LOCAL INVARIANTS FOR BIOLOGICAL MOTION PERCEPTION

by

Dorita Hue Fung Chang

A thesis submitted to the Centre for Neuroscience Studies
In conformity with the requirements for
the degree of Doctor of Philosophy

Queen’s University
Kingston, Ontario, Canada
July, 2010

Copyright ©Dorita Hue Fung Chang, 2010
Abstract

Observers can retrieve the facing direction of a walker from point-light displays that are devoid of structure-from-motion information and retain solely local motion signals. This ability is orientation-dependent and relies on the motions representing the feet of the agent. The experiments described here were designed to investigate visual sensitivity to local cues contained in biological motion. Initial experiments revealed that local biological motion carries information about animacy in addition to the agent’s facing direction in an orientation-dependent manner (Chapter 2). The mechanism underlying the perception of local biological motion can be dissociated from that underlying the retrieval of global structure-from-motion information according to characteristics such as sensitivity to learning and noise (Chapter 3). Further experiments revealed that the orientation-dependency for perceiving local biological motion is carried by vertical acceleration in the foot’s motion (Chapter 4). The importance of acceleration for biological motion perception raises the need to achieve a better understanding of acceleration sensitivity across various parameters such as stimulus size. To this end, Chapter 5 showed that acceleration thresholds for perceiving a linearly accelerating stimulus scale according to mean velocity as predicted by size invariance and are inversely proportional to stimulus duration. An important role for acceleration for the perception of biological motion was further corroborated by findings in an evolutionarily guided psychophysical search for the adequate local motion, defined as one that carries maximal directional information and a large inversion effect (Chapter 6). Finally, although orientation-dependency is a pervasively demonstrated characteristic of biological motion perception, the reference systems in which the stimuli are encoded are unclear. The experiments in Chapter 7 revealed that both global structure and local motion aspects of biological motion, like faces, are primarily coded in an egocentric frame of reference. Unlike
faces however, there is an additional contribution of non-visual information about gravity for the perception of biological motion. These findings are finally discussed in the context of emerging behavioural, neuroimaging, and electrophysiological work that further characterize a local motion mechanism that is proposed to serve as a fundamental first stage towards interpreting animate motion patterns.
Acknowledgements

I am thankful to my supervisor, Niko Troje, whose guidance and encouragement enabled me to develop tremendous knowledge and skills en route to completing this thesis. It is a pleasure to thank additionally all the other members of the BioMotionLab who created a fun environment to work in during the past four years.

I would like to show my gratitude to members of my advisory committee who have always made time in their busy schedules to provide assistance and guidance when approached. I am also grateful to Laurence Harris, who hosted my visit to his laboratory to complete a collaborative project and who was more than willing to hang upside-down for the sake of science! Lastly, I am indebted to my family for their limitless support while I strive to complete my endeavours.
Co-Authorship

I. Co-Authorship Declaration

I hereby declare that this thesis incorporates material that is result of joint research, as follows:

This thesis incorporates the outcome of a joint research undertaken in collaboration with Professor Laurence Harris at York University, Canada, under the supervision of Professor Nikolaus Troje (Appendix A). The collaboration is covered in Chapter 7 of the thesis. In all cases, the experimental designs, data analysis, and interpretation were performed by the author, and the contribution of co-authors was primarily through the provision of providing guidance in their areas of expertise and providing comments and suggestions for the resulting manuscript.

II. Declaration of Previous Publication

This thesis includes four original papers that have been published in peer reviewed journals, as follows:

<table>
<thead>
<tr>
<th>Thesis chapter</th>
<th>Publication title/full citation</th>
<th>Publication status</th>
</tr>
</thead>
</table>
I certify that I have obtained a written permission from the copyright owners to include the above published materials in my thesis (Appendix B). I certify that the above material describes work completed during my registration as a graduate student at Queen’s University.
# Table of Contents

Abstract .......................................................................................................................... ii
Acknowledgements ......................................................................................................... iv
Co-Authorship ................................................................................................................ v
List of Figures ................................................................................................................ xi
List of Tables .................................................................................................................. xiii
Chapter 1 General Introduction ..................................................................................... 1
  1.1 Retrieving information from biological motion ...................................................... 1
  1.2 Distinct inversion effects with global and local sources ......................................... 4
  1.3 Towards a better understanding of a local visual filter .......................................... 7
References .................................................................................................................... 16
Chapter 2 Perception of Animacy and Direction ............................................................ 21
  2.1 Introduction ............................................................................................................ 21
  2.2 Experiment 1 .......................................................................................................... 24
    2.2.1 Methods .......................................................................................................... 25
    2.2.2 Results ........................................................................................................... 28
    2.2.3 Discussion ...................................................................................................... 29
  2.3 Experiment 2 .......................................................................................................... 30
    2.3.1 Methods .......................................................................................................... 30
    2.3.2 Results ........................................................................................................... 31
    2.3.3 Discussion ...................................................................................................... 33
  2.4 Experiment 3 .......................................................................................................... 33
    2.4.1 Methods .......................................................................................................... 33
    2.4.2 Results ........................................................................................................... 34
    2.4.3 Discussion ...................................................................................................... 38
  2.5 General Discussion ................................................................................................. 39
References .................................................................................................................... 43
Chapter 3 Dissociating Global and Local Mechanisms .................................................. 48
  3.1 Introduction ............................................................................................................ 48
  3.2 Methods .................................................................................................................. 54
    3.2.1 Participants .................................................................................................... 54
Chapter 8 General Discussion

8.1 On the importance of local motion ................................................................. 198
8.2 Other emerging knowledge: Attention and peripheral vision ...................................... 199
8.3 Further considerations: Neural concomitants ..................................................... 204
8.4 Closing: Returning to the big picture .............................................................. 207

References.................................................................................................................. 209

Summary and Conclusions ...................................................................................... 213

Appendix A Permission of Co-Authors ........................................................................ 214
Appendix B Copyright Waiver ....................................................................................... 215
Appendix C Supplementary material for Chapter 6: Determining the Fourier series .......... 216
List of Figures

Figure 2.1. Static frames of coherent and spatially scrambled walkers ........................................26
Figure 2.2. Animacy ratings for coherent and scrambled displays ...........................................29
Figure 2.3. Direction discrimination accuracies for coherent and scrambled displays ..............32
Figure 2.4. Correlation between animacy rating and direction discrimination accuracy ..........36
Figure 3.1. Error rates on a walker detection task across walker type and mask density ..........59
Figure 3.2. Error rates on a walker detection task across learning blocks ...............................60
Figure 3.3. Error rates on a direction discrimination task for coherent and scrambled walkers across walker type and mask density .................................................................62
Figure 3.4. Error rates on a direction discrimination task for coherent and scrambled walkers across learning blocks ......................................................................................................63
Figure 4.1. Schematic depictions of the fragments of the human foot motion used to derive the stimuli of Chapter 4 ........................................................................................................................................82
Figure 4.2. Performances on a direction discrimination task for stimuli derived from single or paired fragments of the human foot trajectory ..................................................................................................................................................85
Figure 4.3. Schematic depictions of the fragments of other foot motions ................................ 94
Figure 4.4. Direction discrimination performances for stimuli derived from paired fragments of different foot motions, collapsed across fragment pairs ..................................................96
Figure 4.5. Direction discrimination performances for stimuli derived from paired fragments of different foot motions, across fragment pairs .................................................................................................................................97
Figure 4.6. Correlation between acceleration and the inversion effect ...................................101
Figure 4.7. Direction discrimination performances for naturally-accelerating stimuli and constant speed stimuli .........................................................................................................................................................106
Figure 5.1. Acceleration thresholds for horizontal and vertical axes across the two base velocities and sizes .........................................................................................................................................................130
Figure 5.2. Mean slope from linear regressions applied to the acceleration thresholds of individual participants .........................................................................................................................................................132
Figure 5.3. The duration required to attain a value of the constant of 0.15 for various starting temporal positions along the trajectory of a human walker foot ..................................................134
Figure 6.1. Change in behavioural performance across generations for Run 1 during the search for the adequate local biological motion ........................................................................................................154
Figure 6.2. Change in behavioural performance across generations for Run 2 during the search for the adequate local biological motion. ............................................................................................................. 155
Figure 6.3. The progression of stimuli across generations within the first two principle axes for Run 1 ................................................................................................................................................. 156
Figure 6.4. The progression of stimuli across generations within the first two principle axes for Run 2 ................................................................................................................................................ 157
Figure 6.5. The progression of Runs 1 and 2 within a common space defined by the first two principle components from a pooled analysis .................................................................................................................. 158
Figure 6.6. Loadings of the eight search parameters for the first and second principle axes. .... 159
Figure 6.7. The change in acceleration and amplitudes for both vertical and horizontal components of the stimuli across generations for Runs 1 and 2 and comparisons with the inversion effect .............................................................................................................................................. 161
Figure 6.8. Changes in the distribution of vertical acceleration within a region defined analogously to the swing phase of the gait cycle across generations ................................................................................. 164
Figure 6.9. The proportion of the vertical swing phase trajectory during which acceleration was positive versus negative across generations ...................................................................................................................................... 165
Figure 7.1. Schematic depiction of the “tumbling room”. ................................................................................................. 165
Figure 7.2. Schematic depictions of 12 room/observer/stimulus configurations used to investigate reference frames for biological motion and face perception ......................................................................................... 176
Figure 7.3. Static frames taken from the coherent biological motion, scrambled biological motion, and face stimuli ............................................................................................................................................... 178
Figure 7.4. Direction discrimination performances on the biological motion task, collapsed across all room/observer/stimulus configurations ........................................................................................................... 183
Figure 7.5. Performances on the biological motion and face recognition tasks across the different configurations ...................................................................................................................................................... 185
Figure 7.6. A comparison of the experimental data and model predictions for the different configurations ......................................................................................................................................................... 192
Figure C.1. Performances on a direction discrimination task expressed across the different series and walker types. .............................................................................................................................................. 218
Figure C.2. Performances on the direction discrimination task for the two stimulus organizations and two orientations, collapsed across walker type and series ................................................................................................................................. 219
List of Tables

Table 4.1. Average vertical accelerations and inversion score for each fragment pair from which stimuli were derived in Chapter 4................................................................. 100

Table 7.1. Relative weightings of the egocentric, gravitational, and visual-environmental reference systems for biological motion and face perception........................................ 191
Chapter 1

General Introduction

1.1 Retrieving information from biological motion

A set of moving light-points representing the motions of the major joints of an agent readily evokes the percept of an animate being (Johansson, 1973). A vast collection of literature has highlighted the numerous, additional characteristics that can be extracted from such biological motion displays including the agent’s gender (e.g., Barclay, Cutting, & Kozlowski, 1978; Kozlowski & Cutting, 1977; Troje, 2002) identity (Cutting & Kozlowski, 1977), emotion (e.g., Dittrich, Troschianko, Lea, & Morgan, 1996; Pollick, Paterson, Bruderlin, & Sanford, 2001), and action (e.g., Pollick, Fidopiastis, & Braden, 2001).

The impressive capacity of observers to recognize and extract a wide range of information from biological motion patterns depicted by a mere set of moving points has led a search for specialized neural mechanisms dedicated to processing these stimuli. Indeed, several dissociation studies support the existence of dedicated processes for analyzing animate motions. For example, it has been shown that motion blind patients can discriminate biological motion (McLeod, Dittrich, Driver, Perrett, & et al., 1996; Vaina, Lemay, Bienfang, Choi, & Nakayama, 1990). In turn, patients with intact motion perception have been reported to be unable to perceive biological motion (Schenk & Zihl, 1997). It should be noted, however, that in the latter study it is not possible to discern whether the observed impairments were specific to biological motion, or reflected a general structure-from-motion deficit.
Nonetheless, many efforts have been made to understand the nature of the mechanisms that underlie the retrieval of information from biological motion patterns. Consider that at least two distinct sources of information are offered by a point-light display: local motion carried by the individual dots and global, structure-from-motion information obtained via spatiotemporal integration of the display’s constituent dots. The retrieval of the global structure of the agent has been argued to involve direct arrangement of local motion signals in a hierarchical manner to recover articulated structure (e.g., Johansson, 1973; Webb & Aggarwal, 1982; Hoffman & Flinchbaugh, 1982), or to involve the mapping of body deformations to stored templates (e.g., Beintema & Lappe, 2002; Lange & Lappe, 2006).

The very first model of biological motion processing, a perceptual vector analysis, proposed by Johansson (1973) is of the view that the recovery of global structure proceeds in a hierarchical manner. According to Johansson, common motion components effectively provide a reference frame for the interpretation of deviant components. In the context of a human movement pattern, there is a hierarchical nesting of pendular motions. That is, the shoulders and hips have a common semitranslatory motion and the knees and elbows are perceived as pendular motions relative to those common components. In turn, the ankle describes pendulum motion relative to the knee, much as the wrist describes pendulum motion relative to the elbow. This model then, describes biological motion perception as a process of structuring the movement pattern in terms of hierarchies of common global and deviating local motion components. Other models such as the fixed axis model suggested by Webb and Aggarwal (1982) or the planarity model of Hoffman and Flinchbaugh (1982) differ in the way individual elements are analyzed, but have in common, the proposition that a unified structural percept is formed only after recovering information about individual local components.
Still, experimental data have shown that local motion is not necessary for the recovery of global, biological form (e.g., Beintema & Lappe, 2002; Lange & Lappe, 2006). For example, Beintema and Lappe (2002) tested observers’ ability to perceive a moving human figure that was devoid of local motion. In this study, local motion was removed by reallocating each point of the walker to another randomly selected position on the same limb or segment in successive frames. Under these conditions, observers still spontaneously reported seeing walking persons. Correspondingly, Bertenthal and Pinto (1994) embedded a target walker in a mask of additional walker point-lights that masked local motion and found that observers could still detect a walking figure from global structure-from-motion information. Lange and Lappe (2006) proposed that form information can be extracted from biological motion patterns via the mapping of postures to stored templates. In this study, the authors introduced an algorithm that used static postures from walking humans as stored templates and compared simulated performances from the model to psychophysical results. The results showed that the simulation data were highly correlated with psychophysical data obtained in various experiments that varied stimulus duration and local motion signals. That is, there was a strong dependency on available form information in the stimulus for both the simulation and psychophysical data. The authors interpreted their findings to suggest that local motion signals are not critical for the retrieval of form from biological motion.
1.2 Distinct inversion effects with global and local sources

Regardless of the exact role of local motion for the recovery of global form, it has recently become apparent that local motion carries additional, significant information for biological motion perception. Recent advances in understanding the contributions of local motion were prompted by investigations relating to the orientation-dependency for perceiving biological motion (Troje & Westhoff, 2006).

The perception of many classes of visual stimuli is orientation-dependent. For example, the recognition of faces is impaired if the stimulus is inverted (e.g., Yin, 1969; see Valentine, 1988 for a review). Inverting a face also renders any distortions to the face less apparent (e.g., Bartlett & Searcy, 1993). One striking demonstration of this phenomenon is the Thatcher illusion introduced by Thompson (1980). Thompson showed that even if the orientation of the eyes and mouth are restored to their veridical upright orientation, the expression on the otherwise inverted face is not perceived to be significantly different from one in which all parts are inverted. It is thought that the inversion effect observed for face perception is due to disruptions of configural rather than featural processing with stimulus inversion. To this end, Freire, Lee, and Symons, (2000) presented observers with face images that were altered in terms of the spatial (configural) relationships of the facial features (eyes, nose, etc.), and with images in which the specific facial features were replaced with those of other faces. The authors showed that performances on their same-different face discrimination task were impaired when the faces were inverted, but for the configurally-altered faces only. There was no difference in discrimination ability between upright and inverted stimuli when the faces differed in terms of featural information.

