.$$

Expressing the first order derivatives in terms of slant and tilt we find:

$$\kappa_n(\alpha)\,\frac{1 + \tan^2\sigma\cos^2(\alpha - \tau)}{\cos\sigma} = Z_{XX}\cos^2\alpha + 2Z_{XY}\cos\alpha\sin\alpha + Z_{YY}\sin^2\alpha.$$

This expression can be simplified by introducing the angles ζ and μ defined by [39]:

$$\begin{aligned}\cos\zeta &\equiv \cos\sigma\sqrt{1 + \tan^2\sigma\cos^2(\alpha - \tau)},\\ \cos\mu &\equiv 1/\sqrt{1 + \tan^2\sigma\cos^2(\alpha - \tau)}.\end{aligned}$$

These angles describe the geometry of the normal of the surface patch relative to the plane (denoted by Γ) determined by the fiducial direction $\mathbf{e}_\alpha$ and the line of sight $\mathbf{e}_z$ (see fig. 5–2). The angle between the normal and the plane Γ is ζ. Projecting the normal onto Γ we obtain μ as the angle between this projection and $\mathbf{e}_z$. Using these angles, we find:

$$\lambda \equiv \kappa_n/(\cos\zeta\cos^3\mu) = Z_{XX}\cos^2\alpha + 2Z_{XY}\cos\alpha\sin\alpha + Z_{YY}\sin^2\alpha, \tag{5.5}$$

where I have introduced λ as a new curvature measure. In contrast to κ_n, which only depends on curvature, λ also depends on the attitude of the tangent plane and on the fiducial direction. Later I will show that it is impossible to obtain κ_n directly from the velocity field, but that λ can be obtained directly. Note that

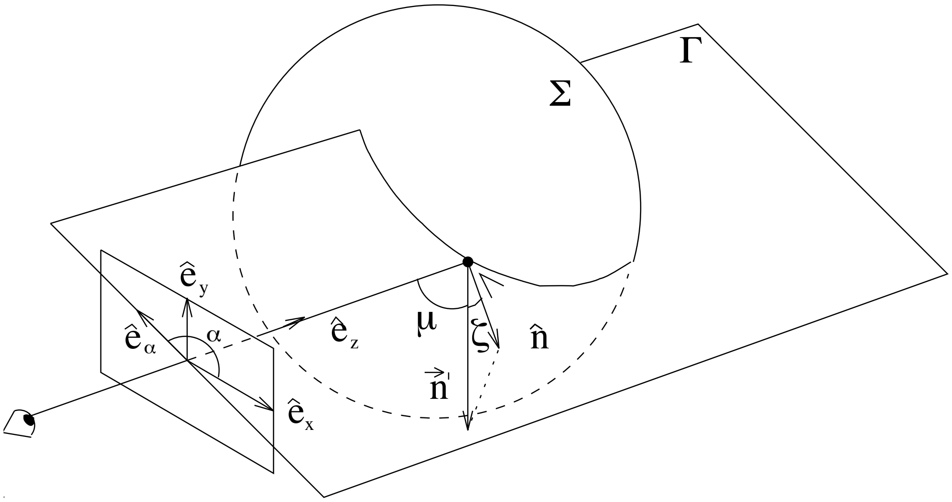


Figure 5–2 The geometry of the normal $\vec{\mathbf{n}}$ to the surface patch Σ relative to the fiducial plane Γ. Γ is the plane through the fiducial direction $\mathbf{e}_\alpha$ and the direction of looking $\mathbf{e_Z}$. ζ denotes the angle between $\vec{\mathbf{n}}$ and Γ. μ denotes the angle between the projection ($\vec{\mathbf{n}'}$) of $\vec{\mathbf{n}}$ on Γ and $\mathbf{e_Z}$.

the last equation has the same structure as Euler's formula (eq. 5.1). One can easily show that for:

$$\tan(2\,\alpha_0) = 2Z_{XY}/(Z_{XX} - Z_{YY}), \tag{5.6}$$

one obtains the maximal and minimal λ. α_0 gives the direction of maximal λ. The maximal and minimal λ are given by:

$$\lambda_{(max,\,min)} = 1/2\,(Z_{XX}+Z_{YY})\pm 1/2\sqrt{Z_{XX}^2 + 4Z_{XY}^2 + Z_{YY}^2 - 2Z_{XX}Z_{YY}}. \tag{5.7}$$

When the tangent plane is fronto-parallel, we have $\sigma = 0$. In that case λ_{max} and λ_{min} in the previous equation become equal to κ_{max} and κ_{min}. Also, when the slant is not too large, $\cos\zeta$ and $\cos\mu$ will be close to 1 and thus λ_{max} and λ_{min} will not deviate very much from κ_{max} and κ_{min} (see section 5.5).

5.3 Invariant decomposition of the second order velocity field

In this section I will decompose the second order velocity field in differential invariants. The complete decomposition of the zeroth and first order velocity field was already done by [37] and leads to four differential invariants: translation of order zero and divergence, rotation and deformation of order one. For reference, I give the expressions of the first order differential invariants in terms of derivatives of the velocity field. I denote the divergence by $\nabla \cdot \mathbf{v}$, the rotation by $\nabla \times \mathbf{v}$ and the deformation by $\nabla \circ \mathbf{v}$. We have [37]:

$$\begin{aligned}
\nabla \cdot \mathbf{v} &= v_x^x + v_y^y, && (5.8)\\
\nabla \times \mathbf{v} &= -v_y^x + v_x^y, && (5.9)\\
\nabla \circ \mathbf{v} &= \begin{pmatrix} v_x^x - v_y^y \\ v_y^x + v_x^y \end{pmatrix}, && (5.10)
\end{aligned}$$

with e.g. v_y^x denoting the first order derivative in the y direction of the x component of the velocity field. It will be an easy exercise to show with the methods introduced below that these are actually invariant. In fig. 5–3 I have plotted some examples of the vector field generated by these invariants. Discussing these will give me a chance to introduce the concept of the *weight* of an invariant. The weight captures the symmetry properties of the invariant under rotation: 2π divided by the weight is the smallest angle over which one has to rotate the invariant so that it becomes equal to itself again (except for weight zero). The translation has weight 1, meaning that a rotation over 2π maps the vectorfield onto itself. The divergence and rotation have weight zero, meaning that every rotation maps the vectorfield onto itself. Thus they are scalars. The deformation has weight 2, meaning that a rotation over π maps the vectorfield onto itself (see fig. 5–3).

The importance of invariance under a rotation stems from the fact that we do not have a preferred direction in the image plane. This is a natural approach because the shape measures listed in the previous section were also constructed so as to be independent of the choice of coordinate system. A nice bonus of calculating invariants is that invariants have weights and these correspond directly to geometrical properties of the observables. Only linear combinations of invariants of the same weight are invariants. It is important to realise that invariance does not mean that an observable does not change its value when the coordinate system is rotated. One can already observe this from the zeroth order invariant: rotating the coordinate system over θ rotates the translation over $-\theta$. The observables that do not change value when the coordinate system is rotated are called *scalar* invariants in this context. Examples are the divergence and the rotation (they have weight zero). The construction of invariants proceeds as follows: we take a set of observables measured relative to some coordinate system.

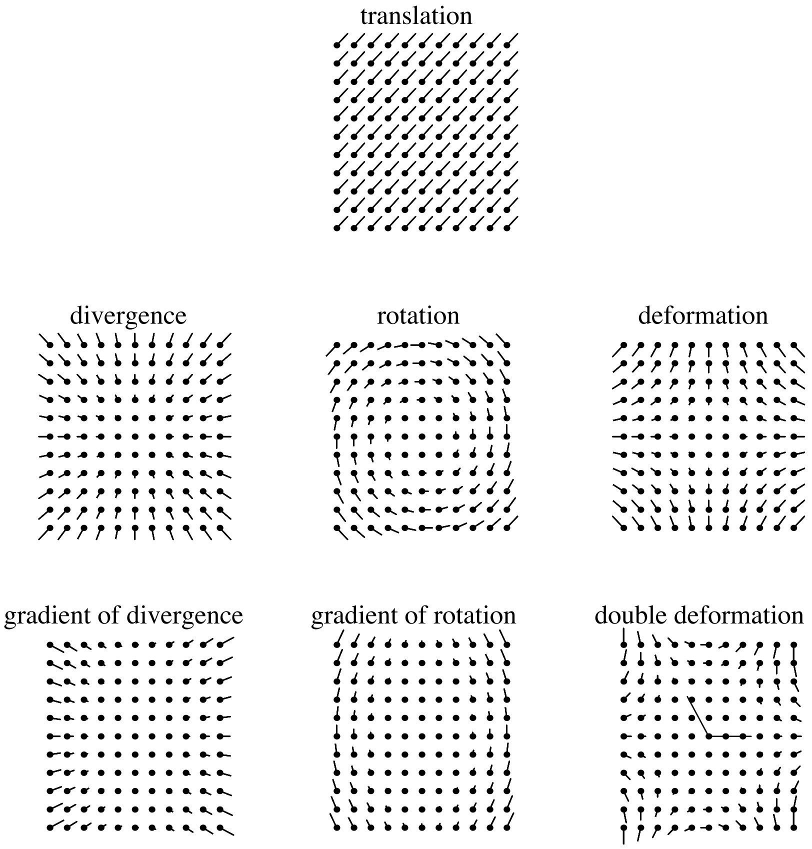


Figure 5–3 Flow fields generated by pure differential invariants. Top row: the zeroth order invariant, translation in direction (1, 1). Middle row: the first order invariants, divergence, rotation and the deformation in direction (1, 0). Bottom row: the second order invariants, all in direction (1, 0). In the flow field of the double deformation I have drawn two lines of equal length, with an angle $2\pi/3$ between them.

We then rotate the coordinate system over an infinitesimal angle[1]. The observables relative to the new coordinate system will be a linear transformation of the observables relative to the old coordinate system. Denote this transformation by T. We now call our set of observables invariant when T is a (complex) diagonal matrix. My notion of invariance is perhaps best illustrated by an example of a quantity that is not invariant (and thus variant). Examples of variants are the

---

[1] For the special case of SO(2) this angle does not have to be infinitesimal.

spatial derivatives in a cartesian coordinate system. Because I will come back to it later, I take the spin variation in the direction of the x axis $SV(0) = v^y_{xx}$ as an example of a variant. Rotating v^y_{xx} over an arbitrary angle θ we find:

$$\begin{aligned} SV(\theta) &= \cos^3\theta(2v^x_{xy} + v^y_{xx} - v^y_{yy}) + \sin^3\theta(-2v^y_{xy} + v^x_{xx} - v^x_{yy}) + \\ &\quad \cos\theta(-2v^x_{xy} + v^y_{yy}) + \sin\theta(2v^y_{xy} + v^x_{xx}). \end{aligned} \tag{5.11}$$

Clearly, when we rotate the coordinate system over an arbitrary angle, v^y_{xx} transforms in a very complicated way. The transformed v^y_{xx} does not only depend on the original v^y_{xx}, but also on all other second order spatial derivatives. Thus the transformation T is not a complex diagonal matrix. This is different for the invariants: a transformed invariant depends only on the original invariant, not on the others.

The image plane is two dimensional and has the nice property that rotations commute i.e. the result of two rotations is independent of the order in which they are performed. Because of this one can show that all invariants are necessarily complex numbers [34], with the exception of the invariants of weight 0, which form a pair of real numbers. Because the group of rotations in the plane (SO(2)) is also easily denoted by a complex number of unit length I will construct the invariants using a complex number notation. I denote the velocity field by $v(z, \bar{z})$ with $z = x + iy$ and the complex conjugate $\bar{z} = x - iy$. v is now a mapping from the complex plane to the complex plane. I write the second order development of the velocity field as:

$$v(z, \bar{z}) = 1/2\, v_{zz}\, z^2 + v_{z\bar{z}}\, z\bar{z} + 1/2\, v_{\bar{z}\bar{z}}\, \bar{z}^2,$$

with v_{zz} the velocity field differentiated twice with respect to z and with similar notation for the other derivatives. Now v_{zz}, $v_{z\bar{z}}$ and $v_{\bar{z}\bar{z}}$ are observables of our velocity field which we measure relative to some coordinate system. To make these observables independent of our choice of coordinate system we rotate it with an arbitrary angle and see what happens. As we will see below, it will turn out that v_{zz}, $v_{z\bar{z}}$ and $v_{\bar{z}\bar{z}}$ are already invariant! This actually means that I have been cheating by starting with a very favourable representation from the outset. I denote an arbitrary rotation by $U(\theta) = \exp(in\theta)$ with $U(\theta) \in SO(2)$, $\theta \in S^1$ and an arbitrary integer n, as the weight. Rotating $v(z, \bar{z})$ over θ we find:

$$\begin{aligned} U(\theta)\, v(U^{-1}(\theta)z,\, U^{-1}(\theta)\bar{z}) &= \exp(in\theta)\{1/2\, v_{zz}\, (\exp(-in\theta)\, z)^2 + \\ &\qquad v_{z\bar{z}}\, (\exp(-in\theta)z)\, (\exp(in\theta)\bar{z}) + \\ &\qquad 1/2\, v_{\bar{z}\bar{z}}\, (\exp(in\theta)\bar{z})^2\}. \end{aligned}$$

This results in:

$$v(z,\bar{z}) = 1/2\ \exp(-in\theta)\, v_{zz}\, z^2 + \exp(in\theta)\, v_{z\bar{z}}\, z\bar{z} + 1/2\ \exp(3in\theta)\, v_{\bar{z}\bar{z}}\, \bar{z}^2.$$

Thus we have three differential invariants of second order: v_{zz} is an invariant of weight -1 , $v_{z\bar{z}}$ is an invariant of weight 1 and $v_{\bar{z}\bar{z}}$ is an invariant of weight 3. The sign of the weight has to do with symmetry under reflection in the x axis and need not concern us here. One can change the sign of an invariant by complex conjugation. The differential invariants are not unique: linear combinations of invariants with the same weight are also invariant.

By rewriting the differential invariants in real coordinates we can get a better picture of them. As before I will use upper indices to denote the components of the velocity field v, e.g. v^x denotes the x component. I will use lower indices to denote spatial derivatives, e.g. v^x_{yy} denotes the x component twice differentiated in the y direction. The reader can easily verify the following relations:

$$2(\bar{v}_{zz} + v_{z\bar{z}}) = \begin{pmatrix} v^x_{xx} + v^y_{xy} \\ v^x_{xy} + v^y_{yy} \end{pmatrix} = \nabla(\nabla \cdot \mathbf{v}), \tag{5.12}$$

$$2i(\bar{v}_{zz} - v_{z\bar{z}}) = \begin{pmatrix} -v^x_{xy} + v^y_{xx} \\ -v^x_{yy} + v^y_{xy} \end{pmatrix} = \nabla(\nabla \times \mathbf{v}), \tag{5.13}$$

$$4v_{\bar{z}\bar{z}} = \begin{pmatrix} v^x_{xx} - v^x_{yy} - 2v^y_{xy} \\ 2v^x_{xy} + v^y_{xx} - v^y_{yy} \end{pmatrix} = \nabla \circ \nabla \circ \mathbf{v}. \tag{5.14}$$

Here I have introduced the three second order differential invariants of the velocity field. The gradient of the divergence $\nabla(\nabla \cdot \mathbf{v})$ and the gradient of the rotation $\nabla(\nabla \times \mathbf{v})$ were already introduced in [37]. The double deformation $\nabla \circ \nabla \circ \mathbf{v}$ is the remaining differential invariant.

In fig. 5–3 (bottom row left panel) I have plotted the vector field of a pure gradient of the divergence. In the next section I will show this invariant to depend both on the second and lower order terms of the geometry of the surface. Although the dependence on the curvature is the most important, we can use the first order term to get an intuitive idea of the vector field generated by this invariant (see eq. 5.16 below): imagine a plane rotating around an axis in the plane. Then part of the plane comes towards the observer (leading to positive divergence) and part goes away from the observer (leading to negative divergence). For the flow pattern of fig. 5–3 (bottom row, left panel) the rotation would be around a vertical axis. Note that the vector field thus generated does not lead to a pure gradient of divergence but also to a nonzero gradient of rotation (see below). Thus the divergence changes with position and we have a gradient.

The gradient of the divergence has weight 1, thus it is a normal vector: only rotation over 2π transforms it into itself.

The vector field of a pure gradient of the rotation is plotted in fig. 5–3 (bottom row middle panel). Just as for the gradient of the divergence this invariant depends both on the second and lower order terms of the geometry of the surface, with the second order term being the important one. Again, we can get an idea of the vector field generated by this invariant by considering a rotating plane: in the part moving towards the observer the velocity field will not only be oriented away from the axis of rotation but will also be slightly curved inward. On the other side of the axis of rotation the velocity field will be slightly curved outward. Thus, the direction of rotation is different at opposite sides of the point of fixation, leading to a gradient in the rotation. This is harder to imagine than for the gradient of divergence because the non-curvature dependent terms are three times as small for the gradient of rotation (see eq. 5.17 below). The gradient of the rotation has weight 1, thus it is a normal vector.

In fig. 5–3 (bottom row right panel) I have plotted the vector field of a pure double deformation. In the next section I will show that this invariant only depends on the shape of the object: it is zero when the object is not curved. The double deformation has weight 3, and a rotation over $2\pi/3$ transforms it into itself. As this is difficult to see directly I have drawn two additional lines under an angle of $2\pi/3$ in the plot of the double deformation.

The tools I have introduced above are very powerful: it is easy to see that the invariants of third order have weights -2, 0, 2 and 4. One should be able to identify the invariants of weight -2 and 2 with the deformation of the gradients of rotation and divergence respectively. The invariant of weight 4 could be identified with the triple deformation. Finally, the pair of scalar invariants of weight zero could be identified with the divergence of the gradient of the divergence or rotation.

5.4 Extraction of shape descriptors from optic flow

In this section I will calculate the velocity field generated by a moving rigid surface patch under perspective projection, and calculate the invariants of the velocity field. From these invariants we will obtain the shape index and some other properties related to shape.