Perceptual inversion effects are not restricted to static stimuli, but extend also to the perception of animate or biological motion. For example, upon the inversion of a display, the
perceived quality or naturalness of movement as reported by observers is degraded (Sumi, 1984). Observers’ abilities to recognize animals (Pavlova, 1989), judge action type (Dittrich, 1993), discriminate gender (Barclay et al., 1978), detect walkers behind a mask of additional walker point-lights (Bertenthal & Pinto, 1994) are also disrupted upon inverting biological motion displays. The inversion effect for biological motion perception has traditionally been interpreted as being related to impaired processing of global or configural information (Bertenthal & Pinto, 1994), much analogous to the inversion effect observed in the domain of face perception (Freire, Lee, & Symons, 2000).

More recent findings have indicated, however, that there is in fact a second source to the inversion effect observed for biological motion perception – one that is not related to configural processing but is instead attributed entirely to local motion processing (Troje & Westhoff, 2006). In this study, Troje and Westhoff (2006) presented observers with intact and scrambled point-light displays in which the walker’s constituent dots were spatially perturbed. The displays were shown in upright and inverted orientations and observers were required to discriminate the facing direction of the walkers. Performances for the coherent displays dropped significantly when the displays were shown upside-down as compared to upright. This particular result is consistent with previous demonstrations with intact walkers and may well be attributable to deficits in global, structure-from-motion processing at the inverted orientation (Bertenthal & Pinto, 1994). Of particular significance, however, are the results for the scrambled displays which did not convey information regarding the structure of the walker. For these displays, observers were still able to extract the facing direction of the walker. Moreover, inverting these displays resulted in a reduction of accuracy rates comparable to that for the coherent walkers – an inversion effect that must be entirely local in nature. Further investigations by inverting specific parts of the display
revealed that the local inversion effect is carried by the motions representing the feet of the walker.

The relevance of local motion and in particular, motion of the feet for biological motion perception has been highlighted in the past (Mather, Radford, & West, 1992). Mather et al. (1992) manipulated their walker’s inter-frame interval and frame increment while observers completed two tasks: a left/right direction discrimination task and a coherence discrimination task. In the second task, observers were required to discriminate between displays in which the upper and lower portions of the walker indicated a coherent or conflicting facing direction. The authors found that performances on these tasks were reliable only when the inter-frame interval and frame increments of the point-light stimulus were short, consistent with the restricted temporal and spatial properties of local motion detecting processes (e.g., Baker & Braddick, 1985). (Note however, a later study by Thornton, Pinto, and Shiffrar, 1998 showing that with sufficiently long trial durations, observers can perceive biological motion even at extended temporal intervals). Critically, the importance of the motions of the distal limbs was further demonstrated by Mather et al. (1992) in subsequent experiments that showed performances on their tasks were significantly affected by the omission of the wrist and ankle dots. By contrast, omission of the shoulder and hip, or elbow and knee dots did not result in significant impairments in performances.

The need to consider the contributions of local motion independently of the retrieval of global structure-from-motion is thus apparent. Troje and Westhoff (2006) and Troje (2008) suggested a role for local biological motion that has particular ecological and social significance. Briefly, local biological motion is hypothesized to carry characteristics that are common to locomoting, terrestrial agents. The perception of such characteristics is thought to be subserved
by a specialized visual filter that facilitates the detection of animate agents in the environment.

What might be the critical information in local motion? Troje and Westhoff suggested the critical cue to be some invariant that results as the foot interacts with gravitational force during locomotion. That is, they suggested that the visual system employs knowledge about the effects of gravity on the biomechanics of terrestrial agents. Indeed, the notion that observers may use inherent knowledge about the effects of gravity for visual perception has been implied in the past both for the perception of inanimate dynamic events (e.g. Hecht, Kaiser, & Banks, 1996; McConnell, Muchisky, & Bingham, 1998; Pittenger, 1985; Stappers & Waller, 1993) and for the perception of biological motion (e.g. Jokisch & Troje, 2003; Runeson & Frykholm, 1981; Shipley, 2003) (see Chapter 4 for detailed discussions). The perception of local biological motion may then be orientation-dependent due to the inherent direction of gravity-defined acceleration. That is, the orientation-specificity observed for perceiving local biological motion may be due to the presence of the invariant in upright, gravitationally-consistent motions that render these motions to be salient. Such information would not be available in inverted motions.

1.3 Towards a better understanding of a local visual filter

The experiments described in this thesis were designed to investigate further the characteristics of the mechanisms proposed to underlie the perception of local biological motion. These experiments (some of which have been published) are presented in the subsequent chapters as independent manuscripts (reprints of the published manuscripts or manuscripts to be submitted for publication).

Motivated by the hypotheses put forth by Troje and Westhoff (2006), the experiments described in Chapter 2 asked whether additional information beyond the facing the direction of a
walker can be retrieved from local biological motion. If the critical information used to solve the
direction discrimination task in the study by Troje and Westhoff is an invariant cue that serves as
a characteristic or signature of animate agents, then the perception of animacy may also be
defined by the availability (orientation) of this information in local biological motion. This
possibility was addressed with a series of experiments in which observers were presented with
coherent and scrambled point-light displays of various animals in both upright and inverted
orientations and with various display durations. The task was to rate each display according to
how animate it appeared. The results showed that overall, coherent walkers were rated to be
more animate than scrambled walkers. Importantly, the impression of animacy from scrambled
displays decreased upon stimulus inversion not only for the coherent displays, but also for the
scrambled displays. The orientation specificity of animacy perception from scrambled displays is
congruent with the orientation dependency observed for discerning direction of motion from these
same displays. In a later experiment, the perception of animacy and direction from scrambled
displays were compared more directly. Here, observers were required to complete both a
direction discrimination task and an animacy rating task with the same set of stimuli. The results
revealed that animacy ratings were correlated with direction discrimination performances.
Moreover, both tasks were not affected by changes in stimulus duration and could be
accomplished if stimuli were shown for just a fraction of the full gait cycle (200 ms). This
particular finding suggests that the two paradigms reflect similar perceptual mechanisms.

In Chapter 3, additional characteristics of the local mechanism for biological motion
perception were examined and contrasted with those of the global, structure-from-motion
mechanism. In this chapter, a proposition is first put forth that the local and global mechanisms
for biological motion perception act analogously to the mechanisms conspec and conlern
proposed by Morton and Johnson (1991) for the development of face perception. Briefly, Morton and Johnson suggested that newborn infants possess an innate mechanism termed *conspec* that responds to the general appearance of a face and serves to guide attention to this important stimulus class. A second mechanism, termed *conlern* then allows the individual to learn further about the stimulus class and extract detailed information (e.g., identity). The local and global mechanisms for biological motion perception are proposed to act in an analogous manner. The local motion signals are suggested to drive an early mechanism that may be evolutionarily old and potentially innate. This mechanism may serve as a general detection system that directs attention to an animate agent in the environment. A second mechanism, suggested to retrieve information as to the structure of the agent, may aid in further extracting characteristics such as its identity, actions, and intentions.

Several predictions can be made from the natures of the local and global mechanisms for biological motion perception as proposed above. If the local mechanism is in fact hard-wired or innate, it may be insensitive to the type of agent being displayed and may be insensitive to learning effects. By contrast, if the global mechanism reflects an acquired process, it may be affected by the familiarity of the agent and may change with training. Additionally, as the local mechanism has been previously proposed to be pre-attentive in nature (Troje, 2008), it may be less susceptible to noise (masking) as compared to the global mechanism. The aforementioned predictions were tested with two psychophysical tasks: a walker detection task that addressed global structure-from-motion processing, and a direction discrimination task designed to address primarily local motion processing. In both tasks, performances changes according to agent type (human vs. non-human), various mask densities, and across time (learning) were investigated. The results were largely consistent with the hypotheses. Performance on the walker detection
task designed to address global structure-from-motion processing was best for the human walker displays, improved with training, and was heavily hindered by increases in masking noise. In contrast, performance on the direction discrimination task designed to address local motion processing was unaffected by walker type, did not improve with training, and was relatively robust to masking noise.

The data presented in Chapter 3 suggest that the retrieval of global, structure-from-motion and local motion information are governed by distinct mechanisms with different properties. Still, the exact information being used by the local motion mechanism is unclear. What are the cues that are being exploited from the foot’s motion and what carries the inversion effect? Earlier suggestions by Troje and Westhoff (2006) that the critical information may be some gravity-defined invariant imply that the crucial information is vertical acceleration. The experiments reported in Chapter 4 provide strong evidence to suggest that this may indeed be the case. As noted earlier, the initial experiments (Chapter 1) revealed that the perception of both animacy and direction from local biological motion can be achieved with display durations that constitute only small fractions of the duration of the full gait cycle of a (human) walker. As such, the experiments of Chapter 4 isolated the trajectory of the foot and asked first if all fragments of the gait cycle are comprised of equally salient directional information and contribute equally to the inversion effect. In the first experiment, novel stimuli that were derived from single fragments of the trajectory or two fragments sampled at counterphase positions of the gait cycle (180 deg apart) were presented to observers in a direction discrimination task. The data obtained for these two types of stimuli were very different. The results revealed that the stimuli permitted the recognition of the walker’s facing direction and carried an inversion effect only if derived from the paired fragments. Specifically, when presented with stimuli derived from only single
fragments of the trajectory, observers merely based direction judgments upon the horizontal displacement of the individual dots. Certainly when averaged across the entire gait cycle, the stimuli contained no net displacement. Accordingly, performances for these stimuli were at chance-level. By contrast, the results for the paired stimuli indicated above chance-level performance and a strong inversion effect. These results suggest that the retrieval of direction from local biological motion involves more than processing the spatiotemporal cues inherent to the foot’s motion – rather, the foot’s elemental cues must be evaluated with respect to some reference elements or at the very least, to other parts of the same element’s trajectory. These data are interpreted to indicate a switch from the perception of solely extrinsic motion in the case of the single fragment stimuli, to the perception of intrinsic motion in the case of the paired stimuli – motion conveyed by the deformation of the walker’s body despite being presented as walking on a treadmill.

In a later experiment in the same chapter, the relevance of acceleration for the perception of local biological motion was investigated directly. Observers were presented with stimuli derived from paired fragments of the naturally accelerating human walker foot and with stimuli that were identical to the natural stimuli but had acceleration removed along the individual trajectories. The results revealed that the inversion effect held only for stimuli that retained the natural acceleration. These findings are congruent with previous literature demonstrating the importance of velocity profiles for the perception of animate and inanimate events (e.g., Bingham, Schmidt, & Rosenblum, 1995).

The experiments of Chapter 4 revealed the importance of acceleration information for the perception of local biological motion. Implications from these findings include the possibility that biological motion sensitivity may consequently vary according to the size of the walker – if
acceleration sensitivity is not size invariant. Chapter 5 subsequently documents an experiment that investigated acceleration sensitivity to a non-biological, linear-accelerating stimulus of varying sizes and average velocities. In this experiment, observers were asked to discriminate an accelerated stimulus from a constant velocity stimulus equated for mean velocity and size, and thresholds, defined as the minimum acceleration discriminated at the 82% correct level, were measured using a staircase procedure. The results revealed that acceleration thresholds scale with mean velocity and assumes Weber-like behaviour, as predicted by size invariance, and are inversely proportional to stimulus duration.

The mechanism proposed to underlie the retrieval of information from local biological motion was further characterized in Chapter 6 by means of an experiment that aimed to search for the local biological motion to which the relevant visual filter is optimally tuned. Here, the adequate stimulus was defined as one that carries maximal directional information when shown upright and a pronounced inversion effect. A genetic algorithm that models the natural process associated with the biological process of evolution was used to drive two separate rounds of psychophysical searches, each spanning a multidimensional search space consisting of parameters (e.g., amplitudes and phases) defining a second-order representation of a “foot’s” motion in a two-dimensional image plane. Conducted via a web-based interface, observers were required to make direction judgments for a particular set of “foot” motions. The “fitness” of each stimulus, computed according to performance rates, determined the probability of which it produced descendent children stimuli for the subsequent generation. Parent stimuli, once selected, were then morphed to produce descendant stimuli that varied the parent’s motion characteristics. In light of findings described in the earlier chapters, it was hypothesized that the optimization of psychophysical performance rates should be matched by the evolution of stimuli towards carrying
greater acceleration information. The results indicated that the genetic algorithm was effective at optimizing behavioural performances on the direction discrimination task. The changes in behavioural responses were matched by the evolution of the stimuli across generations towards exhibiting greater amplitude and acceleration along both the horizontal and vertical components. Notably, the evolution of vertical acceleration was described by a change in the distribution of acceleration within a region defined analogously to the swing phase of the gait cycle, towards carrying greater positive acceleration at the beginning of this region and greater negative acceleration before the “foot” contacts the ground – a pattern that may define the animacy of an object. An asymmetry in the proportion of this region carrying positive versus negative acceleration may also account for the inversion effect. The data described in this chapter further support the notion of a visual mechanism that is tuned to gravity-defined invariants present in local biological motion.

Finally, the experiments described to this point exploited as a tool for investigations, the orientation-specific nature of biological motion perception. Yet, it is unclear as to which reference systems biological motion stimuli are encoded in. That is, what does “upright” mean? Consider that the orientation of a stimulus can be described according to a variety of reference systems. Possible reference frames include egocentric components such as the retina, head, or the rest of the body and allocentric reference frames such as the visual-environment or gravity. Previous reports suggest that the reference frame that is most relevant depends on the particular stimulus class involved.

In his early work, Rock (1956) investigated reference frames for the perception of ambiguous figures designed such that the tops of two versions differed by 90 deg. His observers viewed the figures at various head tilts. The results indicated that the reported percept depended
on the visual-environmental upright. In another experiment, observers tilted 90 deg from visual-environmental upright were presented with fragmented figures that were environmentally upright. Rock noted that observers were still able to recognize the figures based upon environmental direction cues despite the 90 deg discrepancy with the egocentric frame of reference (Rock, 1973).

The egocentric reference frame has been shown to be dominant for the perception of shape from shading, although there are additional contributions from gravitational and visual-environmental frames of reference as well (e.g., Yonas, Kuskowski, & Sternfels, 1979; Howard, Bergstrom, & Ohmi, 1990; Jenkin, Dyde, Jenkin, Howard, & Harris, 2003). Other studies have reported gravity-centered representations for reversible figures (Yamamoto & Yamamoto, 2006) and retinocentric representations for natural scenes (Gaunet & Berthoz, 2000). As for shape-from-shading judgments, both gravity- and body-centered representations have been reported for the visual vertical (Mittelstaedt, 1992) and character recognition (Dyde, Jenkin, & Harris, 2006).

Very little is known as to the reference frame in which biological motion is coded (Bingham et al., 1995; Shipley, 2003). As noted, observers have been shown to use implicit knowledge about the effects of gravity when interpreting animate motion patterns (e.g., Runeson & Frykholm, 1981). Making assumptions with regards to the direction of gravity, however, does not necessarily implicate direct measurements of gravitational acceleration (e.g., via input from the vestibular system). The visual system may simply take advantage of the fact that gravity is typically aligned with egocentric coordinates.

The experiments reported in Chapter 7 investigated reference frames for the perception of biological motion as compared to faces. In these experiments, the relevance of the egocentric, gravitational, and visual-environmental reference coordinates for biological motion and face
perception was teased apart by testing observers on behavioural tasks inside the York “tumbling room” – a facility that permits the observer and the visual-environment to be oriented independently with respect to gravity. The relative contribution of each reference frame was assessed by arranging pairs of frames to be either aligned or opposed to each other while rendering the third neutral by orientation it sideways relative to the stimulus. The results revealed that the perception of both biological motion and faces were optimal when the stimulus was aligned with egocentric coordinates. However, when this dominant reference frame was rendered neutral, the perception of biological motion, but not faces, relied more on stimulus alignment with gravity rather than the visual-environment. The results suggest that while biological motion and faces are predominantly egocentrically coded, non-visual measurements of gravity are integrated for the perception of biological motion.

In closing, the findings from the projects in this thesis are discussed in the context of emerging work from other laboratories towards the understanding of the perception of local biological motion in the visual periphery, the role of attention, and current understanding of the underlying neural concomitants of biological motion perception (Chapter 8).
References


Chapter 2
Perception of Animacy and Direction

2.1 Introduction

The visual mechanisms that underlie the perception of movement patterns characteristic of living animals, collectively termed *biological motion* (Johansson, 1973), are of particular interest given the social implications that accompany identifying and interpreting such patterns. It has been shown that from simple displays of point-lights depicting typically the motions of the joints of an agent, characteristics such as gender (e.g., Kozlowski & Cutting, 1977; Barclay, Cutting, & Kozlowski, 1978), emotion (e.g., Dittrich, Troscianko, Lea, & Morgan, 1996; Pollick, Paterson, Bruderlin, & Sanford, 2001), and identity (Cutting & Kozlowski, 1977) can be extracted reliably. In addition, this perceptual ability is not restricted to human patterns as it has been extended to point-light animations depicting the motions of other legged animals (Mather & West, 1993).

The exact nature of the mechanisms that allow the retrieval of such a wide range of information from biological motion however, is not well agreed upon. While some believe that information processing relies foremost on local motion signals (e.g., Mather, Radford, & West, 1992), others emphasize the importance of global processes, relating to the spatiotemporal

---

organization of the display (e.g., Beintema & Lappe, 2002; Bertenthal & Pinto, 1994; Heptulla Chatterjee, Freyd, & Shiffrar, 1996; Shiffrar, Lichtey, & Heptulla Chatterjee, 1997).

A role for local processes has been demonstrated empirically. Consistent with the restricted temporal and spatial properties of local motion detecting processes (e.g., Baker & Braddick, 1985), Mather, Radford, and West (1992) found that performances on coherence and direction discrimination tasks were reliable only when the inter-frame interval and frame increments of the point-light stimulus were short (see however, Thornton, Pinto, & Shiffrar, 1998, for evidence that mechanisms supporting biological motion perception can operate over extended temporal intervals).

Still, others have demonstrated the relevance of global analyses for biological motion perception. Beintema and Lappe (2002) tested the perception of a moving human figure in the absence of local image motion by re-allocating each point to another randomly selected position on the limb in successive frames and showed that under these conditions, observers still spontaneously reported seeing walking persons. Correspondingly, Bertenthal and Pinto (1994) embedded a target in a mask of additional, randomly-positioned walker point-lights and found that observers could still detect a walking figure. These studies suggest that the perception of structure from a point-light display does not in fact require prior processing of local motions. In addition, the orientation-dependency of biological motion perception, a well-documented characteristic describing an impairment in perceiving inverted point-light displays that otherwise maintain the local relations of their upright versions (i.e., the “inversion effect”), has been often interpreted in favor of global or configural processing (e.g., Bertenthal & Pinto, 1994).

Although proponents for local or global processing argue for very different mechanisms, they share in common the assumption that biological motion perception is a unitary phenomenon.
It has recently become apparent however, that this approach should be abandoned in favor of a multi-level view that could well accommodate the contributions of both local and global information (Troje, 2008). In addition to the inversion-related impairment in biological motion perception that may well be attributable to global form processing, Troje and Westhoff (2006) demonstrated an inversion effect that is associated solely with local motion cues. In their study, coherent and spatially scrambled point-light displays were presented to observers performing a direction discrimination task. While coherent displays contain structural information that can be exploited, scrambled displays in which individual dots are randomly displaced from their veridical positions contain only local information. Significantly, they showed that even in the absence of structural cues, observers could reliably extract directional information from the scrambled displays. However, the ability to discriminate direction from these displays was significantly impaired when the scrambled walkers were shown inverted rather than upright. This effect was independent of the temporal relationships among the individual dots. Further examination by inverting specific parts of the display revealed that the cues for direction of motion were carried by the ankle dots of the walker.

The importance of the ankle motion for biological motion perception has been reported previously by Mather et al. (1992). In their study, point-light walkers with certain dots omitted (shoulder and hip, elbow and knee, or wrist and ankle) were presented to observers. They showed that performances on coherence and direction discrimination tasks were most affected by the omission of wrist and ankle dots. The findings by Troje and Westhoff (2006) that the ankle dots retain their cues to direction even in scrambled displays and that they carry an orientation-specific effect contribute important extensions to the findings by Mather et al. Particularly, their findings suggest that the inversion effect observed for coherent walkers has two entirely distinct
causes. While there seems to be an inversion effect which is due to the configuration or global shape of the walker, there is also a second one, which operates on the local motion of the ankles. Troje and Westhoff suggested that this second inversion effect reflects a visual mechanism that constitutes a general detection system that signals the presence of terrestrial, articulated animals in the visual environment.

If the visual invariants that signal direction in upright, scrambled biological motion displays are also used to signal the presence of an animal in the visual field, then the presence or absence of these invariants should influence the perceived animacy of the display. Here, we tested this hypothesis by exploiting the orientation-specific characteristic of perceiving scrambled point-light displays in three experiments that investigated the perception of both animacy and direction from the same set of point-light stimuli. In Experiment 1, we investigated if scrambled displays that are oriented upright are perceived as more animate than scrambled displays that are inverted, as our hypothesis predicts. As we show that animacy can be perceived within exposure times as short as 200 ms, we replicated the experiment conducted by Troje and Westhoff (2006) with short exposure times and asked if the ability to discriminate walking direction from our stimuli is similarly robust in Experiment 2. Finally, in Experiment 3, we investigated if the animacy percept induced from our stimuli is related to the ability to discriminate their direction of walking.

2.2 Experiment 1

The ability to discriminate direction from spatially scrambled point-light displays is orientation-specific (Troje & Westhoff, 2006). In this first experiment, we asked if the perception of animacy from spatially scrambled displays is similarly orientation-specific.
2.2.1 Methods

Participants

Twelve naïve observers that ranged in age from 17 – 23 years (mean age of 19.25 years; 5 males, 7 females) participated in this experiment. All observers had normal or corrected-to-normal vision.

Stimuli and apparatus

Sample frames of the stimuli are presented in Figure 2.1. The stimuli were derived from point-light sequences of a walking human, cat, and pigeon. The human walker, computed as the average walker from motion-captured data of 50 men and 50 women (Troje, 2002), was depicted by a set of 11 markers used to represent the motions of the head, one shoulder, one hip, two elbows, two wrists, two knees, and two ankles. The cat sequence was constructed by sampling 14 points from single frames of a video sequence showing a cat walking on a treadmill. Finally, the point-light pigeon sequence was created from motion-captured data of a pigeon fitted with 11 markers. For all sequences, the translating component of the walk was removed such that the animals displayed stationary walking. All walkers were presented in sagittal view (i.e., facing rightwards or leftwards) and were shown at their veridical speeds with gait frequencies of 0.93 Hz, 1.7 Hz, and 1.6 Hz for the human, cat, and pigeon, respectively. For each presentation, the starting position of the walker within the gait cycle was selected randomly. The point-light displays were presented upright or inverted about the horizontal axis, and were coherent (all points maintained original organization), or spatially scrambled (local trajectory of each point remained intact but was displaced randomly within areas matched to those occupied by the corresponding coherent versions).
The stimuli were generated using MATLAB (Mathworks, Natick, MA) with extensions from the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997), and were displayed on a 22 inch ViewSonic P220f CRT color monitor with 0.25 mm dot pitch, 1280 x 1024 pixels spatial resolution, and 100 Hz frame rate. All stimuli appeared as white dots on a black background and the full point-light figures subtended visual angles of 2.1 x 4.6 deg, 4.6 x 2.4 deg and 3.6 x 3.6 deg for the human, cat, and pigeon respectively. In addition, all walkers were embedded in a 6.4 x 6.4 deg mask that consisted of 125 randomly positioned stationary dots with a limited lifetime of 125 ms.

![Figure 2.1](image)

Figure 2.1. Static frames from right-walking sequences of a coherent human, cat, and pigeon (left panel) and their corresponding scrambled versions (right panel). In the scrambled displays, the local trajectories are intact but spatially displaced.

**Procedure**

Stimuli were viewed binocularly at a distance of 80 cm as maintained by a chin-rest. A six-point Likert-type scale was used to measure animacy judgments. On each trial, the observer's
task was to make a confidence indication of whether the stimulus displayed was a “living being” by selecting one of six boxes presented on the computer screen with a mouse click. The six boxes were labeled from 0 – 5, where 0 represented “definitely not a living being” and 5 represented “definitely a living being”.

Participants were first instructed on the task both verbally and by printed instructions on the computer screen. These instructions revealed only that a set of moving dots would be displayed on each trial, upon which a response decision is to be based. A practice block of trials was then presented during which participants familiarized themselves with the task. Here, a total of 48 trials were presented: 24 unmasked walkers consisting of all possible combinations of the three animal types (human, cat, and pigeon), two organizations (coherent and scrambled), two display orientations (upright and inverted), and two walking directions (left and right), and 24 masked walkers of these same combinations. For each trial in the practice block, the stimulus was displayed for 1000 ms, after which the six response boxes were presented across the screen until a response was given.

After the practice block, participants completed the experiment proper which consisted of three experimental blocks that differed only with respect to stimulus duration (200 ms, 500 ms, and 1000 ms). The order in which the blocks were tested was counterbalanced among the participants. Within each experimental block, the 24 possible masked stimuli, consisting of all combinations of animal type, organization, orientation, and walking direction were repeated five times. Therefore, 120 trials were presented in each block (the order of which was counterbalanced for the 24 possible stimuli and randomized), resulting in a total of 360 trials for the experiment proper. The resulting design was a within-subject test of factors animal type, organization, orientation, and stimulus duration.
2.2.2 Results

Mean ratings were evaluated with a 3 (animal type) x 2 (organization) x 2 (orientation) x 3 (stimulus duration) repeated-measures analysis of variance (ANOVA) that revealed a significant main effect of organization, $F(1, 11) = 83.82, p < 0.001$, a significant main effect of orientation, $F(1, 11) = 103.25, p < 0.001$, a significant organization x orientation interaction, $F(1, 11) = 13.73, p = 0.007$, but no effect of animal type, $F(2, 22) = 1.51, p = 0.243$, or stimulus duration, $F(2, 22) = 0.97, p = 0.394$. All other interactions were not significant.

An examination of group means indicated that coherent stimuli (mean = 4.03) were rated more animate than scrambled stimuli (mean = 1.98), and upright stimuli (mean = 3.49) were rated more animate than inverted stimuli (mean = 2.51). Figure 2.2 shows the mean ratings for coherent and scrambled stimuli in both upright and inverted conditions. Specifically, mean ratings were 4.69, 3.36, 2.29, and 1.67 for the coherent/upright, coherent/inverted, scrambled/upright, and scrambled/inverted conditions, respectively. Initial Tukey’s post-hoc comparisons crossing the two organizations and the two orientations failed to reveal the source of the interaction. That is, upright stimuli were rated more animate than inverted stimuli in both coherent and scrambled conditions, and coherent stimuli were rated more animate than scrambled stimuli in both upright and inverted conditions ($p < 0.001$ for all). A subsequent analysis was performed on difference scores obtained for each individual by subtracting the mean ratings of inverted stimuli from those of upright stimuli for the coherent and scrambled conditions while collapsing across all other conditions. Thus, for each individual, a measure of the changes in animacy ratings associated with inversion was obtained for both coherent and scrambled conditions. The scores were entered in a paired $t$ test that showed that the mean difference between upright and inverted stimuli was larger for coherent than for scrambled stimuli ($p =$...
0.007). Specifically, the mean differences were 1.33 and 0.62 for coherent and scrambled stimuli, respectively.

![Figure 2.2](image.png)

Figure 2.2. A comparison of mean animacy ratings for coherent and scrambled displays between the upright and inverted orientations. Error bars represent ± 1 standard error of the mean.

### 2.2.3 Discussion

The higher animacy ratings attributed to coherent stimuli as opposed to scrambled stimuli is not surprising as the perception of animate entities surely depends on information beyond just kinematics, including the ability to detect recognizable form. In addition, the decrease in animacy ratings at the inverted orientation is consistent with previous studies that have reported degraded recognition or perceived quality of biological motion patterns upon inverting stimulus displays (Pavlova, 1989; Pavlova & Sokolov, 2000; Sumi, 1984). Significantly, we show here that the perception of animacy decreases upon inversion not only for coherent displays, but also for scrambled displays that contain solely local cues. The inversion effect associated with scrambled stimuli cannot be attributed to impaired global, configural information processing. In
In this respect, the orientation-specificity of animacy perception from scrambled displays is congruent with the orientation-dependency for discerning direction of motion from scrambled displays documented by Troje and Westhoff (2006).

The lack of effect of animal type suggests that the responsible mechanisms are tuned to some invariant characteristic(s) present in the locomotion patterns of all three animals used in this experiment, and possibly in all terrestrial, legged animals. The lack of effect of stimulus duration suggests further that these mechanisms are remarkably robust and are able to convey information about animacy efficiently at very limited exposure times. Is the retrieval of directional information from scrambled displays similarly robust? In Experiment 2, we maintained the design and parameters of this first experiment but replaced the animacy rating task with a direction discrimination task in order to assess discrimination performance under identical conditions. It is important to note that the stationary walking stimuli do not contain any extrinsic motion (that is, a translatory component referred to as “common motion” by Johansson, 1973). The cues that imply a particular direction are in contrast entirely object-centered and therefore intrinsic. Indeed, this paradigm has been used in numerous studies of biological motion perception in the past (e.g., Neri, Morrone, & Burr, 1998; Beintema & Lappe, 2002; Thornton, Rensink & Shiffrar, 2002).

2.3 Experiment 2

2.3.1 Methods

Participants

Twelve observers that ranged in age from 18 – 27 years (mean age of 19.83 years; 6 males, 6 females) participated in this experiment. All observers had normal or corrected-to-
normal vision. In addition, all observers were naïve to the task and had not previously participated in Experiment 1.

**Stimuli and apparatus**

The stimuli and apparatus were identical to those described for Experiment 1.

**Procedure**

Here, a two-alternative, forced-choice direction discrimination paradigm was used whereby the observer’s task was to indicate whether the stimulus appeared to be moving leftwards or rightwards by pressing one of two arrow keys on the keyboard. Feedback was not given for correct/incorrect responses. All other design parameters and the testing procedures were identical to those described for Experiment 1.

**2.3.2 Results**

Direction discrimination accuracies, measured in terms of proportions of correct responses, were analyzed with a 3 (animal type) x 2 (organization) x 2 (orientation) x 3 (stimulus duration) repeated-measures ANOVA that showed a significant main effect of organization, $F(1, 11) = 143.29, p < 0.001$, a significant main effect of orientation, $F(1, 11) = 126.72, p < 0.001$, and a significant organization x orientation interaction, $F(1, 11) = 13.53, p = 0.004$. There was no effect of animal type, $F(2, 22) = 1.22, p = 0.314$, or stimulus duration, $F(2, 22) = 1.20, p = 0.318$, and all other interactions were not significant.