The velocity field and the second order invariants

The expression of the velocity field on a planar camera is derived in many studies e.g. [44, 84]. I consider a moving observer viewing a stationary object. I locate the origin of the coordinate system at the vertex of perspective projection and the positive Z-axis along the line of sight. Representing the patch by $Z(X,Y)$ and subjecting the observer to a translation $\mathbf{V} = (V^X, V^Y, V^Z)^T$ and rotation $\mathbf{\Omega} = (\Omega^X, \Omega^Y, \Omega^Z)^T$ we find for the velocity field on a planar camera at unit focal distance from the point of projection:

$$\begin{pmatrix} v^x \\ v^y \end{pmatrix} = \begin{pmatrix} V^Z/Z\ x - V^X/Z + \Omega^X\ xy - \Omega^Y\ (1+x^2) + \Omega^Z\ y \\ V^Z/Z\ y - V^Y/Z + \Omega^X\ (1+y^2) - \Omega^Y\ xy - \Omega^Z\ x \end{pmatrix}, \tag{5.15}$$

with $x = X/Z, y = Y/Z$ cartesian coordinates on the camera. Here I have introduced the 3D translation $\mathbf{V}$ and the 3D rotation $\mathbf{\Omega}$ as independent to keep the results general. In the simulation section below I will assume fixation. Now, using the definition of the surface patch in eq. 5.4 we approximate $1/Z$ in camera coordinates [71]:

$$Z^{-1}(x,y) = (1 - Z_X\,x - Z_Y\,y)/Z_0 - 1/2\,Z_{XX}\,x^2 - Z_{XY}\,xy - 1/2\,Z_{YY}\,y^2 + O^3(x,y).$$

Substituting this relation in eq. 5.15 and then using eqs 5.12 to 5.14, we get for the second order differential invariants:

$$\begin{aligned}
\nabla(\nabla\cdot\mathbf{v}) &= \begin{pmatrix} Z_{XX} & Z_{XY} \\ Z_{XY} & Z_{YY} \end{pmatrix} \mathbf{V}_{\|} + 3J\,\mathbf{\Omega}_{\|} - 3V_{\perp}/Z_0\ \nabla Z, && (5.16)\\
\nabla(\nabla\times\mathbf{v}) &= -\begin{pmatrix} Z_{XX} & Z_{XY} \\ Z_{XY} & Z_{YY} \end{pmatrix} J\,\mathbf{V}_{\|} - \mathbf{\Omega}_{\|} - V_{\perp}/Z_0\ J\,\nabla Z, && (5.17)\\
\nabla\circ\nabla\circ\mathbf{v} &= \begin{pmatrix} Z_{XX} - Z_{YY} & -2Z_{XY} \\ 2Z_{XY} & Z_{XX} - Z_{YY} \end{pmatrix} \mathbf{V}_{\|}, && (5.18)
\end{aligned}$$

where I have used the notation $\mathbf{V}_{\|} \equiv (V^X, V^Y)^T$ for translation parallel to the camera and $V_{\perp}$ for translation orthogonal to the camera and identical notation for $\mathbf{\Omega}$. The matrix:

$$J = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix},$$

denotes a rotation over $\pi/2$. The first two expressions were already derived in [37] albeit in a spherical coordinate system. Note that the gradients of divergence and rotation depend both on curvature of the patch and on lower order terms and that the double deformation only depends on curvature.

The shape index

The double deformation is related in an interesting way to the curvature of the surface patch. To see this we will calculate the length of the double deformation and the angle between double deformation and $\mathbf{V}_{\|}$:

$$\begin{aligned} |\nabla\circ\nabla\circ \mathbf{v}| &= \sqrt{(Z_{XX}-Z_{YY})^2+4Z_{XY}^2}\,|\mathbf{V}_{\|}|, &(5.19)\\ \tan\angle(\mathbf{V}_{\|},\,\nabla\circ\nabla\circ \mathbf{v}) &= \frac{-2Z_{XY}}{Z_{XX}-Z_{YY}}. &(5.20)\end{aligned}$$

These relations are easily understood from eqs. 5.6 and 5.7. We have:

$$\begin{aligned} |\nabla\circ\nabla\circ \mathbf{v}| &= |\lambda_{max}-\lambda_{min}|\,|\mathbf{V}_{\|}|, &(5.21)\\ \angle(\mathbf{V}_{\|},\,\nabla\circ\nabla\circ \mathbf{v}) &= -2\alpha_0. &(5.22)\end{aligned}$$

Geometrically the operation of the double deformation on $\mathbf{V}_{\|}$ can be viewed as a rotation over $-2\alpha_0$ followed by a scaling with size $\lambda_{max}-\lambda_{min}$. For slants that are not too large, $\cos\zeta$ and $\cos\mu$ are close to 1 and it follows from eq. 5.5 that λ_{max} and λ_{min} are close to κ_{max} and κ_{min} (see simulation section). Looking back at the definition of the shape index (eq. 5.2) we see that we also need the sum of the two principal curvatures. This can be found from a linear combination of the gradients of divergence and rotation. We define [37]:

$$\boldsymbol{\beta}\equiv\nabla(\nabla\cdot\mathbf{v})+J\,\nabla(\nabla\times\mathbf{v})=\mathbf{v}_{xx}+\mathbf{v}_{yy}.$$

This is actually the differential invariant $v_{z\bar{z}}$ used before. We get:

$$\boldsymbol{\beta}=\begin{pmatrix} Z_{XX}+Z_{YY} & 0\\ 0 & Z_{XX}+Z_{YY}\end{pmatrix}\mathbf{V}_{\|}+2J\,\boldsymbol{\Omega}_{\|}-2V_{\perp}/Z_0\,\nabla Z. \qquad (5.23)$$

Neglecting the last two terms for the moment (which is exact under parallel projection, see below), we have

$$|\boldsymbol{\beta}|=|\lambda_{max}+\lambda_{min}|\,|\mathbf{V}_{\|}|.$$

Geometrically the operation of $\boldsymbol{\beta}$ on $\mathbf{V}_{\|}$ can be viewed as a scaling with size $\lambda_{max}+\lambda_{min}$. Now, we can extract the absolute value of the shape index by:

$$|S_e|=2/\pi\ \arctan(\frac{|\boldsymbol{\beta}|}{|\nabla\circ\nabla\circ\mathbf{v}|}). \qquad (5.24)$$

The sign of the shape index can be obtained from the relation between $\boldsymbol{\beta}$ and $\mathbf{V}_{\|}$. Because $\boldsymbol{\beta}$ is always parallel to $\mathbf{V}_{\|}$, I give S_e a positive sign when $\boldsymbol{\beta}$ and

$\mathbf{V}_\parallel$ point in the same direction and I give S_e a negative sign when the angle between $\boldsymbol{\beta}$ and $\mathbf{V}_\parallel$ is π. So, all we need is an estimate of the direction of $\mathbf{V}_\parallel$. Unfortunately, this cannot be obtained from second order optic flow when we neglect the non-curvature dependent terms. This can be seen geometrically by introducing the sister of $\boldsymbol{\beta}$ (called $\boldsymbol{\alpha}$) which we define by:

$$\boldsymbol{\alpha} \equiv M\left(\nabla(\nabla \cdot \mathbf{v}) - J\,\nabla(\nabla \times \mathbf{v})\right),$$

with M a reflection in the x axis ($\boldsymbol{\alpha}$ is the differential invariant v_{zz} used before). We neglect the non-curvature dependent terms, just as we did for $\boldsymbol{\beta}$, but it should be noted that they are twice as large as for $\boldsymbol{\beta}$. It is easy to show that the operation of $\boldsymbol{\alpha}$ on $\mathbf{V}_\parallel$ can be viewed as a rotation over $2\alpha_0$ followed by a scaling with size $\lambda_{max} - \lambda_{min}$. Thus one could also construct an estimate of the shape index from $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ at the expense of some extra approximations. More interestingly, we find the axial direction of $\mathbf{V}_\parallel$ as the bisector of $\boldsymbol{\alpha}$ and $\nabla \circ \nabla \circ \mathbf{v}$. Unfortunately, we do not know which direction to take on the bisector, leaving the sign of $\mathbf{V}_\parallel$ undetermined. For now, we will assume that we obtained the direction of $\mathbf{V}_\parallel$ by some other means, but I will come back to this below. We obtain a signed estimate of the shape index from:

$$S_e = \operatorname{sign}(\boldsymbol{\beta} \cdot \mathbf{V}_\parallel)\; 2/\pi\; \arctan\left(\frac{|\boldsymbol{\beta}|}{|\nabla \circ \nabla \circ \mathbf{v}|}\right). \tag{5.25}$$

It turns out to be possible to derive the shape index *directly* from the velocity field, without having to calculate any other parameter, not even the absolute value of the 3D translation velocity. It is important to realise that S_e is only defined for curved surfaces and does not signal the *presence* of any significant curvature. Thus, the shape index is not relevant for tasks where curved surfaces have to be discriminated from planar ones, like in the previous chapter.

The direction of maximal curvature and the curvedness

In the derivation above we have also found another shape characteristic: the direction of maximal curvature α_0 (see eq. 5.6). Because $\boldsymbol{\beta}$ and $\mathbf{V}_\parallel$ are parallel, we obtain an estimate of α_0 by:

$$2\alpha_e = \angle(\nabla \circ \nabla \circ \mathbf{v},\, \operatorname{sign}(\boldsymbol{\beta} \cdot \mathbf{V}_\parallel)\, \boldsymbol{\beta}).$$

Thus, we can also find the direction of maximal curvature directly from the velocity field. Note that when we do not know the direction of $\mathbf{V}_\parallel$ we can only obtain the orientation of the principal curvatures, but we do not know which orientation is the direction of maximal curvature and which of minimal curvature. Below I will show that this ambiguity is connected with the sign ambiguity of the shape index.

Furthermore, it is easy to derive the following estimate for the velocity scaled curvedness, denoted by Γ_e:

$$\Gamma_e \equiv C_e\, |\mathbf{V}_{||}| = 1/2\, \sqrt{|\nabla \circ \nabla \circ \mathbf{v}|^2 + |\boldsymbol{\beta}|^2}.$$

As I stated already in the introduction, it is impossible to obtain metric information from the velocity field. Here we obtain the curvedness scaled with the absolute value of the 3D velocity parallel to the camera.