A comparison of means for the main effects revealed that performance was higher for coherent (mean = 0.79) versus scrambled (mean = 0.58) stimuli, and for upright (mean = 0.75) versus inverted (mean = 0.61) stimuli. The mean discrimination accuracies for coherent and scrambled stimuli in both upright and inverted conditions are displayed in Figure 2.3. The proportions of correct responses were 0.87, 0.70, 0.64, and 0.52 for the coherent/upright,
coherent/inverted, scrambled/upright, and scrambled/inverted conditions, respectively. As for Experiment 1, the organization x orientation interaction was first analyzed with Tukey’s post-hoc comparisons. These comparisons failed to reveal the source of the interaction, indicating that discrimination accuracies were higher for coherent than for scrambled stimuli in both upright and inverted conditions and higher for upright than for inverted stimuli in both coherent and scrambled conditions ($p < 0.001$ for all). As a result, difference scores were obtained for each individual by subtracting accuracies of inverted stimuli from those of upright stimuli, in coherent and scrambled conditions while collapsing across all other factors. The mean difference was 0.17 for coherent stimuli and 0.12 for scrambled stimuli. A paired $t$ test used to analyze these scores indicated that the reduction in discrimination accuracies due to inversion was higher for coherent than for scrambled stimuli ($p = 0.004$).

![Figure 2.3](image.png)

Figure 2.3. Mean direction discrimination accuracies, expressed as the proportions of correct responses for coherent and scrambled displays at upright and inverted orientations. Error bars represent ± 1 standard error of the mean.
2.3.3 Discussion

The results indicate that under identical conditions to Experiment 1, the ability to discriminate direction from scrambled displays, as from coherent displays, is orientation-specific. The present results are consistent with those reported by Troje and Westhoff (2006) but more importantly, show further that direction retrieval is highly robust, as demonstrated by the stable performance across the stimulus durations tested – including one duration (200 ms) that corresponds to only a small fraction of the full gait cycle duration of all three animals.

The pattern of results obtained in this experiment for the perception of direction is strikingly similar to the pattern of results obtained in Experiment 1 where we obtained instead measurements of perceived animacy. For example, animacy ratings and direction discrimination accuracies decreased upon inversion both more substantially for coherent displays than for scrambled displays. In addition, both measurements were stable across different animal types and changing stimulus durations. It is worth asking then if the two paradigms can be used interchangeably in experiments investigating the cues contained in local biological motions. That is, is the ability to retrieve walking direction from the present stimuli related to the animacy percept induced by these same stimuli? In Experiment 3, we probed this possibility by testing a new group of naïve observers on both the animacy rating and direction discrimination paradigms in order to permit within-subject comparisons. It should be noted that while this approach exploits inter-individual variability, it allows observers to carry strategies from one task to the other. Such carry-over effects were controlled for in Experiments 1 and 2.

2.4 Experiment 3

2.4.1 Methods
Participants
A new group of 16 naïve observers that ranged in age from 18 – 24 years (mean age of 19.24 years; 7 males, 9 females) were tested in this experiment. All observers had normal or corrected-to-normal vision.

Stimuli and apparatus
The stimuli and apparatus were identical to those described for Experiments 1 and 2.

Procedure
In this experiment, all participants completed both the Likert-type animacy rating task (see Experiment 1) and the 2AFC direction discrimination task (see Experiment 2). The order of testing was fixed such that the rating task always preceded the discrimination task.

For both tasks, all stimuli were presented with a stimulus duration of 500 ms. Participants completed two blocks of experimental trials (one block per task) and each block comprised of 360 trials, consisting of 15 replications of the 24 possible stimuli that were counterbalanced and randomized. All other design parameters and procedures were identical to those described for Experiments 1 and 2.

2.4.2 Results
Animacy rating
Animacy ratings were analyzed with a 3 (animal type) x 2 (organization) x 2 (orientation) repeated-measures ANOVA. The analysis showed significant main effects of organization, $F(1, 15) = 106.65, p < 0.001$, and orientation, $F(1, 15) = 65.42, p < 0.001$, and a significant organization x orientation interaction, $F(1, 15) = 6.30, p = 0.024$. There was no effect of animal type, $F(2, 30) = 0.20, p = 0.980$. 
An examination of group means indicated that coherent stimuli (mean = 3.38) were rated more animate than scrambled stimuli (mean = 1.83), and upright stimuli (mean = 2.98) were rated more animate than inverted stimuli (mean = 2.22). The mean ratings for coherent and scrambled stimuli in both upright and inverted orientations are conveyed by the asterisks in Figure 2.4. Specifically, mean ratings were 3.87, 2.89, 2.10, and 1.55 for the coherent/upright, coherent/inverted, scrambled/upright, and scrambled/inverted conditions, respectively. Tukey’s post-hoc comparisons indicated that upright stimuli were rated more animate than inverted stimuli in both coherent and scrambled conditions and coherent stimuli were rated more animate than scrambled stimuli in both upright and inverted conditions ($p < 0.001$ for all). Difference scores obtained for each individual by subtracting the mean ratings of inverted stimuli from those of upright stimuli for the coherent and scrambled conditions were entered in a paired $t$ test that showed that the mean difference between upright and inverted stimuli was larger for coherent than for scrambled stimuli ($p = 0.024$). Here, the mean differences were 0.98 and 0.55 for coherent and scrambled stimuli, respectively.

**Direction discrimination**

A comparable analysis of discrimination accuracies showed significant main effects of organization, $F(1, 15) = 185.03, p < 0.001$, and orientation, $F(1, 15) = 43.61, p < 0.001$, a significant organization x orientation interaction, $F(1, 15) = 5.54, p = 0.033$, but no effect of animal type, $F(2, 30) = 1.50, p = 0.240$.

A comparison of means indicated that performance was higher for coherent (mean = 0.78) versus scrambled (mean = 0.58) stimuli, and for upright (mean = 0.75) versus inverted (mean = 0.61) stimuli. The mean discrimination accuracies for coherent and scrambled stimuli in both upright and inverted conditions are also conveyed by the asterisks in Figure 2.4.
Specifically, the proportions of correct responses were 0.87, 0.69, 0.63, and 0.52 for the coherent/upright, coherent/inverted, scrambled/upright, and scrambled/inverted conditions, respectively. Tukey’s post-hoc comparisons indicated that discrimination accuracies were higher for coherent than for scrambled stimuli in both upright and inverted conditions and higher for upright than for inverted stimuli in both coherent and scrambled conditions ($p < 0.001$ for all). A subsequent paired $t$ test based upon difference scores obtained for each subject by subtracting accuracies of inverted stimuli from those of upright stimuli within coherent and scrambled conditions indicated that inversion impaired performance more substantially for coherent than for scrambled stimuli ($p = 0.033$). Specifically, the mean difference was 0.18 for coherent stimuli and 0.11 for scrambled stimuli.

Figure 2.4. Correlation between animacy rating and discrimination accuracy (expressed as the proportion of correct responses) across individual participants for four conditions that result from crossing the two stimulus organizations and two orientations. Individual data points are overlaid with their respective linear regression lines. Asterisks superimposed on the regression lines convey the corresponding mean ratings and mean accuracies.
Cross-task analyses

Data from this experiment were analyzed further by comparing animacy ratings with respect to discrimination accuracies by means of simple linear regressions. Overall, a significant correlation was found between animacy ratings and discrimination accuracies in the positive direction ($r^2 = .84$; $p < 0.001$). According to a 2 (organization) x 2 (orientation) factorial model, four separate linear fits were imposed on the data (Figure 2.4, overlaid). The analyses indicated that the positive correlation between animacy rating and discrimination accuracy was significant for conditions of coherent/upright ($r^2 = .40$; $p = 0.008$), coherent/inverted ($r^2 = .43$; $p = 0.006$), and scrambled/upright ($r^2 = .43$; $p = 0.006$), but not scrambled/inverted ($r^2 = .09$; $p = 0.25$).

Further comparisons of these linear fits with a two-way ANOVA on the regression slopes showed no effects of organization and orientation, and no interaction ($p > 0.069$ for all). Similarly, a comparable analysis for regression intercepts revealed no effects of organization and orientation, and no interaction ($p > 0.329$ for all).

Finally, a linear model fitted to the current data that describes animacy rating as predicted from discrimination performance within the factorial arrangement is represented by the following equation:

$$r = 4.56d + 0.32a + 0.06b - 0.49$$ (2.1)

Here, animacy rating, $r$, can be predicted from discrimination performance, $d$, conditional upon stimulus organization and orientation as represented by binary variables $a$ and $b$, respectively. The model indicates that animacy rating can be almost entirely predicted by direction discrimination performance. The small size of the coefficients $a$ and $b$ show that the independent effects of stimulus organization and orientation are comparatively minor.
2.4.3 Discussion

The results of this experiment replicated those of Experiments 1 and 2 suggesting that the observers’ strategies for the second task were independent of prior experience with the first task. Importantly, we found here a strong linear relationship between the ability to discriminate direction and the animacy percept induced by our stimuli that was stable across conditions. That is, it appears that those observers who attributed a strong percept of animacy to a particular stimulus were well able to discern the apparent direction of motion of this same stimulus. Conversely, those observers who attributed a weak percept of animacy to a given stimulus were less able to discern the direction of that stimulus. We therefore conjecture that within our context, the two paradigms elucidate similar perceptual strategies and thus, address similar visual mechanisms. It is possible that the correlation in our data can be explained if the two tasks are not perceptually independent; that is, observers may have been able to discriminate walking direction only if an animate agent was first recognized (or more unlikely, vice versa). Nonetheless, it is clear that the manner in which both tasks were solved depended on the orientation of the relevant cues. In consideration of the findings of Troje and Westhoff (2006), we propose that the relevant local cues exploited for scrambled stimuli in the two tasks were contained in the motion of the limbs. However, as we did not isolate specifically the limb motion in the three experiments, we must also acknowledge the possibility that observers at least when asked to attribute animacy, did not rely on the ankle trajectories but used instead other arbitrary local cues. This explanation is unlikely however, given the similarity in the pattern of results obtained from the two paradigms.
2.5 General Discussion

We showed that despite lacking coherent form information, scrambled point-light displays still elicit a significant degree of perceived animacy. However, the perception of animacy from scrambled displays is orientation-specific (Experiment 1). The decrease in the perceived animacy of inverted scrambled stimuli is analogous to the impaired ability to discriminate walking direction from these same displays (Experiment 2). By exploiting inter-individual variability, we showed also a substantial correlation between measures of perceived animacy and direction from our stimuli (Experiment 3). Together, these results provide novel insights into the properties of the perceptual mechanisms responsible for processing local biological motion signals.

Of particular interest is the orientation-specific nature of our data for scrambled stimuli. As these stimuli do not carry coherent structural information, the associated inversion effect cannot be explained by global mechanisms (e.g., template-matching) and rather must be attributed to the processing of local motions. Significantly, the local ankle motions have been shown to be particularly important in the perception of biological motion (Mather et al., 1992; Troje & Westhoff, 2006). Electromyographic data have indicated that leg muscles are almost inactive during the swing phase (e.g., Crowninshield & Brand, 1981). As such, it has been proposed that the foot, after being set into motion by muscles during the stance phase, moves entirely under the influence of gravity (Mochon & McMahon, 1980) – a locomotive strategy that minimizes energetic cost (Nakamura, Mori, & Nishii, 2004). The visual system may be particularly sensitive to the motion that results as the foot interacts with gravitational force – a unique cue that may signal the presence of other terrestrial, legged animals. Indeed, an examination of the ankle dot’s motion during the swing phase reveals that it exhibits large
changes in velocity along the vertical axis, presumably due to gravitational influence. The orientation of the ankle dot dictates the manner of velocity change. We speculate that the visual system interprets velocity changes with respect to gravitational acceleration. Inverted displays exhibit the same velocity changes as their upright versions, but in a manner that is not congruent with the direction of gravity. As such, the inverted displays may be less salient for information retrieval than the upright displays, which contain motions that are deemed plausible (i.e., gravitationally-consistent).

A role for gravity for the interpretation of biological motion is supported by Runeson and Frykholm (1981), who showed that from point-light sequences of an actor lifting and carrying a box, observers were readily able to judge the box’s mass. They concluded that the mass of the box was cued by the relationship between postural changes and the motion of the box – a relationship that can be exploited only with an understanding of gravitational force. In another study, Jokisch and Troje (2003) presented point-light displays of dogs walking with varying frequencies and showed that observers judged the size of the dog to be smaller at high stride frequencies than at low frequencies. They suggested that the visual system uses implicit knowledge of a fixed relation between spatial and temporal parameters that exists in a gravity-based environment in order to retrieve size information from the point-light gait patterns. The use of gravitation-based heuristics for interpreting biological motion is further corroborated by Shipley (2003) who presented a point-light figure walking on his hands in both upright and inverted orientations and found that walker detection was better for the upright display which contained familiar dynamic relations.

Nonetheless, orientation effects are interesting to consider in light of a recent finding that newly hatched chicks possess an orientation bias (Vallortigara & Regolin, 2006). In their study,
the authors showed that visually naïve chicks presented with point-light sequences of a walking hen aligned their bodies with the apparent direction of movement of upright walking hens, but oriented randomly with inverted hens. The apparently innate orientation bias in chicks raises an intriguing possibility that other vertebrates, such as humans, may also have a similar predisposition. Developmental studies have shown that three-month-old infants can discriminate an upright walker from an inverted walker (Bertenthal, Proffitt, & Cutting, 1984; Fox & McDaniel, 1982) but little is known about sensitivity to biological motion patterns in younger infants. An orientation-bias associated with local biological motion signals that may be innate to humans requires further empirical work with human newborns (Johnson, 2006).

The apparent difference in saliency between upright and inverted versions of local signals contained in biological motion sets forth a contrast that should be investigated when exploring possible neural concomitants. Neurophysiological and neuroimaging accounts have often implicated the posterior superior temporal sulcus (STS) in the perception of coherent biological motion (e.g., Bonda, Petrides, Ostry, & Evans, 1996; Grossman et al., 2000; Oram & Perrett, 1994; Vaina, Solomon, Chowdhury, Sinha, & Belliveau, 2001). Interestingly, this region has also been shown to be activated by simple moving objects whose interactions appear causal or intentional, thereby inducing a percept of animacy (Schultz, Friston, O’Doherty, Wolpert, & Frith 2005). Consistent with psychophysical impairments in interpreting coherent biological motion displays that are inverted rather than upright, Grossman and Blake (2001) have shown relatively smaller STS activity in response to inverted displays as compared to upright displays. This reported difference in STS activation however, may reflect only the inversion effect proposed to be associated with global form processing. The neural structures that underlie orientation-dependency in response to local biological motion signals remain to be teased out, although
recent findings have shown that the extrastriate areas V3 and V3A are differentially responsive to upright and inverted versions of scrambled biological motion displays (Jiang & He, 2007).

Finally, a better understanding of the substrates relevant to interpreting local biological motion signals may well have implications that extend to autistic individuals who are characterized by social deficits (Kanner, 1943). The abilities to detect and interpret animate activity are fundamental to social interactions. As such, it is appealing to ask if deficits in social function may be attributed at least in part, to impairments in the neural mechanisms that subserve these perceptual skills. Indeed, it has been shown that autistic individuals are impaired in the recognition of point-light animations depicting human activity (Blake, Turner, Smoski, Pozdol, & Stone, 2003). Blake et al. (2003) interpreted their results in terms of an impaired ability to integrate local motion signals into coherent form. However, we have shown here that information about animacy can also be retrieved from local biological motion signals that cannot be integrated into meaningful form. It is worth considering then, if the mechanisms underlying the perception of local biological cues may too be implicated in such disorders of social function.
References


3.1 Introduction

Visual mechanisms that underlie the perception of biological motion, that is, the motion patterns of animate beings (Johansson, 1973), seem to be in place early in human development. For example, 2-day-old infants prefer to look at a point-light walking hen rather than a display of randomly moving dots. These infants also show a looking preference for point-light hens that are oriented upright rather than inverted (Simion, Regolin, & Bulf, 2008). Infants aged 3 – 4 months can discriminate an upright human movement pattern as depicted by a set of moving point-lights from an inverted version of the same pattern or from a pattern of randomly moving dots (Bertenthal, Proffitt, & Cutting, 1984; Fox & McDaniel, 1982). At 5 months of age, infants can discriminate a point-light walker from one in which the dots’ spatial organization and temporal phase are disrupted (Bertenthal, Proffitt, & Kramer, 1982) and can discriminate left- from right-facing point-light walkers (Kuhlmeier, Troje, & Lee, submitted). Infant sensitivity to biological motion has also been demonstrated electrophysiologically. To this end, Hirai and Haraki (2005) have shown that the amplitudes of event-related potentials (ERPs) are higher for intact than for scrambled point-light animations in 8-month-old infants. At this age, ERP amplitudes are also higher for point-light displays that are shown in the upright rather than in the inverted orientation.

---

(Reid, Hoehl, & Striano, 2006). Considered together, the developmental literature suggests that humans have the capacity to process biological motion within the first year of life and that at least some facilities are present in infants as young as several days. This raises the intriguing possibility that the mechanisms subserving the perception of biological motion may at least in part, be innate rather than acquired.

Innate mechanisms for processing biologically-relevant stimuli have been suggested in research of filial preference behavior of chicks. Newly-hatched, visually naive chicks raised in the dark prefer to approach objects resembling adult conspecifics (i.e., stuffed hens) rather than less naturalistic objects (Johnson, Bolhuis, & Horn, 1985). Newly hatched chicks also respond to point-light animations and prefer to approach displays of intact or scrambled walking hens rather than displays of a rigidly rotating hen or of randomly moving dots (Vallortigara, Regolin, & Marconato, 2005). Furthermore, newly hatched chicks prefer to align with a display of an intact, upright walking hen rather than one in which the intact walker is inverted (Vallortigara & Regolin, 2006). Together, these results suggest that preference and orienting behaviors of newborn chicks are guided by visual invariants that are likely to emanate from their mother.

Morton and Johnson (1991) proposed that the epigenetic mechanisms underlying the development of face perception in infants consist of two systems. They suggested that newborn infants possess an innate, ancestrally acquired mechanism termed conspec which contains some knowledge as to the general visual characteristics of conspecific faces. An innate predisposition for attending to faces is supported by findings that newborn infants prefer to track a moving schematic face rather than a scrambled face or a blank head outline (Goren, Sarty, & Wu, 1975; Maurer & Young, 1983). Importantly, this mechanism is presumed to guide attention towards faces ensuring that the developing visual system receives ample exposure to a stimulus class that
it must learn so much about. A proposed second mechanism termed *conlern*, is responsible for exactly that: learning the sophisticated details that convey specific information about identity, emotion, and other aspects of human faces.

It is conceivable that both innate and acquired visual mechanisms may also contribute to the perception of animate motion patterns in adult humans. This matter has received little consideration in the literature, however, perhaps because of the tendency to view biological motion perception as a single-mechanism phenomenon. Biological motion perception has generally been regarded as either a local phenomenon, relying foremost on motion signals of the individual dots (e.g., Mather, Radford, & West, 1992) or a global, form-from-motion phenomenon, relying predominantly on the display’s spatiotemporal organization (e.g., Beintema & Lappe, 2002; Bertenthal & Pinto, 1994; Heptulla Chatterjee, Freyd, & Shiffrar, 1996; Shiffrar, Lichtey, & Heptulla Chatterjee, 1997). Troje (2008) suggested that biological motion perception should instead be regarded as a multi-level phenomenon allowing for distinct contributions of a mechanism that exploits the local motion of the individual dots or body parts, and a second mechanism that uses motion to retrieve the global, dynamically changing shape of a body in motion. This view is supported by the finding of at least two distinct sources for the characteristic inversion effect in biological motion perception; that is, the impairment in perceiving point-light displays that are presented upside down (Troje & Westhoff, 2006).

Impaired perceptual performance with display inversion has been demonstrated consistently in biological motion perception. For example, the abilities to recognize a walking dog (Pavlova, 1989), judge human action type (Dittrich, 1993), and discriminate gender (Barclay, Cutting, & Kozlowski, 1978) are disrupted upon inverting point-light displays. The inversion effect associated with biological motion perception has often been attributed to impaired
processing of global, configural information (e.g., Bertenthal & Pinto, 1994). However, Troje and Westhoff (2006) have demonstrated that in addition to an inversion effect that may be attributable to the inversion of global form, there is a second inversion effect that relies on local motion conveyed by dots representing the feet of a walker. In their study, observers were presented with upright and inverted versions of point-light displays that were organized coherently or spatially scrambled (i.e., with individual dot trajectories displaced randomly) in a direction discrimination task. They found that observers were well able to discern walking direction not only from coherently organized displays, but also from spatially scrambled displays that lack any configural information. Critically, they showed that display inversion causes impairment in performance for both the coherent and scrambled displays. The inversion effect associated with scrambled displays cannot be attributed to impaired global form processing and must be local in nature. Further investigations revealed that the cues for direction of motion in scrambled displays and the associated inversion effect are carried by the feet of the walker. This finding is consistent with a previous report by Mather et al. (1992) regarding the importance of the foot motions in biological motion perception. Troje and Westhoff proposed that the visual mechanisms responsible for the local inversion effect may constitute an innate and non-specific life detection system that is distinct from an acquired system responsible for processing global shape that is required for more specific identification of an agent and its action (see also Johnson, 2006; Troje, 2008). To this end, we have shown that like the ability to discriminate direction, the perception of animacy from spatially scrambled point-light displays is also orientation-specific suggesting that the relevant local mechanisms may in fact convey information about animacy (Chang & Troje, 2008). The walking direction of scrambled displays can be well-discriminated in the visual periphery as well (Gurnsey, Roddy, Ouhnana, & Troje, 2008; Thompson, Hansen,
Hess, & Troje, 2007). In addition, a recent study by Jiang, Zhang, and He (submitted) showed that upright, but not inverted, scrambled biological motion attracts attention and that it is processed automatically.

Based on findings with neonate infants, chicks, and the data obtained from human adult observers, we hypothesize that the two mechanisms work together in a manner similar to that of the mechanisms proposed by Morton and Johnson (1991) for face perception. We suggest that local motion drives an early mechanism which is evolutionarily old and possibly innate. It may serve as a general detection system that directs attention to a stimulus class of great importance. That there may be evolutionarily acquired mechanisms towards identifying humans and animals has been suggested previously (New, Cosmides, & Tooby, 2007). New et al. (2007) hypothesized that humans have evolved a higher level of spontaneous recruitment of attention to humans and other animals, due to their ancestral values for survival and social opportunities, as compared to objects. Indeed, they found that performances on a change detection task in which observers were required to detect a difference between scenes that are identical except for a change in one target, were better when changes involved an animate rather than an inanimate target (e.g., a vehicle).

Once an animate target is detected, it can then be foveated and other mechanisms can be employed to further investigate it. We suggest that a crucial step at this level is the retrieval of the articulated, dynamically changing shape of the body of an animal or human. In the case of biological motion point-light displays, information about the articulation of the distinct dots is carried by the partly rigid motion of the segments of the body. The motion-mediated shape is eventually analyzed to obtain more detailed information about the nature of the animal and its actions.
Here, we present an experiment aimed to characterize the mechanisms responsible for the retrieval of global, motion-mediated shape, on the one hand, and for the retrieval of information contained in the local motion of individual dots, on the other hand. Note that we use the terms “local” and “global” in a very restricted and specific way. Our distinction between global, motion-mediated shape and local motion is not related to the usage of the terms “local motion” versus “global motion” when distinguishing between (local) object motion and (global) self-induced optic flow. Our usage of these terms is also not related to the intrinsic non-rigid (local) deformation of an articulated body and its overall (global) translation or rotation in space.

We tested observers on one of two tasks. One group of observers was tested on a walker detection task that addressed the global aspect of biological motion perception – requiring the observer to segregate the coherent structure of a walker from a mask of dots with similar local motion (i.e., a scrambled walker mask). The use of such a mask in this task renders the individual local trajectories of the walker uninformative and requires the observers to detect the structure of the walker. A second group of observers was asked to discriminate walking direction from both coherent displays (containing global and local information) and spatially scrambled displays (containing local information only). Walkers in this second task were embedded in a mask of randomly positioned flickering dots in order to retain informative local trajectories, in particular for the scrambled walker which, in contrast to the coherent walker, does not contain global structural information. Note that although coherent displays were included in order to render the task more rewarding for observers, the second task was designed to investigate primarily direction retrieval from local biological motions (scrambled displays). In both tasks, we investigated the effects of walker type, mask density, and learning on task performance. In
addition to examining these factors and comparing their effects between the two tasks, we also compared their effects between coherent and scrambled displays in the second task. In consideration of the claimed distinction between local and global mechanisms and their proposed innate and acquired natures, respectively (Troje, 2008), several predictions can be made regarding the behavior of mechanisms underlying the retrieval of local and global information contained in biological motion. If in fact, the relevant local mechanisms are ancestrally acquired and therefore innate, they may be insensitive to how familiar an observer is with a particular type of walker and may be insensitive to learning effects. In contrast, the relevant global mechanisms, presumed to be acquired individually, may be affected by the familiarity of the walker and may be subject to learning. Furthermore, if the role of the local mechanisms is to guide attention to animate motions, they may be pre-attentive in nature (Troje, 2008). As such, the local mechanisms, as compared to the global mechanisms may be more tolerant of masking noise. Importantly, any differences between the patterns of results for detecting/discriminating direction of coherent walkers and discriminating direction of scrambled displays would suggest that local and global information contained in biological motion patterns are processed by neural mechanisms with differing properties.

3.2 Methods

3.2.1 Participants

Two groups of naïve observers participated in this experiment. Group 1 consisted of 12 observers that ranged in age from 17 – 26 years (mean age of 19.2 years; 4 males, 8 females). Group 2 consisted of 12 different observers that ranged in age from 17 – 26 years (mean age of 19.3 years; 4 males, 8 females). All observers had normal or corrected-to-normal vision.
3.2.2 Stimuli and apparatus

The stimuli were derived from point-light sequences of a walking human, cat, and pigeon. The human walker was computed as the average walker from motion-captured data of 50 men and 50 women (Troje, 2002) and was represented by a set of 11 dots. The cat sequence was created by sampling 14 points from single frames of a video sequence showing a cat walking on a treadmill. The pigeon sequence was created from motion-captured data obtained from a pigeon fitted with 11 markers. All walkers were presented in sagittal view (i.e., facing rightwards or leftwards) and displayed stationary walking as if walking on a treadmill. Point-light sequences were shown at their veridical speeds with gait frequencies of 0.93 Hz, 1.7 Hz, and 1.6 Hz for the human, cat, and pigeon, respectively. On each trial, the starting position of the walker within its gait cycle was selected randomly.

The point-light walkers were presented upright and could be coherently organized (with all points maintaining veridical organization) or spatially scrambled (with points displaced randomly within areas matched to those occupied by the corresponding coherent versions). For observers in Group 1, right-facing coherent point-light walkers were embedded in a 6.4 x 6.4 deg scrambled walker mask. This mask was comprised of walker dots (i.e., dots carrying veridical trajectories of the walker) displaced randomly in space. The number of walker dots were sampled in five steps logarithmically from 50 – 150. For observers in Group 2, right- and left-facing walkers were embedded in a 6.4 x 6.4 deg random dot mask. This mask was comprised of randomly positioned stationary dots with a limited lifetime of 125 ms after which they were redrawn at new, randomly assigned locations. The number of dots in the random dot mask was sampled in five steps logarithmically from 50 – 750. For this group, target walkers were either
coherently organized or spatially scrambled. Positions of the dots for the spatially scrambled walker were randomly selected on each trial.

The stimuli were generated using MATLAB (Mathworks, Natick, MA) with extensions from the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997), and were displayed on a 22 inch ViewSonic P220f CRT color monitor with 0.25 mm dot pitch, 1280 x 1024 pixels spatial resolution, and 100 Hz frame rate. All stimuli appeared as white dots on a black background and the point-light figures subtended visual angles of 2.1 x 4.6 deg, 4.6 x 2.4 deg, and 3.6 x 3.6 deg for the human, cat, and pigeon, respectively.

3.2.3 Procedure

Observers completed one of two tasks: observers in Group 1 completed a detection task while observers in Group 2 completed a direction discrimination task. For both groups, participants were first instructed on the task both verbally and by printed instructions on the computer screen. A practice block was then presented during which participants familiarized themselves with the task. After the practice block, participants completed the experiment proper. For both groups, stimuli were viewed binocularly at a distance of 80 cm as maintained by a chin-rest.

Coherent walker detection task

Using a two-interval alternative forced-choice paradigm, observers in Group 1 were presented with two consecutive intervals on each trial: one interval containing a coherent walker and an alternate interval containing a scrambled walker, each embedded in a mask of additional scrambled walker dots. The task was to indicate the interval containing the coherent walker. On
each trial, the two 1 s display intervals were separated by a 0.5 s intertrial interval during which the screen remained blank. All walkers presented here were right-facing.

The practice block comprised of 6 trials that consisted of unmasked versions of all three animal types and 6 trials which showed the same displays embedded in a scrambled walker mask of 10 dots. The experiment proper consisted of 300 trials in total that were completed across three different experimental blocks of 100 trials, each featuring one of the three animal types. The order of test blocks was counterbalanced among participants. Each experimental block consisted of five sub-blocks of trials, each of which was comprised of two repetitions of all possible combinations of the interval of appearance (first or second) and mask density (50, 66, 87, 114, 150 scrambled walker dots). Within each repetition, the order of stimulus presentation was randomized. Feedback was not given for correct/incorrect responses.

**Direction discrimination task**

Observers in Group 2 were presented with a single 1 s display that consisted of a coherent or a scrambled walker embedded in a random dot mask. Here, a direction discrimination paradigm was used whereby the task was to decide whether the walker appeared to be facing leftwards or rightwards.

The practice block here comprised of 12 trials consisting of unmasked versions of all possible combinations of the three animal types, two stimulus organizations (coherent and scrambled), and two facing directions, and 12 trials consisting of masked versions of these same combinations. For the practice block, the random dot mask consisted of 10 dots. The experiment proper consisted of 300 trials that were completed across three different experimental blocks of 100 trials, each featuring one of the three animal types. The order of test blocks was
counterbalanced among participants. Each experimental block consisted of five sub-blocks of trials, each consisting of all possible combinations of stimulus organization, mask density (50, 98, 194, 381, 750 random dots), and facing direction. Within each sub-block, the order of stimulus presentation was randomized. Feedback was not given for correct/incorrect responses.