The slant and the tilt

In order to find relations for the attitude we need the first order differential invariants. We have from eq. 5.15 and eqs 5.8 to 5.10:

$$\begin{aligned} \nabla \cdot \mathbf{v} &= 1/Z_0\, (\nabla Z \cdot \mathbf{V}_{||} + 2V_{\perp}), && (5.26)\\ \nabla \times \mathbf{v} &= 1/Z_0\, (\nabla Z \cdot J\, \mathbf{V}_{||}) + 2\Omega_{\perp}, && (5.27)\\ \nabla \circ \mathbf{v} &= 1/Z_0 \begin{pmatrix} Z_X & -Z_Y \\ Z_Y & Z_X \end{pmatrix} \mathbf{V}_{||}, && (5.28) \end{aligned}$$

Just as for the double deformation it is interesting to look at the length of the deformation and its angle with $\mathbf{V}_{||}$:

$$\begin{array}{rclcl} |\nabla \circ \mathbf{v}| &=& 1/Z_0\, \sqrt{Z_X^2 + Z_Y^2}\, |\mathbf{V}_{||}| &=& 1/Z_0\, \tan\sigma\, |\mathbf{V}_{||}|, \\ \angle(\nabla \circ \mathbf{v},\, \mathbf{V}_{||}) &=& \angle(\nabla Z,\, \mathbf{e}_x) &=& \tau. \end{array} \qquad (5.29)$$

The first of these equations gives the velocity scaled slant. We obtain an estimate of the tilt from:

$$\tau_e = \angle(\nabla \circ \mathbf{v},\, \mathrm{sign}(\boldsymbol{\beta} \cdot \mathbf{V}_{||})\, \boldsymbol{\beta}).$$

Note that when we do not know the direction of $\mathbf{V}_{||}$ we can only obtain the axis to which τ is restricted: both τ and $\pi + \tau$ lead to the same estimate of the tilt. Below I will show that this ambiguity is connected with the sign ambiguity of the shape index.

The velocity field under parallel projection

In the derivation of a linear estimate for the shape index we had to make the assumption that the non-curvature dependent terms in $\boldsymbol{\beta}$ are small and that we can obtain the direction of $\mathbf{V}_\parallel$. I also indicated an ambiguity in the shape index, the orientation of the principal curvatures and the tilt when the direction of $\mathbf{V}_\parallel$ is unknown. I will show that the assumption that the non-curvature dependent terms in $\boldsymbol{\beta}$ are zero is identical to using parallel projection in the derivation of the velocity field. From the velocity field under parallel projection it is easy to understand the ambiguities.

The expression of the velocity field of a rigid moving second order surface patch under parallel projection is derived in [39]. Because it is important to my argument I will give a summary of this derivation. Representing the patch by $Z(X, Y)$ and subjecting the observer to a translation $\mathbf{V} = (V^X, V^Y, V^Z)^T$ and rotation $\boldsymbol{\Omega} = (\Omega^X, \Omega^Y, \Omega^Z)^T$ we find for the velocity field on a planar camera under parallel projection:

$$\begin{pmatrix} v^x \\ v^y \end{pmatrix} = \begin{pmatrix} -V^X - \Omega^Y\, Z + \Omega^Z\, y \\ -V^Y + \Omega^X\, Z - \Omega^Z\, x \end{pmatrix}.$$

Note that the roles of $\mathbf{V}$ and $\boldsymbol{\Omega}$ are reversed when we compare this velocity field with the velocity field we obtained under perspective projection: under parallel projection $\boldsymbol{\Omega}$ generates information about the structure of the object, under perspective projection $\mathbf{V}$ generates this information. Introducing polar coordinates r, α in the image plane, we can write this succinctly as:

$$\mathbf{v}(r, \alpha) = -\mathbf{V}_\parallel - \Omega_\perp J\, \mathbf{e}_\alpha\, r + J\, \boldsymbol{\Omega}_\parallel Z.$$

Now, substituting the Taylor development of Z (eq. 5.4) and the expressions for the attitude and principal curvatures (eq. 5.5) and rewriting them in polar coordinates we find:

$$\begin{aligned} \mathbf{v}(r, \alpha) \;\; = \;\; & -\mathbf{V}_\parallel + J\boldsymbol{\Omega}_\parallel Z_0 + (-\Omega_\perp J\mathbf{e}_\alpha + (\nabla Z, \mathbf{e}_\alpha)\, J\boldsymbol{\Omega}_\parallel)\, r + \\ & \kappa_n(\alpha) \frac{1 + \tan^2\sigma\, \cos^2(\alpha - \tau)}{\cos\sigma} J\boldsymbol{\Omega}_\parallel\, r^2/2. \end{aligned}$$

It is now easy to calculate the second order invariants as in eqs. 5.16 to 5.18. They have only the curvature dependent terms, but are otherwise similar. Further, one can easily show that the zeroth order disappears when the observer fixates and that the first order and second order velocity field $\mathbf{v}(r, \alpha)$ do not change when we make the following substitutions:

$$\begin{aligned} \boldsymbol{\Omega}_\parallel & \rightarrow -\boldsymbol{\Omega}_\parallel, \\ \kappa_n(\alpha) & \rightarrow -\kappa_n(\alpha), \\ \tau & \rightarrow \pi + \tau. \end{aligned}$$

The change of the tilt by 180 deg is equivalent to a change of the sign of ∇Z. Because of Euler's formula (eq. 5.1), the substitution $\kappa_n(\alpha) \rightarrow -\kappa_n(\alpha)$ is equivalent to:

$$\begin{aligned} \kappa_{max} &\rightarrow -\kappa_{min}, \\ \kappa_{min} &\rightarrow -\kappa_{max}, \\ \alpha_0 &\rightarrow \pi/2 + \alpha_0. \end{aligned}$$

Thus, the velocity field of a surface patch under parallel projection is ambiguous in exactly the same way as we found before when we neglected the non-curvature dependent terms in $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$. This should come as no great surprise as I have shown that the second order invariants under parallel projection only contain curvature dependent terms. The ambiguities are easily imagined one by one: the concave/convex ambiguity is easily understood for rotating spheres with a fronto-parallel tangent plane, the interchange of the directions of maximal and minimal curvature in easily understood for a rotating symmetric saddle with a fronto-parallel tangent plane and the tilt ambiguity is easily understood with a rotating plane.

As I showed above, the ambiguities can be lifted once we know the direction of $\mathbf{V}_{||}$. A detailed discussion of how this direction can be obtained is beyond the scope of this paper. There are two possibilities how the information about the direction of $\mathbf{V}_{||}$ could be obtained. The first is the use of extra-retinal signals, like the knowledge that your left eye is on the left of your right eye in stereo vision, or knowledge of movement direction in active vision, as in the previous chapter. The second possibility is to use a third view, as described in [38].

5.5 Simulations

In deriving the various estimates related to shape I made some approximations. In this section I will show the effects of these approximations to be generally small. I will not give analytic expressions for the effect of the approximations because they tend to get unwieldy but I will rely on computer simulations instead.

In this section I will assume fixation and zero torsion i.e. I take:

$$\boldsymbol{\Omega}_{||} = -J\,\mathbf{V}_{||}/Z_0, \tag{5.30}$$

$$\Omega_{\perp} = 0. \tag{5.31}$$

Further, I will take quadratic surface patches. It should be noted that these patches generally do not have a constant shape index: except for the parabolic (cylinder-like) patches, shape index varies with position. This is not a big problem though, as long as the field of view is small. Further, for parallel projection to be reasonable, I also need a small field of view. To make a comparison with experimental results I took most of the parameters from [78]: a square field of view of 6 by 6 degrees, a distance of 2.5 m, slant zero, a curvedness of 5 m^{-1}, direction of maximal curvature in the x direction and a velocity of 1 m/s in the x direction. When I take a non-zero slant, I will always take zero tilt (and thus rotate the tangent plane around a horizontal axis). I will use this set of parameters in all simulations reported below, unless otherwise noted. I took a square grid of 5 by 5 in the image plane and calculated the velocity in each grid point. I fitted the velocity field with a polynomial up to second order, using a linear least squares algorithm. From the fitted parameters I calculated the differential invariants and from there the estimates of the shape index S_e, the velocity scaled curvature Γ_e and the angle between direction of maximal curvature and the x axis, α_e. In the experiment of van Damme et al. [78], the subject was shown many views of the object and was actively moving. Thus, we can assume the subject knows the direction of movement and we take the correct direction of $\mathbf{V}_{\|}$ in the simulations.

The vector $\boldsymbol{\beta}$ enters in all shape measures. Therefore, it is important to show the effect of the non-curvature dependent terms in $\boldsymbol{\beta}$. Substituting eq. 5.30 in eq. 5.23, we find:

$$\boldsymbol{\beta} = (Z_{XX} + Z_{YY} + 2/Z_0)\, \mathbf{V}_{\|} - 2(V_{\perp}/Z_0)\, \nabla Z. \tag{5.32}$$

There are two terms in this equation which I previously neglected. A zeroth order term $2/Z_0$, which I assumed to be small relative to $Z_{XX} + Z_{YY}$, and a first order term $2(V_{\perp}/Z_0)\, \nabla Z$, which I assumed to be small relative to $(Z_{XX} + Z_{YY})\, \mathbf{V}_{\|}$. Note that the zeroth order term only changes the length of $\boldsymbol{\beta}$, never its direction. The first order term has a more complicated influence on $\boldsymbol{\beta}$, depending on the tilt (which gives the direction of ∇Z) and $\mathbf{V}_{\|}$. When ∇Z and $\mathbf{V}_{\|}$ are parallel the first order term only changes the length, not the direction of $\boldsymbol{\beta}$. When ∇Z and $\mathbf{V}_{\|}$ are orthogonal the first order term not so much influences the length as the direction of $\boldsymbol{\beta}$.

The shape index

First, I will show the effect of neglecting the zeroth order term in eq. 5.32 on the estimate of the shape index S_e. In fig. 5–4 I have plotted the bias in shape index (difference between the estimated shape index and the shape index in the fixation point) for several values of the curvedness. Note that we would have obtained the same curves for different values of Z_0: keeping the curvedness constant at 5 m^{-1}

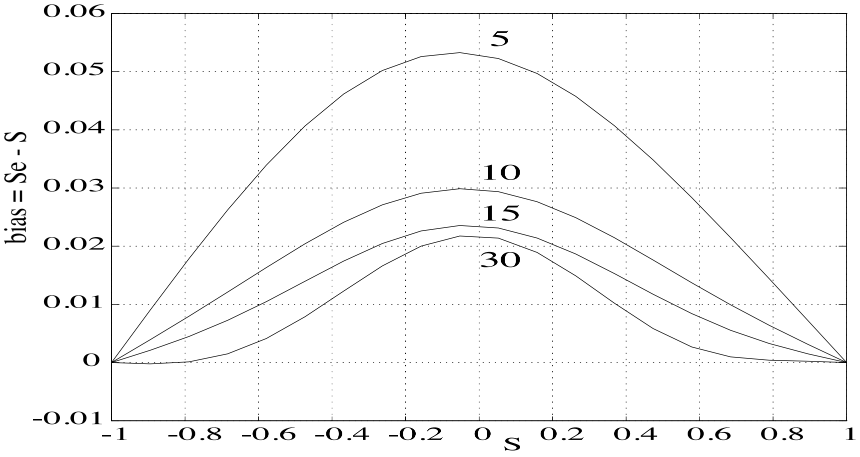


Figure 5–4 The bias (difference between the estimated shape index and the shape index at the fixation point) as a function of shape index at the fixation point. The different curves are for different values of the curvedness: 5, 10, 15 and 30 m^{-1}.

and Z_0 equal to 15, 7.5, 5 and 2.5 m results in the same curves. The bias is zero at the extremes of the shape index scale and increases towards the middle. One can show that the effect that the bias is zero for the spherical shapes, is caused by the vanishing of the double deformation at the extremes of the shape index scale. The bias is smaller for higher values of the curvedness. Incidently, one can also see the problems one would have for a series development of the bias: the lowest curve (for the highest value of curvedness) dips below zero for S around -0.9 indicating that one would have to go to fourth order in this series development (this effect is even stronger for higher values of curvedness). The effect that the bias is slightly larger for cylindrical and hyperbolical shapes than for elliptical shapes is precisely the effect reported experimentally in [78, 15]. Also, increasing the curvature decreases the bias, a weak effect that was also reported. Finally, note that the bias is slightly larger for concave than for convex surfaces [78].

Second, I will show the effect of neglecting the first order term in eq. 5.32 on the estimate of the shape index S_e. In fig. 5–5 I have plotted the bias in shape index for different values of the slant and of the sign of $V_\perp$. The middle curve is for zero slant and is the same as the upper curve in the previous figure. We see that when $V_\perp > 0$ i.e. when the observer is moving towards the surface, the bias decreases with increasing slant because the zeroth and first order terms have

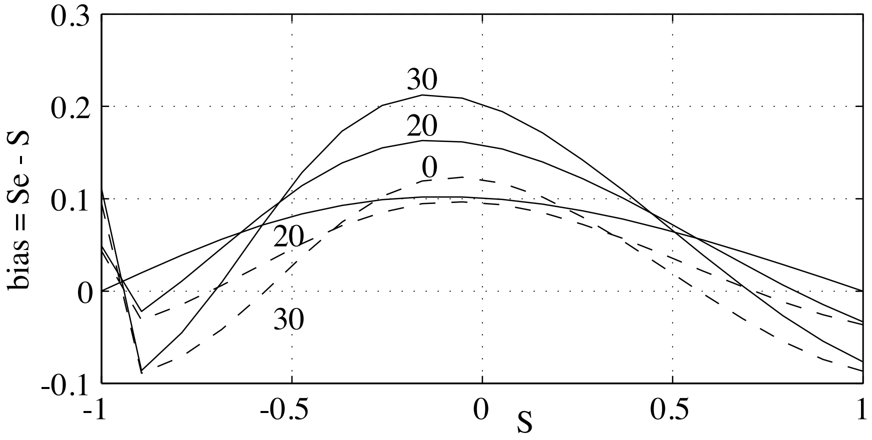


Figure 5–5 The bias in shape index as a function of shape index at the fixation point. The different curves are for different combinations of the slant and $V_\perp$. The two solid curves labelled 30 and 20 are for slant 30 and 20 degrees with $V_\perp = 1$, the remaining solid curve is for slant zero ($V_\perp$ does not matter). The dashed curves are for slant 30 and 20 degrees with $V_\perp = -1$. The other parameters are: curvedness 2.5 m^{-1}, tilt zero, distance 2.5 m and $\mathbf{V}_\parallel = (1, 0)$ m/s.

different signs. When $V_\perp < 0$ we see the opposite effect. These are the best and worst case situations: when ∇Z and $\mathbf{V}_\parallel$ are not parallel we find intermediate values of the bias. The sharp discontinuity in the bias around shape index -0.9 will be discussed in the next paragraph.

A third approximation I made in deriving S_e was to neglect the slant i.e. I assumed λ_{max} and λ_{min} to be close to κ_{max} and κ_{min}. In fig. 5–6 I have plotted the bias in shape index for different values of the slant. This shows the effect of slant to be very small for the cylindric shapes and somewhat larger for the hyperbolic and spherical shapes. The sharp discontinuity in the bias around shape index -0.9 is caused by a change of the sign of the double deformation. Also the bias cannot be negative there because the shape index cannot become smaller than -1. Still, the bias does not exceed 0.1 around shape index -0.9. Even for a slant of 30 degrees there is not much influence of the slant on the shape index, a result reported in [16].

To illustrate the stability of my estimate of the shape index against noise I added 10% multiplicative gaussian white noise to the velocity field (independent for the x and y direction). I ran 250 simulations for each value of the shape index and calculated the mean and standard deviation of S_e. In fig. 5–7 I have

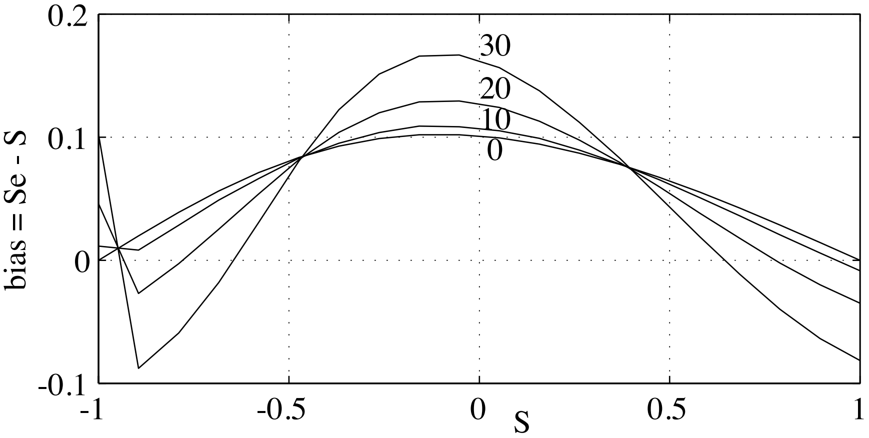


Figure 5–6 The bias in shape index as a function of shape index at the fixation point. The different curves are for different values of the slant: 0, 10, 20 and 30 degrees. The other parameters are: curvedness 2.5 m^{-1}, tilt zero, distance 2.5 m and $\mathbf{V} = (1, 0, 0)$ m/s.

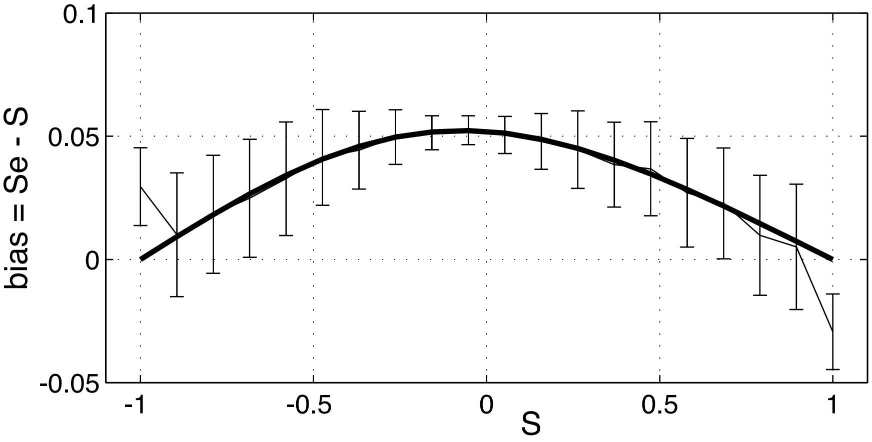


Figure 5–7 The bias in shape index as a function of shape index at the fixation point. The thick curve gives the bias without noise, the thin curve gives the mean bias of 250 simulations with 10% multiplicative gaussian white noise. The error bars denote the standard deviation. The other parameters are: curvedness 5 m^{-1}, slant zero, distance 2.5 m and $\mathbf{V} = (1, 0, 0)$ m/s.

plotted the difference between the mean of S_e and the shape index in the fixation point. Because I used the term bias for the deterministic difference I denote this difference as the statistical bias. In fig. 5–7 I have also plotted the bias in shape index from the noiseless simulation. Except for $S = \pm 1$ the statistical bias does not deviate from the bias, showing my estimate to be statistically unbiased. The deviation at $S = \pm 1$ is caused by the arctan function in the shape index, which is very nonlinear close to $S = \pm 1$. The standard deviation seems to be larger for concave surfaces than for convex ones. I performed some more simulations with noise (multiplicative and additive), which showed my estimate of the shape index to be highly resistant against noise. Variations of slant, the direction of $\mathbf{V}_{\|}$ and the curvedness had negligible influence beyond the deterministic bias. Only when I added a large movement orthogonal to the camera ($V_{\perp} \neq 0$) in combination with multiplicative noise the estimates became very noisy, because the velocities are high in the periphery and so is the noise.