3.3 Results

3.3.1 Coherent walker detection

Due to the dependent and continuous natures of the factors sub-block and mask density, the walker detection data were entered in a multiple regression analysis adjusted for repeated-measures data (Lorch & Myers, 1990) that evaluated mask density, sub-block, and animal type as predictors for detection errors. Briefly, a multiple linear regression model was first fitted to the data of each individual. Individual coefficients obtained from these fits were then evaluated with two-tailed t tests. The linear model as given by the mean coefficients obtained from observers in Group 1 is

\[ e = 0.49 + 0.46 \log m - 0.03s - 0.05a_1 + 0.01a_2 , \]  

(3.1)

where error rate (e) is predicted by the logarithm of mask density (m), sub-block (s), and animal type (\(a_1\) and \(a_2\) in this particular equation binary-code for the human and cat stimulus types, respectively). The analyses indicated that mask density, \(t(11) = 7.360, p < 0.001\), and sub-block, \(t(11) = -2.805, p = 0.017\) were significant predictors of error rate. The contribution of the human walker stimulus type to the model was also different from that of the cat and the pigeon walkers,
$t(11) = -3.395, p = 0.006$, which did not differ, $t(11) = 0.561, p = 0.586$. Specifically, error rates were lower for the human walker stimulus (mean = 0.26) than for the cat (mean = 0.32) and pigeon (mean = 0.32) stimuli.

Detection performances, expressed in terms of error rates for the three animal types at the five mask densities, are presented in Figure 3.1. As reflected by Equation 3.1 and figure, error rates generally increase with increases in mask density.

![Figure 3.1](image.png)

Figure 3.1. Walker detection performances, expressed in terms of error rates for the three walker types and five mask densities. Error bars represent ± 1 standard error of the mean.

Detection performances for the different sub-blocks collapsed across animal type and mask density are presented in Figure 3.2. Here, decreased error rates with the later sub-blocks are evident.
3.3.2 Direction discrimination

The direction discrimination data were analyzed with a repeated-measures multiple regression model that evaluated factors of mask density, sub-block, animal type, and stimulus organization as predictors for error rate. In order to further examine the effect of stimulus organization on the relationships between error rate and each of mask density, sub-block, and animal type, three interaction terms were also included in this analysis. The linear model as given by the mean coefficients obtained from observers in Group 2 is

\[
e = -0.12 + 0.18 \log m + 0.01 s + 0.02 a_1 + 0.04 a_2 - 0.26 c + 0.08 c \log m - 0.02 c s - 0.02 c a_1 - 0.05 c a_2,
\]

(3.2)
where error rate (e) is predicted by the logarithm of mask density (m), sub-block (s), organization (c), and animal type (human and cat stimulus types binary-coded here by $a_1$ and $a_2$, respectively). Individual two-tailed $t$ tests for the regression coefficients indicated that mask density, $t(11) = 7.134, p < 0.001$, stimulus organization, $t(11) = -2.260, p = 0.043$, and the stimulus organization by mask density interaction, $t(11) = 2.399, p = 0.035$, were significant predictors of error rate. All other coefficients were not significantly different from zero.

Direction discrimination performances, expressed in terms of error rates for the three animal types at the two stimulus organizations and five mask densities are presented in Figure 3.3A. As reflected in the regression model and figure, the error rates for coherent stimuli (mean = 0.12) were lower than error rates for scrambled stimuli (mean = 0.32). An examination of this figure reveals also a general increase in error rates with increases in mask density, as predicted by Equation 3.2.

Figure 3.3B re-plots the data for coherent and scrambled conditions across the various mask densities, collapsed across animal type. To better illustrate the interaction reflected by the regression model, a further analysis of these two factors was performed with a two-tailed, paired $t$ test comparing coherent and scrambled conditions in terms of regression slopes for individual participant data plotted on a logarithmic scale. That is, for each participant, linear regression slopes were obtained by plotting error rates against mask densities on a logarithmic scale, separately for coherent and scrambled conditions and collapsed across all other factors. The mean slopes are depicted by the regression lines in Figure 3.3B. As evident in the figure, the increase of errors with increasing mask density was larger for coherent than for scrambled stimuli, $t(11) = 2.399, p = 0.035$. 

61
Figure 3.3. (A) Direction discrimination performances, expressed in terms of error rates for coherent and scrambled versions of the three walker types across five mask densities. (B) Mean error rates for coherent and scrambled conditions collapsed across animal type plotted against mask densities. Superimposed linear regression lines carry the mean slopes obtained by plotting error rates against mask density on a logarithmic scale for individual participants. Error bars represent ± 1 standard error of the mean.
Direction discrimination performances of coherent and scrambled displays for the different sub-blocks collapsed across animal type and mask density are presented in Figure 3.4. As noted earlier, stimulus organization did not significantly alter the relationship between sub-block and error rate with all other variables held constant. Although not confirmed by the statistical analyses, the figure implies that some degree of learning might have occurred over the first three blocks for the coherent stimuli only.

Figure 3.4. Direction discrimination performances, expressed in terms of error rates for coherent and scrambled stimuli for the five learning sub-blocks, collapsed across walker type and mask densities. Error bars represent ± 1 standard error of the mean.

3.4 Discussion

With two tasks, we investigated the effects of walker familiarity (human vs. non-human), mask density, and learning on the ability to discriminate a coherent walker from a scrambled walker (walker detection task) and the ability to discriminate direction from both coherent and scrambled walker displays. While the first task relied on the observer’s ability to retrieve the global motion-mediated shape of the walker, the second task (at least where the scrambled walker
was involved) allowed us to assess the observer’s ability to retrieve information from the local motion of the individual dots.

The effects of the three factors on the two tasks were very different. Walker type strongly affected the detection task, which requires the retrieval of global form information. Specifically, error rates were lower for the familiar human walker stimuli than for the cat and pigeon stimuli. In contrast, walker type had no effect on the direction discrimination task, which was designed to address primarily visual processing of local motions. The two tasks were also differentially affected by learning. While error rates decreased at the later test blocks for the detection task, no learning effect was observed for the direction discrimination task. For the coherent displays in the direction discrimination task, performance was generally very high and the lack of effects of animal type and learning may be due to a ceiling effect. More important, however, is the lack of such effects for the scrambled displays. Here, performance was lower and the lack of effects of walker type and learning cannot be explained by ceiling performance. Comparing the effect of masking between the detection and direction discrimination tasks is less straightforward as we used very different masks in the two tasks. In the detection task, the mask consisted of additional, scrambled walker dots, whereas in the direction discrimination task, the mask consisted of stationary flickering dots. Common to both tasks is an effect of mask density.

Considered alone, the increase in error rates with increases in mask density on the walker detection task replicates previous findings involving a comparable mask and task (Hiris, 2007). In this study, Hiris (2007) measured detection sensitivity for a coherent point-light walker, a non-biological structured rotating stimulus, and a non-biological unstructured rotating stimulus, all of which were either embedded in a random-mask or in a scrambled motion mask (in which individual local motions were identical to those of the targets) with varying densities.
Interestingly, for the scrambled motion mask, detection sensitivity did not differ between the point-light walker and the non-biological structured stimulus. The author therefore argued that differences between the ability to detect a biological motion stimulus and a non-biological motion stimulus can be explained by the fact that biological motion always contains underlying structure.

Critically, the comparison of interest for the present purposes is between the coherent and scrambled stimuli in regards to the effect of mask density within the direction discrimination task. As noted earlier, for the coherent walkers, observers could potentially employ both configural information and local motion information as cues to the facing direction of the walker. As such, there is no question that more information is contained in the coherent walker (in the form of additional structure) than in the scrambled walker, where only local motion information is available. Correspondingly, the error rates for the coherent walkers were significantly lower than those for the scrambled walkers. As depicted in Figure 3.3B, the rate of increase in errors with increases in mask density was different for coherent than for scrambled walkers. Specifically, the errors increased more substantially for coherent than for scrambled walkers with increasing mask density. It should be noted that the manner in which the random-dot mask affects the information contained in the two types of walkers is unknown however. It is possible that increases in density of the random-dot mask may more readily disrupt configural information contained in the coherent walkers as compared to local motion information contained in the scrambled walkers.

The differences in the patterns of effects observed between detecting/discriminating the direction of a coherent walker and discriminating the direction of a scrambled walker suggest that global and local cues contained in biological motion are retrieved from visual mechanisms with differing properties. Our findings for the two tasks fit well with the hypotheses set forth by Troje and Westhoff (2006) and Troje (2008); that is, the proposition of an innate, ancestrally acquired
mechanism that is sensitive to invariants contained in the local motion of individual dots and that is distinct from an ontogenetically trained mechanism that retrieves global form in biological motion perception. The direction discrimination results, in particular for the scrambled walkers that contain solely local cues, are consistent with the hypothesis that the local mechanism is non-specific, hard-wired, and possibly innate. Consistent with our predictions, there were no effects of walker type or learning. In addition, performances were relatively tolerant of masking noise, supporting the hypothesis of a robust and possibly pre-attentive local system. In contrast, an individually acquired global processing system may be sensitive to walker type and should exhibit learning effects. The results for our walker detection task are consistent with these latter predictions. We show further that the global system is also comparatively more susceptible to masking noise.

It is worth noting that an expertise-dependent nature for the global mechanism in biological motion perception is supported by several experimental findings (Fox & McDaniel, 1982; Grossman, Blake, & Kim, 2004; Jastorff, Kourtzi, & Giese, 2006). In one study, Fox and McDaniel (1982) tested the sensitivity of infants aged 2, 4, and 6 months to coherent biological motion patterns using a forced preferential looking technique. On each trial of the relevant experiment, two stimuli were presented side by side: one display of a point-light human runner organized coherently and an alternate display of randomly-moving dots. The authors found visual preference for the coherent biological motion pattern as opposed to the random-dot display in 4- and 6- month-olds but not in 2-month-old infants. This suggests then that the global mechanism in biological motion perception does not develop until some time between 2 – 4 months of postnatal life. At first glance, the findings of Fox and McDaniel may appear to be inconsistent with those reported by Simion et al. (2008). However, the fact that the newborn
Infants in the study by Simion et al. preferred to look at point-light animations of a non-human animal (hen) suggests that their preferences may have been guided by form-invariant, local cues rather than global shape.

Using functional magnetic resonance imaging (fMRI), perceptual learning of coherent biological motion patterns and its effects on neural activity have also been investigated in adults (Grossman et al., 2004). In this study, the authors trained observers to discriminate coherent from scrambled biological motion patterns embedded in noise dots and measured neural activity before and after training. Consistent with our results reported here, behavioral performances of the adults in this study improved from pre- to post-training. Importantly, the behavioral improvements were paralleled by increased activity after training in the posterior superior temporal sulcus (STS) in response to the coherent patterns as compared to pre-training neural activity.

The posterior STS has often been implicated in the perception of coherent biological motion patterns (e.g., Bonda, Petrides, Ostry, & Evans, 1996; Grossman & Blake, 2001; Grossman et al., 2000; Oram & Perrett, 1994; Vaina, Solomon, Chowdhury, Sinha, & Belliveau, 2001). For example, using fMRI, Grossman et al. (2000) showed that the STS is responsive to coherent point-light displays but is not activated by scrambled point-light displays, coherent motion, or kinetic boundaries. Using positron emission tomography (PET), Bonda et al. (1996) have also shown this area to be active when observers view point-light animations. Thus, this region may well be implicated in the system that processes global form information from biological motion patterns. The possible neural sites for the processing of local biological motions are less clear. It has been shown recently however that the extrastriate areas V3 and
V3A are differentially responsive to upright and inverted versions of scrambled biological motion displays (Jiang & He, 2007).

Nonetheless, it is clear that the local and global mechanisms underlying biological motion perception can be regarded as distinct processing modules. While it is assumed that the global mechanism is sensitive to the shape of the walker, it is still unclear as to which motion properties the local mechanisms are tuned to. Troje and Westhoff (2006) showed that the inversion effect associated with spatially scrambled walkers is carried by the motions of the feet. In an earlier study, Mather et al. (1992) showed that performances on coherence and direction discrimination tasks involving point-light walkers were most affected by the omission of wrist and ankle dots as opposed to the omission of the shoulder and hip, elbow and knee dots. Thus, converging lines of evidence suggest that the cues of interest to the relevant local mechanism are contained in the motions of the feet. What is the nature of these cues? To this end, we have investigated the local inversion effect in biological motion perception with novel stimuli that display solely foot-specific information (Chang & Troje, 2009). In one experiment, we compared direction discrimination performances for displays containing naturally accelerating foot motions with those containing constant speeds (i.e., with accelerations removed along the trajectory paths). We found that the inversion effect holds for the naturally accelerating stimuli but not the constant speed stimuli suggesting that acceleration contained in the foot motion carries the local motion-based inversion effect in biological motion perception. In light of these findings, we then conjecture that the local processing system in biological motion perception is based on knowledge about the characteristic acceleration pattern contained in foot motions as an animal moves through the environment which is constrained by gravity, inertia, and the general kinetics of moving bodies.
In sum, recent findings of at least two sources for the inversion effect in biological motion perception have prompted the need to consider separately, the contributions of global and local motion processes (Troje & Westhoff, 2006). Here, we have characterized these two processes at the behavioral level. While performances presumed to rely on the global mechanism are sensitive to walker type and learning effects, and are heavily affected by masking noise, performances presumed to exploit the local mechanism are not affected by varying walker types and learning effects, and are relatively robust to masking noise. These findings suggest that the human visual system exploits both information contained in global, motion-mediated form, and information contained in the local motion of individual body parts, and that the retrieval of these two types of information are governed by distinct mechanisms with very different properties.
References


Chapter 4

Acceleration and the Local Inversion Effect

4.1 Introduction

Visual perception of animal movement patterns, or biological motion patterns (Johansson, 1973), has traditionally been investigated with displays comprised of a number of light-points depicting the motions of an agent’s major joints. When the points are organized coherently, such a display is sufficient to allow the recognition of an agent and its action. Moreover, higher-order characteristics such as gender (e.g., Kozlowski & Cutting, 1977; Barclay, Cutting, & Kozlowski, 1978; Mather & Murdoch, 1994; Troje, 2002) and identity (e.g., Cutting & Kozlowski, 1977; Hill & Pollick, 2000; Troje, Westhoff, & Lavrov, 2005) can also be extracted.

Orientation-specificity is a widely demonstrated characteristic of biological motion perception. Perceptual performance is generally impaired if the point-light displays are inverted. For example, observers in a study by Sumi (1984) reported that the perceived quality or naturalness of movement was degraded with inversion. Pavlova (1989) showed that observers were impaired in the recognition of a walking dog when the point-light sequence was inverted. Bertenthal and Pinto (1994) found that the detection of a coherent point-light target in a mask of additional walker point-lights dropped to chance when the target was inverted. Observers’

---

abilities to judge human action type (Dittrich, 1993), and the agent’s gender (Barclay, Cutting, &
Kozlowski, 1978) were also disrupted upon inverting point-light displays. These inversion-
related impairments in perceiving coherent point-light displays have often been attributed to
impaired processing of global, configural information as conveyed by the display’s overall
spatiotemporal organization (e.g., Bertenthal & Pinto, 1994). Indeed, the importance of
configural information for biological motion perception is supported by findings that walking
persons can be perceived from point-light figures in which each point is reallocated to another
randomly selected position on the limb in successive frames or successive sets of frames and thus
have little or no local motion (Beintema & Lappe, 2002; Neri, 1998).