The direction of maximal curvature and the curvedness

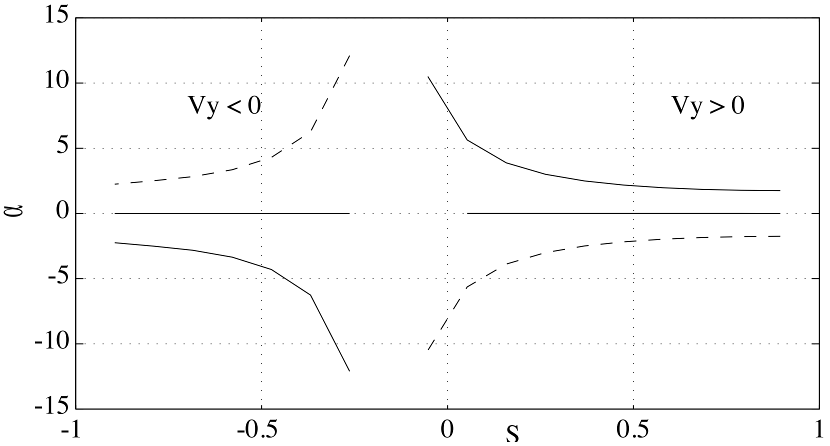


Figure 5–8 The bias in direction of maximal curvature relative to the x axis in degrees as a function of shape index at the fixation point. The middle curve is for slant zero, the dashed discontinuous curve for slant 30 degrees and $\mathbf{V} = (0, -1, 1)$ m/s and the solid discontinuous curve for slant 30 degrees and $\mathbf{V} = (0, 1, 1)$ m/s. The other parameters are: curvedness 5 m^{-1}, tilt zero and distance 2.5 m.

In fig. 5–8 I have plotted the bias in the direction of maximal curvature (difference between α_e and α_0 in the fixation point) for different values of the slant and velocity in the y direction (and $V_{\perp} > 0$). I showed that α_e can be calculated from the angle between $\boldsymbol{\beta}$ and the double deformation. When either of these becomes very small the angle cannot be determined accurately. Therefore, I did not calculate α_e when the smaller of $\boldsymbol{\beta}$ and the double deformation was less than 10% in length than the larger of the two. This happens around $S = \pm 1, 0$ which is why the curves shown are discontinuous at these values of S. For $S = \pm 1$ this is not a problem as α_0 is not defined. That the direction of maximal curvature is hard to determine for hyperbolic shapes is an interesting prediction from my theory. The horizontal curve is for slant zero, the broken line is for slant 30 degrees and $V^Y < 0$, the solid line is for slant 30 degrees and $V^Y > 0$. In discussing the effects of neglecting the zeroth and first order terms in $\boldsymbol{\beta}$ we already noted that only the first order term can change the direction of $\boldsymbol{\beta}$, when ∇Z and $\mathbf{V}_{\|}$ are not parallel. I have chosen ∇Z and $\mathbf{V}_{\|}$ orthogonal (the worst case) and still the difference does not exceed 13 degrees.

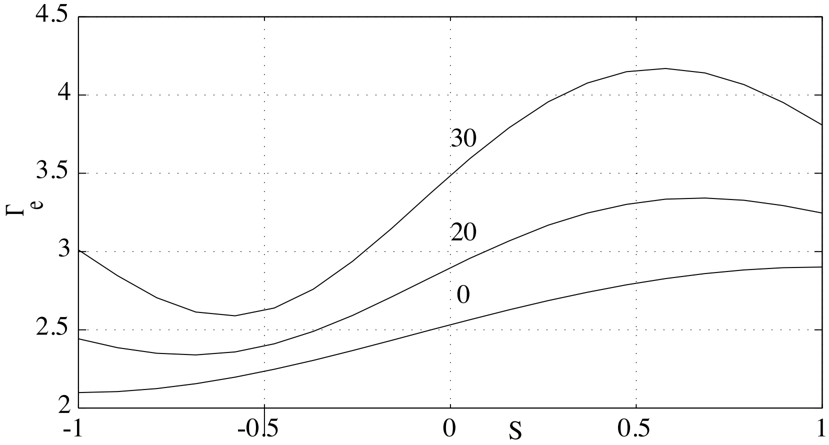


Figure 5–9 The velocity scaled curvature Γ_e as a function of the shape index at the fixation point. The correct value is 2.5 s^{-1}. The different curves are for different values of the slant: 0, 20 and 30 degrees. The other parameters are: curvedness 5 m^{-1}, tilt zero, distance 2.5 m and $\mathbf{V} = (1, 0, 0)$ m/s.

In fig. 5–9 I have plotted the estimate of velocity scaled curvedness Γ_e for different values of the slant. The value I used for the simulation was 2.5 s^{-1}. The figure shows Γ_e to be quite sensitive for slant. For higher values of the slant Γ_e leads to a considerable overestimation of the true value. Further Γ_e increases with S for slant zero. One can easily show this to be caused by the zeroth order term in $\boldsymbol{\beta}$.

5.6 Discussion

We found that we can obtain the shape index, a scale independent descriptor of 3D shape, directly from the velocity field generated by a curved patch moving rigidly relative to a planar camera. I have tried to strike a middle ground between human perception and machine vision. With the results I am able to explain some of the outcomes of psychophysical experiments (see below). I think my model is relevant for machine vision for those cases where one does not know the movement of the camera through 3D space.

Comparison with psychophysical findings

The shape index and curvedness have been used in a number of psychophysical experiments, employing both optic flow [78, 79] and stereo vision [15, 16]. In these experiments it was established that human observers can use the shape index and curvedness quite independently from one another. In a detection experiment [78, 15], where subjects were asked to classify a surface patch of unknown shape index in one of eight shape index categories, it was found that subjects could more readily classify the elliptical shapes. Performance was lower for the cylindrical and hyperbolical shapes. Performance increased only slightly with increasing curvedness. Furthermore, it was remarked [78] that the velocity of the subject did not seem to have a significant influence. In another experiment [16] subjects where shown two patches and their task was to detect which one was the reference patch. Using this paradigm and stereo vision, it was established that the discrimination of shape index is independent of the slant of the surface patch for slants up to 30 deg. All these results are in qualitative agreement with the results of the simulations in the previous section.

Comparison with Koenderink and van Doorn's approach

As stated in the introduction my approach is close to Koenderink and van Doorn's [37, 39]. I improved on their 1975 paper by introducing the remaining invariant of the second order velocity field: the double deformation. I showed this invariant to be proportional to the difference of the maximal and minimal curvature. From this I was able to construct the shape index directly from the velocity field. In their original approach they could only calculate the sign of the gaussian curvature, which would be the equivalent of calculating whether the absolute value of the shape index is smaller or larger than 0.5.

My approach is somewhat different from the approach of [39]. First Koenderink and van Doorn employed parallel projection. I used both perspective and parallel projection, which allowed me to estimate the difference. For parallel projection we found the same expressions for the second order differential invariants except for the gradients of divergence and rotation, which retained only the curvature dependent term. In particular, $\boldsymbol{\beta}$ had no zeroth order term. But it was precisely the influence of this term that allowed me to explain the experimental results of van Damme et al. [78] and de Vries et al. [15]. A second difference is that the approach of Koenderink and van Doorn results in the projected indicatrix of Dupin, scaled by the slant and the 3D velocity, whereas some of my shape measures (the shape index and the orientation of the principal curvatures) are independent of both.

Comparison with spin variation theory

A recent theory on the perception of curvature from optic flow is based on the concept of spin variation [12, 21] i.e. the bending of lines in the image. The spin variation is not a geometrical object like the differential invariants, but it is defined independent of a coordinate system. Therefore, it can be expressed in differential invariants as follows. The spin variation in the direction of the x-axis is given by:

$$SV(0) = v^y_{xx}.$$

By rotating this over an arbitrary angle θ we obtain the spin variation function (see eq. 5.11). Because we know that the spin variation depends only on second order spatial derivatives of the range function we construct differential invariants that have this property too. The double deformation already has this property: it depends only on second order spatial derivatives of the range function. We define:

$$\boldsymbol{\delta} \equiv \nabla(\nabla \cdot \mathbf{v}) + 3\, J\, \nabla(\nabla \times \mathbf{v}),$$

as a second differential invariant only dependent on second order derivatives of the range function. Using eq. 5.11, it is easy to show that:

$$SV(\theta) = -1/4\, \nabla \circ \nabla \circ \mathbf{v}\, J \begin{pmatrix} \cos 3\theta \\ \sin 3\theta \end{pmatrix} - 1/4\, \boldsymbol{\delta}\, J \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}. \tag{5.33}$$

This shows again nicely the invariance of $\nabla \circ \nabla \circ \mathbf{v}$ (weight 3) and $\boldsymbol{\delta}$ (weight 1).

Spin variation theory has been compared with the findings of psychophysical experiments where human subjects had to discriminate cylinders and planes. A central prediction of spin variation theory was the asymmetry in detection thresholds depending on the axis of the cylinder relative to the direction of movement of the observer. Movement parallel to the cylinder axis leads to lower detection thresholds than movement orthogonal to the cylinder axis. This effect has indeed been observed both in studies employing optic flow [12, 52] and employing stereo vision [63].

From the expression of the spin variation in differential invariants (eq. 5.33) it easy to see that this asymmetry for cylinders must be due to $\boldsymbol{\delta}$, because the double deformation is symmetric (take $Z_{XY} = 0$ in eq. 5.18). The vector field $\boldsymbol{\delta}$ is indeed asymmetric as can be seen by substituting eqs 5.16 and 5.17. We obtain:

$$\boldsymbol{\delta} = \begin{pmatrix} Z_{XX} + 3Z_{YY} & -2Z_{XY} \\ -2Z_{XY} & 3Z_{XX} + Z_{YY} \end{pmatrix} \mathbf{V}_{\|}.$$

By taking $Z_{XY} = 0$ and $\mathbf{V}_{\|} = (1, 0)$ we find that for a horizontal cylinder ($Z_{XX} = 0$) $\boldsymbol{\delta}$ is three times larger than for a vertical cylinder ($Z_{YY} = 0$). The measure I proposed for curvedness, Γ_e is constructed from the symmetric vector fields $\boldsymbol{\beta}$ and $\nabla \circ \nabla \circ \mathbf{v}$. Thus, assuming that this effect can be explained on the level of the velocity field, my curvedness measure cannot be relevant for this particular aspect of human perception. Precisely because of their symmetry they might be interesting for machine vision though.

Thus, the spin variation theory is not so different from the current approach in its mathematics. The difference is more in the emphasis. Where spin variation theory stresses more the computational aspects and the detection of curved vs planar surfaces, my emphasis is more on geometry and on the detection of shape as given by the shape index. Beyond that my formulation solves a few of the problems that are in the spin variation theory: the shape index is almost independent of 3D velocity and slant, whereas the spin variation function depends on both.

Comparison with discrete algorithms

I have employed the velocity field as input to my calculations i.e. I assumed a small disparity between two views of a rigid curved surface patch. The discrete algorithms of structure from motion do not make the assumptions of small motion and of a smooth surface patch. Two recent studies in this field are [35, 85]. For several reasons it is hard to compare my results directly to the results of these studies. First, the discrete algorithms start with the calculation of the velocity and rotation of the camera and given these calculate the 3D position of the points. They do not calculate an explicit structure of the environment, although one could do this in an extra step by fitting a model to the 3D points. This contrasts with my approach where I directly calculate the structure of the environment. Second, the discrete algorithms employ a two step approach: in the first step they use a linear algorithm to estimate the velocity and rotation of the camera. Because this estimate is statistically biased they use this estimate as starting point for an iteration, which minimises the discrepancy between fitted image points and observed image points. In contrast to this, my approach is linear in the parameters. Notwithstanding these differences there is an interesting point in the study by Weng et al. ([85], Fig. 16), where they did some simulations for small image motion. Using parameters comparable to theirs, I found the statistical bias in the estimates to be small, generally of the order of a few percent. This contrasts with their estimate of the bias in the velocity, which can be of the order of 100%, for small image motion.

Consequences of the theory

From my approach I am able to derive new hypotheses that could be tested in psycho-physical experiments. Most of them are related to the approximations I used to calculate the estimates of shape index, orientation of the principal curvatures, velocity scaled curvedness and tilt. These approximations are reasonable for many real-life situations but could cause a breakdown of shape perception in experimental situations. Especially the effects of the zeroth and first order term in $\boldsymbol{\beta}$ lead to testable predictions.

ambiguities I showed that my way of calculating shape amounts to using parallel projection. This introduces ambiguities in $\mathbf{V}_{\|}$, the shape index, the direction of maximal curvature and the tilt. In contrast, the curvedness and the slant can be extracted without ambiguity. In particular, this means that the difference between a planar and a curved surface can be extracted unambiguously from a two frame motion sequence. This is relevant since most psycho-physical studies searching for an effect of sequence length have used the curvedness in their tasks (e.g. [74] and the references therein). For a two-frame motion sequence I would predict ambiguities in $\mathbf{V}_{\|}$, the shape index, the direction of maximal curvature and the tilt but not in the curvedness and slant. These ambiguities could be lifted by employing longer sequences.

shape index In the calculation of the absolute value of the shape index I neglected the non-curvature dependent terms in $\boldsymbol{\beta}$. Especially the first order term is interesting because it would predict a slant dependent influence of movement orthogonal to the camera ($V_{\perp}$) on the shape index. More specifically, I would predict no influence of $V_{\perp}$ for zero slant and an increasing influence with increasing slant depending on the sign of $V_{\perp}$. For $V_{\perp} > 0$

and for ∇Z and $\mathbf{V}_{\|}$ parallel, I would predict a decreasing bias with increasing slant because the zeroth and first order terms have different signs. For $V_{\perp} < 0$ I would predict the opposite effect (see fig. 5–5).

principal directions The principal directions could be recovered from the angle between $\boldsymbol{\beta}$ and the double deformation. When either of these is small, the angle is hard to determine. The double deformation is small for spheres which is not a big problem as the principal directions are not defined for spheres (all directions have the same curvature). The prediction that the principal directions are also hard to determine for hyperbolic surfaces, where $\boldsymbol{\beta}$ is small and where the two curvatures are of opposite sign, is a surprising consequence of my theory.

Since the zeroth order term does not change the direction of $\boldsymbol{\beta}$, only its size, I would predict the estimate of the principal directions to be independent of 3D velocity, curvedness and distance. Just as for the shape index I predict a slant dependent influence of movement orthogonal to the camera ($V_{\perp}$). In this case I expect this effect to be strongest when ∇Z and $\mathbf{V}_{\|}$ are orthogonal (see fig. 5–8).

velocity scaled curvedness I already discussed that the velocity scaled curvedness is a symmetric quantity and thus cannot be used to explain the asymmetry found in the detection of curvature of cylinders. Still, it could be used by human observers for the detection of symmetric surfaces. An interesting prediction of the theory is that the velocity scaled curvedness would be overestimated for slanted objects and that this overestimation would depend on shape index (see fig. 5–9).

tilt The tilt of the tangent plane could be recovered from the angle between $\boldsymbol{\beta}$ and $\nabla \circ \mathbf{v}$. I would predict the pattern of dependencies to be the same as for the principal directions.

Bibliography

[1] A Azarbayejani, T Starner, B Horowitz and A Pentland, Visually controlled graphics. IEEE Trans Patt Anal Mach Intell, **15**, 602–605 (1993).

[2] K Ball and M Pierowsky, A modified direct linear transformation method (DLT) calibration procedure to improve the accuracy of 3D reconstruction for large volumes. In: G de Groot, AP Hollander, PA Huijing and GJ van Ingen Schenau (Eds), Biomechanics XI-B, vol 7-B in Int Series on Biomechanics. Free University, Amsterdam (1987).

[3] E Batschelet, Circular Statistics in Biology. Academic Press, London (1981).

[4] A Berthoz, M Lacour, J Soechting and P Vidal, The role of vision in the control of posture during linear motion. Prog Brain Res **50**, 197–209 (1979).

[5] PJ Besl and RC Jain, Invariant surface characteristics for 3D object recognition and range images. Comp Vis Graph Imag Proc **33**, 33–80 (1986).

[6] T Bohr, P Bak and MH Jensen, Transition to chaos by interaction of resonances in dissipative systems. II. Josephson junctions, charge-density waves, and standard maps. Phys Rev A **30**, 1970–1981 (1984).

[7] DM Bramble and FA Jenkins jr, Mammalian locomotor-respiratory integration: implications for diaphragmatic and pulmonary design. Science, **262**, 235–240 (1993).

[8] ML Braunstein and JS Tittle, The observer-relative velocity field as the basis for effective motion parallax. J of Exp Psych: Human Perc and Perf, **14**, 582–590 (1988).

[9] ML Braunstein and J Todd, On the distinction between artifacts and information. J of Exp Psych: Human Perc and Perf, **16**, 211–216 (1990).

[10] AM Bronstein and MA Gresty, Short latency compensatory eye movement responses to transient linear head acceleration: a specific function of the otolith-ocular reflex. Exp Brain Res, **71**, 406–410 (1988).

[11] GC Carter, Coherence and time delay estimation. In: JL Lacoume, TS Durrani and R Stora (Eds), Signal Processing, vol 65 in Les Houches, Session XLV. Elsevier Science Publishers, Amsterdam (1987).

[12] V Cornilleau-Pérès and J Droulez, Visual perception of surface curvature: psychophysics of curvature detection induced by motion parallax. Percept Psychophys **46**, 351–364 (1989).

[13] V Cornilleau-Pérès and J Droulez, Stereo-motion cooperation and the use of motion disparity in the visual perception of 3-D structure. Percept Psychophys, **54**, 223–239 (1993).

[14] V Cornilleau-Pérès and J Droulez, The visual perception of 3D shape from self-motion and object-motion. submitted to Vision Res (1994).

[15] SC de Vries, AML Kappers and JJ Koenderink, Shape from stereo: a systematic approach using quadratic surfaces. Percept Psychophys **53**, 71–80 (1993).

[16] SC de Vries, AML Kappers and JJ Koenderink, Influence of surface attitude and curvature scaling on discrimination of binocularly presented curved surfaces. submitted to Vision Res (1994).

[17] J Dichgans and T Brandt, Visual-vestibular interaction: Effects on self-motion perception and postural control. In: R Held, HW Leibowitz and HL Teuber (Eds), Perception. vol 8 in Handbook of Sensory Physiology. Springer, Berlin (1978).

[18] TMH Dijkstra, CCAM Gielen and BJM Melis, Postural responses to stationary and moving scenes as a function of distance to the scene. Hum Mov Science **11**, 195–203 (1992).

[19] TMH Dijkstra, E Argante and CCAM Gielen, Motion parallax from catastrophies in scale-space. In: S. Gielen and B. Kappen (Eds), Proceedings of the International Conference on Artificial Neural Networks, Amsterdam, The Netherlands, Springer, London (1993).

[20] PC Dodwell, The Lie transformation group model of visual perception. Percept Psychophys **34**, 1–16 (1983).

[21] J Droulez and V Cornilleau-Pérès, Visual perception of surface curvature. The spin variation and its physiological consequences. Biol Cybern **62**, 211–224 (1990).

[22] FB Eiseman jr, Bali: sekala & niskala, volume I. Periplus Editions, Berkeley CA (1989).

[23] CJ Erkelens, J Van der Steen, RM Steinman and H Collewijn, Ocular vergence under natural conditions. I. Continuous changes of target distance along the median plane. Proc R Soc Lond B, **236**, 417–440 (1989).