Notably, it has been shown recently that in addition to an inversion effect that may well
be attributable to the inversion of global form, there is a second inversion effect in biological
motion perception that cannot be attributed to configural processing and is inherent to properties
of the local motion signals (Troje & Westhoff, 2006). In their study, Troje and Westhoff (2006)
presented observers with walking animals that were organized coherently or spatially and
temporally scrambled thereby rendering any global spatio-temporal information useless. All
stimuli displayed stationary walking; that is, both the coherent and scrambled variations of the
walkers did not contain any overall translatory motion. The displays were presented at both
upright and inverted orientations and observers were asked to indicate the direction in which the
walkers were facing. For the coherent displays, they found that direction discrimination
performance decreased at the inverted orientation relative to the upright orientation. This
particular result is consistent with many previous studies demonstrating an inversion-related
impairment for perceiving coherent point-light patterns (e.g., Sumi, 1984; Pavlova, 1989). Of
particular relevance however, is their finding that from the spatially scrambled displays that
contain solely local motion cues, observers could still discriminate facing direction reliably. Critically however, performances dropped to chance level when display orientation was inverted. As scrambled displays do not contain coherent structure, this inversion effect cannot be attributed to the processing of global configuration. An additional experiment where only specific parts of the display were inverted revealed that the cues for direction of motion and the associated inversion effect were carried by the dots representing the feet of the walkers. This finding corroborated an earlier study by Mather, Radford, and West (1992) that showed that performances on coherence and direction discrimination tasks were most affected by the omission of the wrist and ankle dots, as compared to the omission of the shoulder and hip, or elbow and knee dots.

The findings by Troje and Westhoff (2006) suggest that two inversion effects need to be considered in biological motion perception. While there seems to be an inversion effect which relies on the global configuration of the walker, there is also a second one, which relies on the local motion of the distal parts of the lower body. The authors suggested that this second inversion effect reflects a visual mechanism that sub-serves the ability to detect terrestrial, articulated animals in the visual environment (see also, Troje, 2008).

In support of the hypothesis put forth by Troje and Westhoff (2006), we have shown that the perception of animacy from spatially scrambled point-light displays is also orientation-specific (Chang & Troje, 2008). In this study, we presented observers with coherent and scrambled displays at both upright and inverted orientations and asked observers to rate each display on a six-point Likert scale according to whether it was a “living being”. Critically, we found that upright, scrambled displays were rated more animate than inverted, scrambled displays, suggesting that the mechanisms responsible for processing local biological motion
signals not only retrieve locomotive direction but also aid in identifying animate agents in the environment.

According to Troje and Westhoff (2006), the critical invariant in local biological motions is based on the fact that gravitational acceleration imprints orientation-specific signatures on the movements of a terrestrial animal in locomotion (see also, Shipley, 2003). Indeed, Bingham, Schmidt, and Rosenblum (1995) have shown a role for dynamics for the perception of both inanimate and animate events. In this study, observers were presented with point-light displays of events that reflected varying types of dynamics (i.e., biodynamics, hydrodynamics, aerodynamics, rigid-body). Both the orientation of the display and observer orientation were manipulated and observers were asked to describe the events with free-responses or by circling properties from a given list. Results indicated that observers were well able to identify the various events as depicted by point-light motions. Importantly, observers’ descriptive patterns reflected the various underlying types of dynamics. In addition, perception seemed to depend on the orientation of the event with respect to gravitational direction and seemed unaffected by the relative orientation of the observer with respect to the display. From these results, the authors concluded that underlying dynamics play a role in the perceptual significance of motions and specifically that variations in velocity contained in the motion, or trajectory form, contributes to event identification (for original work on event identification from motion or trajectory forms, see Runeson, 1974; Bingham, 1987).

Our previous finding that both the ability to discriminate walking direction and the perception of animacy from scrambled biological motion displays could be well-elicited within stimulus durations as short as 200 ms (Chang & Troje, 2008) suggests that the visual system is able to exploit the relevant cues from only a fraction of the full gait cycle (~1100 ms for a human
walker). Here, we ask whether all fragments of the gait cycle contain equally salient directional information and contribute equally to the associated inversion effect.

We present three experiments that investigated the cues underlying the local inversion effect by testing direction discrimination from novel biological motion stimuli that contain only small fragments of different types of foot trajectories. In Experiment 1, we first asked if the cues contained in the local foot motion can be assessed from a single dot trajectory by isolating the local motion of the human foot and presenting displays derived from one foot dot tracing an individual fragment of the trajectory or displays derived from one pair of foot dots that trace counterphase fragments of the gait cycle (as in the case for the two feet of a full walker). We show that the inversion effect occurs only for displays derived from the paired fragments. In Experiment 2, we investigated direction retrieval from displays derived from paired fragments of the gait cycle belonging to other foot types (human walker, human runner, cat, and pigeon) in order to gain insight into the type of cues that may be relevant for the inversion effect. Finally, in Experiment 3, we investigated specifically the importance of acceleration for the local inversion effect by comparing direction discrimination for the human walker stimuli of Experiment 2 and stimuli that were identical to these naturally accelerating stimuli, but had accelerations removed along the trajectory fragments (i.e., moved with uniform speeds).

4.2 Experiment 1

Troje and Westhoff (2006) showed that for full walker displays that are spatially scrambled, the cues for direction and the inversion effect are carried by the motion of the dots representing the feet. These cues are exploited by the visual system independently of the temporal relationships among the individual dots and can thus be thought to be completely local
in nature. Additionally, we have shown that the relevant cues can be retrieved within as short as 200 ms (Chang & Troje, 2008). Together, these findings imply that observers can retrieve the facing direction of the display from a stimulus that exposes only a single 200 ms window of the foot trajectory. Yet, it is difficult to imagine that such a stimulus would be sufficient. Recall that a single moving dot if taken from a stationary walker, exhibits no net translatory motion if displayed for the duration of a full gait cycle. In contrast, a display consisting of an isolated short fragment (e.g., 200 ms) of the foot’s trajectory contains a clear net displacement. Depending on the part of the trajectory it is sampled from, the translation is either coherent or incoherent with the facing direction of the walker. If observers base their decision on the displacement of the individual dots, they would perform at chance level on average and an inversion effect would not be expected. It is hard to imagine that observers would ignore displacement as a cue to direction, and it is hard to imagine what other could be left in the short fragments that would signal the facing direction of the walker.

Perhaps then, single short trajectory fragments do not provide the critical information – despite the finding that spatial and temporal structure is not critical. Perhaps it is necessary to present the trajectories within a context where the local translatory motion of the individual dots is equated by adding other dots with local translatory motion in the opposite direction, thus rendering these cues non-discriminative and forcing the observer to use intrinsic cues in order to determine the facing direction of a stationary walker.

In Experiment 1, we employed two sets of stimuli. We presented one group of observers with displays derived from a single foot dot that traced a particular individual fragment of the gait cycle, and a second group of observers with displays derived from a pair of foot dots that traced counterphase fragments of the gait cycle. According to the reasoning above, observers in the first
group would be expected to use primarily translatory cues resulting in poor performance and no
inversion effect in the direction discrimination task whereas observers in the second group would
show behaviour similar to that obtained previously with full scrambled walker displays (i.e.,
elevated performance rates and a pronounced inversion effect).

4.2.1 Methods

Participants

Two groups of naïve observers participated in this experiment. Group 1 consisted of six
observers that ranged in age from 18 – 26 years (mean age of 20.50 years; 2 males, 4 females).
Group 2 consisted of 12 observers that ranged in age from 18 – 23 years (mean age of 19.03
years; 5 males, 7 females). All observers had normal or corrected-to-normal vision.

Stimuli and apparatus

The motion of the foot of a human walker as defined by the point of one ankle isolated
from the original human point-light sequence used in Troje and Westhoff (2006) was used to
derive two sets of stimuli for the current experiment. The foot’s motion, shown at its veridical
speed with a gait frequency of 0.93 Hz, was displayed in sagittal view either from the left or from
the right side. The full gait cycle was sampled by specifying 10 start frames, spaced at
approximately equal intervals along the cycle, from which the fragment could begin\(^4\). From these
starting positions, fragments were created that extended 100 ms, 150 ms, or 200 ms in length.
Figure 4.1 depicts the 10 start frame positions along the full gait trajectory of the foot (A), and the
individual trajectories of the 10 fragments at the three window lengths (B).

\(^4\) In Experiment 1, a cumulative rounding error that became apparent only after data
collection resulted in a slightly shorter distance between starting points of fragments 1
and 2 as compared to the distance between the starting points of all other adjacent
fragments. The error was corrected in Experiment 2.
Figure 4.1. The trajectory of the human foot across a full gait cycle with the 10 starting positions (A), and the foot trajectories outlined for each fragment (1 – 10) at the three temporal lengths (B). Trajectories shown here are taken from a right-facing walker. Thus, while fragments 1 – 4 displace to the right, fragments 6 – 10 displace to the left. For fragment 5, the direction of displacement reverses as the window length increases.

Observers in Group 1 were presented with displays 1000 ms in duration that consisted of 10 copies of one specified foot fragment (identical start-frame and length) that were displayed at random positions within a 3.6 x 3.6 deg area in the centre of a computer screen. Each copy had its own randomly chosen onset time and was visible only during the specified time window. The stimuli were presented upright or inverted about the horizontal axis.

Observers in Group 2 were presented with stimuli derived from five fragment pairs created by pairing fragments labeled 1/6, 2/7, 3/8, 4/9, and 5/10 in Figure 4.1. Thus, each fragment pair consisted of two foot point motions, one of which traced a trajectory that was 180 deg phase-shifted relative to the other along the gait cycle. Each display consisted of 10 sets of a specified fragment pair (identical start-frame and length) that were displayed at random positions...
within a 3.6 x 3.6 deg area in the centre of the screen. Each set had its own randomly chosen onset time and was visible only during the specified time window. Again, stimuli were presented at both upright and inverted orientations.

The stimuli were generated using MATLAB (Mathworks, Natick, MA) with extensions from the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997), and were displayed on a 22 inch ViewSonic P220f CRT color monitor with 0.25 mm dot pitch, 1280 x 1024 pixels spatial resolution, and 100 Hz frame rate. All stimuli appeared as white dots on a black background and each display subtended a visual angle of 3.6 x 3.6 deg.

Procedure

The stimuli were viewed binocularly at a distance of 80 cm as maintained by a chin-rest. A two-alternative, forced-choice direction discrimination paradigm was used where the observers’ task was to decide the direction (left or right) in which the perceived entity (whatever it may be) seemed to face. Feedback was not given for observer responses.

For both groups, participants were first instructed on the task both verbally and by printed instructions on the computer screen. A practice block was then presented during which participants familiarized themselves with the task. After the practice block, participants completed the experiment proper.

For Group 1, the practice block comprised of 40 trials that consisted of left and right, and upright and inverted versions of the 10 gait cycle fragments of 200 ms length. The experiment proper consisted of 960 trials in total that were completed across two identical experimental blocks of 480 trials each. Each experimental block consisted of all possible combinations of the two directions, two orientations, 10 fragments, and three window lengths, repeated four times. Within each repetition, the 120 possible stimuli were presented in random order.
For Group 2, the practice block comprised of 20 trials that consisted of left and right, and upright and inverted versions of the five fragment pairs of 200 ms length. A total of 960 trials were completed for the experiment proper across two identical experimental blocks of 480 trials each. A single experimental block consisted of all possible combinations of the two directions, two orientations, five fragment pairs, and three window lengths, repeated eight times. Within each repetition, the 60 possible stimuli were presented in random order.

The resulting design of this experiment for both groups was a within-subject test of factors orientation (i.e., upright or inverted), fragment (Group 1) or fragment pair (Group 2), and fragment length.

4.2.2 Results

The data were analyzed both in terms of the proportions of responses consistent with the facing direction of the walker from which the stimuli were derived and in terms of $d'$ sensitivity measures (Macmillan & Creelman, 2005) derived from these rates. The two sets of analyses yielded identical results. For greater clarity, data are presented here in terms of the proportions of responses consistent with walking direction.

**Single fragments**

Results obtained from Group 1 are illustrated in Figure 4.2A. Overall performance was at chance level (0.45 proportion of responses consistent with the facing direction of the walker). It is evident that observers simply responded to the horizontal translatory displacement of the individual dots. For fragments 1 – 4 where dots moved in the direction in which the walker was originally facing, rates of responses that were consistent with the facing direction of the walker were high. For fragments 4 – 9 where dots moved in the opposite direction, observers consistently reported the opposing direction. For fragments 5 and 10 where the horizontal
direction of the dot switches, responses depended on display length in a manner consistent with the assumption that observers based their responses on mere horizontal translation. Additionally, the data from this group show no inversion effect.

Figure 4.2. Performance expressed in terms of the mean proportion of responses consistent with the facing direction of the walker. Thus, response rates above and below 0.5 indicate larger proportions of responses that were consistent, $C$, and inconsistent, $I$, with walking direction, respectively. (A) Response rates for the 10 fragments across the three temporal lengths and between two orientations obtained for Group 1. (B) Response rates for the five fragment pairs across three fragment lengths and between two orientations for Group 2. (C) Response rates of five fragment pairs across three fragment lengths and between two orientations computed from the individual fragment response rates obtained for Group 1. Error bars represent ± 1 standard error of the mean.

Statistics support the above observations. A 2 (test block) x 2 (orientation) x 10 (fragment) x 3 (length) repeated-measures analysis of variance (ANOVA) showed a significant
main effect of fragment, $F(9, 45) = 879.73, p < 0.001$, a significant fragment x length interaction, $F(18, 90) = 19.26, p = 0.007$, but no effect of orientation, $F(1, 5) = 0.33, p = 0.589$, or length, $F(2, 10) = 1.95, p = 0.222$. All other interactions were not significant and performance between the two test blocks did not differ, $F(1, 5) = 0.64, p = 0.460$. Thus, data from the two blocks were pooled for all further analyses.

Tukey’s post-hoc comparisons of the response rates of the 10 fragments, collapsed across orientation and length, indicated that the rates for fragments 2 – 4 did not differ ($p > 0.900$ for all comparisons), but were higher than rates for all other fragments ($p < 0.020$ for all). The rates of fragments 6 – 9 did not differ ($p > 0.900$ for all), but were substantially lower than the rates of fragments 1, 5, and 10 ($p < 0.001$ for all), which differed from each other ($p < 0.001$ for all).

The fragment x length interaction was investigated with individual one-way ANOVAs that compared the three lengths for each fragment. The analyses yielded a significant main effect of length for fragments 5, $F(2, 10) = 38.25, p < 0.001$, and 10, $F(2, 10) = 25.68, p < 0.001$, only. Post-hoc Tukey comparisons of the three lengths for fragment 5 showed that the rate at 100 ms was significantly higher than rates at 150 ms and 200 ms ($p < 0.001$ for both), and the rate at 150 ms was significantly higher than the rate at 200 ms ($p < 0.001$). Comparisons for fragment 10 showed that while rates at 100 ms and 150 ms did not differ ($p = 0.574$), they were significantly lower than the rate at 200 ms ($p < 0.001$ for both).

**Paired fragments**

The results for Group 2 are displayed in Figure 4.2B. Overall performances were above chance level (0.59 proportion of responses consistent with facing-direction of the walker) with rates generally higher for the upright (0.64 proportion of responses) than for the inverted displays (0.53 proportion of responses). Closer inspection reveals that at both orientations, observers
perceived the direction consistent with the facing direction of the walker for pairs 2/7, 3/8, and 4/9 (0.57 – 0.90 proportion of responses) with rates increasing systematically with increasing window length, but perceived the opposite direction for pair 5/10 (0.15 – 0.36 proportion of responses) with the rate of responses in the opposite direction increasing systematically with increasing window length. For pair 1/6, the perceived direction changed with changes in window length (0.25 – 0.70 proportion of responses).

Response rates for Group 2 were analyzed with a 2 (test block) x 2 (orientation) x 5 (fragment pair) x 3 (length) repeated-measures ANOVA. The analysis showed significant main effects of orientation, $F(1, 11) = 361.98, p < 0.001$, fragment pair, $F(4, 44) = 158.00, p < 0.001$, and length, $F(2, 22) = 62.89, p < 0.001$, a significant orientation x length interaction, $F(2, 22) = 16.10, p < 0.001$, and a significant fragment pair x length interaction, $F(8, 88) = 11.98, p < 0.001$. Performances between the two test blocks did not differ, $F(1, 11) = 0.26, p = 0.621$ and were thus pooled. A $t$ test using the pooled response rates confirmed that overall rates of responses consistent with the facing direction of the walker were higher than chance-level, $t(11) = 7.712, p < 0.001$.