[24] L Ferman, H Collewijn, TC Jansen and AV Van den Berg, Human gaze stability in the horizontal, vertical and torsional direction during voluntary head movements, evaluated with a three-dimensional scleral induction coil technique. Vision Res, **27**, 811–828 (1987).

[25] SK Fisher and KJ Ciufredda, Accommodation and apparent distance. Perception **17**, 609–621 (1988).

[26] JL Gallant, J Braun and DC Van Essen, Selectivity for polar, hyperbolic and cartesian gratings in macaque visual cortex. Science **259**, 100–103 (1993).

[27] J Guckenheimer and P Holmes, Nonlinear oscillations, dynamical systems, and bifurcations of vector fields. Springer Verlag, New York (1983).

[28] K Hayashibe, Reversals of visual depth caused by motion parallax. Perception, **20**, 17–28 (1991).

[29] WL Hays, Statistics, 4th edition. Holt, Rinehart and Winston, Orlando FL (1988).

[30] R Held and A Hines. J Comp Physiol Psychology **56**, 872 (1963).

[31] WC Hoffman, The Lie algebra of visual perception. J Math Psychology **3**, 65–98 (1966).

[32] BKP Horn, Closed-form solution of absolute orientation using unit quaternions. J Opt Soc Am A **4**, 629–642 (1987).

[33] MH Jensen, P Bak and T Bohr, Transition to chaos by interaction of resonances in dissipative systems. I. Circle maps. Phys Rev A **30**, 1960–1969 (1984).

[34] K Kanatani, Group-theoretical methods in image understanding. Springer, Berlin (1990).

[35] K Kanatani, Unbiased estimation and statistical analysis of 3-D rigid motion from two views. IEEE Trans Patt Anal Mach Intell **15**, 37–50 (1993).

[36] JAS Kelso, JD Delcolle and G Schöner, Action-Perception as a pattern formation process. In: M Jeannerod (Ed), Attention and Performance XIII. L Erlbaum Assoc, Hillsdale NJ (1990).

[37] JJ Koenderink and AJ van Doorn, Invariant properties of the motion parallax field due to the movement of rigid bodies relative to an observer. Optica Acta **22**, 773–791 (1975).

[38] JJ Koenderink and AJ van Doorn, Affine structure from motion. J Opt Soc Am A **8**, 377–385 (1991).

[39] JJ Koenderink and AJ van Doorn, Second-order optic flow. J Opt Soc Am A **9**, 530–538 (1992).

[40] JJ Koenderink and AJ van Doorn, Surface shape and curvature scales. Image and Vis Comput **10**, 557–565 (1992).

[41] DN Lee and JR Lishman, Visual proprioceptive control of stance. J Hum Mov Studies **1**, 87–95 (1975).

[42] DN Lee, The optic flow field: the foundation of vision. Philos Trans R Soc Lond (Biol) **290**, 169–179 (1980).

[43] F Lestienne, J Soechting and A Berthoz, Postural readjustments induced by linear motion of visual scenes. Exp Brain Res **28**, 363–384 (1977).

[44] HC Longuet-Higgins and K Prazdny, The interpretation of a moving retinal image. Proc R Soc London Ser B **208**, 385–387 (1980).

[45] NA Macmillan and CD Creelman, Detection theory: a user's guide. Cambridge University Press, Cambridge, U.K. (1991).

[46] LT Maloney, Confidence intervals for the parameters of psychometric functions. Percept Psychophys **47**, 127–134 (1990).

[47] PZ Marmarelis and VZ Marmarelis, Analysis of physiological systems: The white-noise approach. Plenum, New York (1978).

[48] SL Marple, Digital spectral analysis with applications. Prentice Hall, Englewood Cliffs NJ (1987).

[49] D Marr, Vision. WH Freeman, New York (1982).

[50] D Mok, A Ro, W Cadera, JD Crawford and T Vilis, Rotation of Listing‘s plane during vergence. Vision Res, **32**, 2055–2064 (1992).

[51] K Nakayama, Differential motion hyperacuity under conditions of common image motion. Vision Res, **21**, 1475–1482 (1981).

[52] JF Norman and JS Lappin, The detection of surface curvatures defined by optical motion. Percept Psychophys **51**, 386–396 (1992).

[53] MJ Norušis, SPSS for windows, advanced statistics, release 5. SPSS Inc, Chicago (1992).

[54] B O'Neill, Elementary differential geometry. Academic Press, New York (1966).

[55] W Paulus, A Straube, S Krafczyk and T Brandt, Differential effects of retinal target displacement, changing size and changing disparity in the control of anterior/posterior and lateral body sway. Exp Brain Res **78**, 243–252 (1989).

[56] RB Post, Circular vection is independent of retinal eccentricity. Perception, **17**, 737–744 (1988).

[57] WH Press, BP Flannery, SA Teukolsky and WT Vetterling, Numerical recipes in C. Cambridge University Press, Cambridge (1988).

[58] B Raßler, S Waurick and D Ebert, Einfluß zentralnervöser Koordination im Sinne von Holsts auf die Steuerung von Atem- und Extremitätenmotorik des Menschen. Biol Cybern **63**, 457–462 (1990).

[59] LE Reichl, A modern course in statistical mechanics. Edward Arnold, Great Brittain (1980).

[60] GE Riccio, Information in movement variability about the qualitative dynamics of posture and orientation. In: KM Newell and DM Corcos (Eds), Variability and motor control, Human Kinetics, Champaign IL (1993).

[61] BJ Rogers and M Graham, Motion parallax as an independent cue for depth perception. Perception, **8**, 125–134 (1979).

[62] BJ Rogers and M Graham, Similarities between motion parallax and stereopsis in human depth perception. Vision Res, **22**, 261–270 (1982).

[63] BJ Rogers and R Cagenello, Disparity curvature and the perception of three-dimensional surfaces. Nature **339**, 135–137 (1989).

[64] S Rogers and BJ Rogers, Visual and nonvisual information disambiguate surfaces specified by motion parallax. Percept Psychophys, **52**, 446–452 (1992).

[65] G Schöner and JAS Kelso, Dynamic pattern generation in behavioural and neural systems. Science **239**, 1513–1520 (1988).

[66] G Schöner, Dynamic theory of action-perception patterns: the "moving room" paradigm. Biol Cybern **64**, 455–462 (1991).

[67] D Schuster and RE Talbott, Optimal and adaptive control in canine postural regulation. Am J Physiol **239**, R93–R114 (1980).

[68] G Sperling, BA Dosher and MS Landy, How to study the kinetic depth effect experimentally. J of Exp Psych: Human Perc and Perf, **16**, 445–450 (1990).

[69] TA Stoffregen, Flow structure versus retinal location in the optical control of stance. J of Exp Psych: Human Perc and Perf, **11**, 554–565 (1985).

[70] TA Stoffregen, The role of optical velocity in the control of stance. Percept Psychophys **39**, 355–360 (1986).

[71] M Subbarao, Interpretation of image flow: rigid curved surfaces in motion. Int J Comp Vision **2**, 77–96 (1988).

[72] RE Talbott, Postural reactions of dogs to sinusoidal motion in the peripheral visual field. Am J Physiol **239**, R71–R79 (1980).

[73] RE Talbott and RE Brookhart, A predictive model study of the visual contribution to canine postural control. Am J Physiol **239**, R80–R92 (1980).

[74] JT Todd and JF Norman, The visual perception of smoothly curved surfaces from minimal apparent motion sequences. Percept Psychophys **50**, 509–523 (1991).

[75] JR Tresilian, Emperical and theoretical issues in the perception of time to contact. J of Exp Psych: Human Perc and Perf **17**, 865–876 (1991).

[76] WNJC van Asten, CCAM Gielen and JJ Denier van der Gon, Postural adjustments induced by simulated motion of differently structured environments. Exp Brain Res **73**, 371–383 (1988a).

[77] WNJC van Asten, CCAM Gielen and JJ Denier van der Gon, Postural movements induced by rotations of visual scenes. J Opt Soc Am A **5**, 1781–1789 (1988b).

[78] WJM van Damme and WA van de Grind, Active vision and the identification of three-dimensional shape. Vision Res **33**, 1581–1587 (1993).

[79] WJM van Damme, FH Oosterhoff and WA van de Grind, Discrimination of 3D shape and 3D curvature in active vision. submitted to Percept Psychophys (1994).

[80] AV van den Berg and H Collewijn, Human smooth pursuit: effects of stimulus extent and of spatial and temporal constraints of the pursuit trajectory. Vision Res, **26**, 1209–1222 (1986).

[81] ALH van der Meer, Arm movements in the neonate: establishing a frame of reference for reaching. In: PJ Beek, RJ Bootsma and PCW van Wieringen (Eds), Studies in perception and action, ICEPA VI, Amsterdam, The Netherlands (1991).

[82] E von Holst, On the nature of order in the nervous system. In: R Martin (trans) The behavioural physiology of animals and man, vol 1. Methuen, London (1937/1973).

[83] WH Warren and KJ Kurtz, The role of central and peripheral vision in perceiving the direction of self-motion. Percept Psychophys, **51**, 443–454 (1992).

[84] AM Waxman, B Kamgar-Parsi and M Subbarao, Closed-form solutions to image flow equations for 3D structure and motion. Int J Comp Vision **1**, 239–258 (1987).

[85] J Weng, N Ahuja and TS Huang, Optimal motion and structure estimation. IEEE Trans Patt Anal Mach Intell **15**, 864–884 (1993).

[86] P Werkhoven and JJ Koenderink, Extraction of motion parallax structure in the visual system I & II. Biol Cybern **63**, 185–199 (1990).

[87] R Wiesenfeld and I Satija, Noise tolerance of frequency-locked dynamics. Phys Rev B **36**, 2483–2491 (1987).