Tukey’s post-hoc comparisons of the five fragment pairs, collapsed across orientation and length, indicated that the rates for pairs 2/7 and 3/8 did not differ ($p = 0.820$), and were both higher than the rate for pair 4/9 ($p < 0.005$ for both). The rates of pairs 1/6 and 5/10 differed from those of all other pairs ($p < 0.001$ for all comparisons). Finally, Tukey’s comparisons of the three lengths revealed that rates at all three lengths differed ($p < 0.001$ for all). Specifically, rates increased with increasing length.

The orientation x length interaction was analyzed with a one-way ANOVA that evaluated difference scores, obtained for each participant by subtracting rates of the inverted condition from
those of the upright condition while collapsing for orientation and fragment pair, across the three lengths. The analysis showed a significant main effect of length, $F(2, 22) = 16.16, p < 0.001$. Tukey’s post-hoc comparisons revealed that the inversion effect, measured in terms of the difference in rates of responses consistent with walker direction between upright and inverted orientations, was smaller for length 100 ms than for 150 ms and 200 ms ($p < 0.005$ for both), and the inversion effects for 150 ms and 200 ms did not differ ($p = 0.194$).

The fragment pair x length interaction was examined with individual one-way ANOVAs that compared the three lengths for each pair. The analyses showed a significant main effect of length for pair 1/6, $F(2, 22) = 31.16, p < 0.001$, pair 2/7, $F(2, 22) = 10.21, p < 0.001$, pair 3/8, $F(2, 22) = 6.32, p = 0.007$, pair 4/9, $F(2, 22) = 4.57, p = 0.022$, and pair 5/10, $F(2, 22) = 13.31, p < 0.001$. Tukey’s comparisons for pair 1/6 indicated that rates at all three lengths differed ($p < 0.005$ for all) with rates increasing with increasing window length. Comparisons for pair 2/7 indicated that the rate for 100 ms was significantly lower than the rates for 150 ms and 200 ms ($p < 0.003$ for both), and the rates for 150 ms and 200 ms did not differ ($p = 0.967$). Similarly for pair 3/8, the rate for 100 ms was significantly lower than the rates for 150 ms and 200 ms ($p < 0.017$ for both), which did not differ ($p = 0.987$). Comparisons for pair 4/9 showed that the rate at 100 ms was statistically lower than the rate at 200 ms only ($p = 0.020$), although the rates for 150 ms and 200 ms did not differ ($p = 0.687$). Finally, pair 5/10 showed a reversed trend from that of pairs 2/7 and 3/8. Specifically, the rate of responses consistent with the direction of the walker for 100 ms was significantly higher than the rates for 150 ms and 200 ms ($p < 0.001$ for both), which did not differ ($p > 0.900$).
4.2.3 Discussion

The results for Group 1 are qualitatively different from those of Group 2. As expected, observers in Group 1 who were presented with displays derived from a single fragment merely based direction judgments upon the horizontal displacement of the individual dot elements. Across a full gait cycle, the displacement averages out and there is no net displacement. Accordingly, observers’ overall response rates were at chance level. Interpreting the displays solely based upon dot translatory movement also explains the lack of an inversion effect in the results of Group 1.

In contrast to Group 1, the results for observers in Group 2 who were presented with displays derived from paired fragments show above chance-level response rates and a strong inversion effect. These results are consistent with the orientation-dependency for discriminating walking direction from scrambled displays derived from full walkers and corroborate the finding of the critical role of the foot motion (Troje & Westhoff, 2006). The response behaviour of observers in Group 2 is qualitatively different from that of observers in Group 1 and cannot be explained by a simple linear integration of the responses given to the individual fragments. Figure 4.2C re-plots the results of Group 1 collapsing the data over fragment pairs. Average response rates for these pairs are close to chance level and there is no inversion effect when the fragments are shown in isolation (Group 1). These data are very different from the data obtained when the fragment pairs are presented simultaneously (Group 2).

Critically, these results suggest that retrieving direction of motion from scrambled biological motion displays involves more than processing the local spatiotemporal cues inherent to the isolated motion of the foot. Rather, we must conclude that even within the context of a full scrambled walker display that cannot be resolved into a percept of coherent form, the foot’s
elemental cues are evaluated with respect to the motions of other elements, or at least to other parts of the same element’s trajectory, as we have shown here. But, why are additional reference dots important?

Critical to the discussion of this question is the distinction between the direction of explicit translatory motion of an object (or a walker, in our case), and the implicit cues contained in the deformation of the body. We call the former extrinsic motion and the latter intrinsic motion (see also, Kersten, 1998, who used these terms in a slightly different, yet analogous context). In his early work, Johansson (1973, 1974) demonstrated how the human visual system decomposes the kinematics of the display of a rolling wheel into common translatory (i.e., extrinsic) motion and deviant circular (i.e., intrinsic) motion in order to understand the mechanics of the scene. Presented alone, a point representing the rim of a rolling wheel is interpreted as moving along a cycloidal path (i.e., a path resembling periodic arches). Only when a second point is added (e.g., on the axle of the wheel) can observers perceive the rolling wheel. In the context of our stimuli, subtraction of the common translatory (extrinsic) component of the motion results in the absolute cyclic motion of the foot and retains the walker’s intrinsic motion, but only if the full gait cycle can be considered. The qualitative difference in response behaviour between the two groups in the present experiment is most likely due to a switch from the perception of solely extrinsic motion with the single fragments in Group 1 to the use of intrinsic motion cues with the paired fragments in Group 2. Only the latter matches the reality from which the stimuli were derived: the stationary walker that lacks any extrinsic motion but contains intrinsic motion cues revealing its facing direction.

It should be noted at this point that participants in this experiment were not informed as to the nature of the stimuli (i.e., that they were derived from a human walker). It is clear then,
that the cues that are relevant for retrieving direction are independent of higher-level knowledge of the nature of the stimuli and particularly, of whether they derive from animate or inanimate objects. The exact nature of these cues however, and their contributions to the observed effects remain unclear.

Most of the fragment pairs employed in this experiment consist of a fragment that represents a fraction of the foot trajectory’s stance phase and a fragment that represents a fraction of its swing phase. During the stance phase, the foot is planted on the ground and is more or less motionless. The motion of the dots that correspond to the stance phase is rather due to the motion of a panning camera that keeps the walker in the center of the display, or alternatively, to the motion of the belt of an invisible treadmill. The velocity of the dot is approximately constant and identical to the translatory (extrinsic) motion of the walker, but points in the opposite direction. The other dot, which represents a fragment from the swing phase exhibits a velocity profile that is much more variable. During the swing phase, the foot accelerates along both the horizontal and vertical dimensions due to muscle activity and gravitational acceleration. If the visual system could reliably identify stance-phase fragments and discriminate them from other types of fragments, it could safely assume that the walker is facing in the direction opposite to the direction in which the stance phase fragments move.

What, then, accounts for the inversion effect? Inversion, that is, the mirror flipping of stimulus about the horizontal axis, affects only vertical components of the trajectory. The linear, horizontal motion of the stance phase fragments is largely unaffected by inversion. We must conclude then that the observed inversion effect is due to vertical asymmetries in the swing phase segments. In principle, such asymmetries can be due to the polarity of vertical velocity, vertical acceleration, or even higher order derivatives. It is unlikely that the visual system is merely
differentially sensitive to downwards and upwards vertical motion. In fact, this possibility is precluded by our results. For example, a superior sensitivity for downwards vertical motion would predict higher rates of responses consistent with the facing direction of the walker for the paired displays containing an inverted fragment 1 than for displays containing its upright version. The results reflected the opposite pattern. Instead, we speculate that the visual system is sensitive to the vertical acceleration exhibited by the swing phase fragments. As noted earlier, such variations in velocity along the motion path, or trajectory forms, have been proposed to facilitate the perception of a variety of events (e.g., Bingham et al., 1995).

In order to investigate the possible relationship between acceleration and the local inversion effect, we extended the use of our novel foot displays to a larger number of stimulus samples (foot motions of other animals) in Experiment 2. In this next experiment, we created paired fragment stimuli (as those presented to Group 2 in Experiment 1) from the foot motions of a human walker, human runner, cat, and pigeon and presented them at both upright and inverted orientations to observers in a direction discrimination task.

If acceleration carries the local motion-based inversion effect in biological motion perception, the size of the effect should vary with different animal foot motions which undoubtedly carry very different motion profiles. In both Troje and Westhoff (2006) and Chang and Troje (2008), there were no significant differences among the inversion effects for the different animals tested. Note however that in those experiments, the foot dots were presented within the context of a full walker. The paired fragment stimuli used in Experiment 1 display solely foot-specific motion information and may be more sensitive to any differences in orientation effects among different stimulus types.
4.3 Experiment 2

4.3.1 Methods

Participants

Twenty new naïve observers that ranged in age from 17 – 24 years (mean age of 18.8 years; 11 females, 9 males) participated in this experiment. All observers had normal or corrected-to-normal vision.

Stimuli and Apparatus

The stimuli of this experiment were derived from the motion of one foot belonging to a human walker (defined by the ankle point), human runner (defined by the ankle point), cat, and pigeon. The foot motions were shown at their veridical speeds corresponding to gait frequencies of 0.93 Hz, 1.3 Hz, 1.7 Hz, and 1.6 Hz for the human walker, human runner, cat, and pigeon, respectively. The motions were shown in sagittal view from either the right or left side. As in Experiment 1, the full gait cycle for each stimulus type (walker, runner, cat, pigeon) was sampled by specifying 10 start frames, spaced at equal intervals along the cycle, from which the fragment could begin. The length that each fragment extended from the start frame was 150 ms. Figure 4.3 depicts the 10 start frame positions along the full gait trajectory (A), and the individual trajectories of the 10 fragments (B), for each stimulus type. Note that for the human walker, starting positions were slightly different from those defined in Experiment 1.
Figure 4.3. The full gait trajectory with the 10 starting positions (A), and the individual 150 ms trajectories of the 10 fragments (B), for the foot of a human walker, human runner, cat, and pigeon. Trajectories shown here are taken from right-facing sequences.

For each stimulus type, five fragment pairs were created by pairing fragments labeled 1/6, 2/7, 3/8, 4/9, and 5/10. Each display lasted 1000 ms and consisted of 10 sets of a specified fragment pair that were randomly displaced within a 4.8 x 4.8 deg area in the centre of the screen. Each set had its own randomly chosen onset time and was visible only during the 150 ms time window. The stimuli were presented at both upright and inverted orientations. Each display subtended a visual angle of 4.8 x 4.8 deg. The apparatus was identical to that of Experiment 1.
Procedure

Stimuli were viewed binocularly at a distance of 60 cm as maintained by a chin-rest. As in Experiment 1, participants were asked to indicate the facing direction of the perceived entity (left or right).

The practice block comprised of 80 trials that consisted of left and right, and upright and inverted versions of the five fragment pairs for each of the four stimulus types. For the experiment proper, a total of 640 experimental trials were completed across two identical experimental blocks of 320 trials each. In a single block, all possible combinations of the four stimulus types, five fragment pairs, two orientations, and two directions were repeated four times. Within each repetition, the 80 possible stimuli were presented in random order. All other testing procedures were identical to Experiment 1.

4.3.2 Results

As in Experiment 1, the data were analyzed both in terms of the proportions of responses consistent with the facing direction of the walker from which the stimuli were derived and in terms of $d'$ sensitivity measures derived from these rates. Both analyses yielded identical results and the results are presented here in terms of the proportions of responses consistent with walking direction.

A comparison of the mean proportion of responses consistent with the facing direction of the walker for each stimulus type at both orientations, collapsed across fragment pairs is shown in Figure 4.4. An inspection of the response rates indicated that the rates were higher for the cat (upright = 0.64 proportion of responses; inverted = 0.58 proportion of responses) than for the human walker (upright = 0.61 proportion of responses; inverted = 0.47 proportion of responses),
human runner (upright = 0.66 proportion of responses; inverted = 0.42 proportion of responses), and pigeon (upright = 0.55 proportion of responses; inverted = 0.46 proportion of responses) stimuli. Overall, rates for the upright stimuli (0.62 proportion of responses) were higher than rates for the inverted stimuli (0.48 proportion of responses). As can be seen in Figure 4.4 however, the difference in rates between upright and inverted orientations varied with stimulus type: the difference was largest for the runner stimuli and smallest for the cat stimuli.

Figure 4.5 shows the mean proportion of responses consistent with walking direction of five fragment pairs at the two orientations separately for each stimulus type. From this figure, it is evident that for certain stimulus types, the difference in rates between upright and inverted orientations also depended on the particular fragment pair.

![Graph showing mean proportion of responses for different types of stimuli in upright and inverted orientations.](image)

Figure 4.4. The mean proportion of responses consistent with the facing direction of the walker for each stimulus type at both orientations, collapsed across fragment pairs. Response rates above and below 0.5 indicate larger proportions of responses that were consistent, \( C \), and inconsistent, \( I \), with walking direction, respectively. Error bars represent ± 1 standard error of the mean.
Figure 4.5. The mean proportion of responses consistent with the facing direction of the walker for five fragment pairs at the two orientations for the human walker (A), human runner (B), cat (C), and pigeon (D). Response rates above and below 0.5 indicate larger proportions of responses that were consistent, $C$, and inconsistent, $I$, with walking direction, respectively. Error bars represent ± 1 standard error of the mean.

The above observations are supported by statistical analyses. The data were first entered in a 2 (block) x 4 (type) x 2 (orientation) ANOVA. The analysis showed significant main effects of stimulus type, $F(3, 57) = 12.60$, $p < 0.001$, orientation, $F(1, 19) = 143.95$, $p < 0.001$, and a significant type x orientation interaction, $F(3, 57) = 21.00$, $p < 0.001$. As the performance between the two test blocks did not differ, $F(1, 19) = 0.04$, $p = 0.851$, data from the two blocks were pooled for further analyses.

Tukey’s post-hoc comparisons of the different stimulus types showed that overall, the proportions of responses consistent with walking direction were higher for the cat stimuli than for
all other stimulus types \( (p < 0.001 \text{ for all}) \). Average rates for the walker, runner and pigeon did not differ \( (p > 0.200 \text{ for all}) \).

Individual comparisons of the two orientations for each stimulus type indicated that the rates for upright stimuli were higher than for inverted stimuli for all stimulus types (Bonferroni-corrected \( t \) tests, \( p < 0.002 \) for all). The type x orientation interaction was subsequently analyzed with a one-way ANOVA on difference scores, obtained for each participant by subtracting rates for the inverted stimuli from those of the upright stimuli. The analysis showed a significant main effect of stimulus type, \( F(3, 57) = 21.00, p < 0.001 \). Tukey’s post-hoc comparisons indicated that the inversion effect, measured in terms of the difference between rates for the upright and inverted stimuli, was larger for the runner stimuli than all other types \( (p < 0.001 \text{ for all}) \). The inversion effects for the walker and pigeon stimuli did not differ \( (p = 0.188) \) but the inversion effect for the walker was significantly larger than the one for the cat \( (p = 0.006) \). The comparison between the inversion effects for the pigeon and cat was not significant \( (p = 0.507) \).

Individual two-way ANOVAs were used to analyze the different fragment pairs between the two orientations for each of the four stimulus types. The analysis for the human walker revealed significant main effects of orientation, \( F(1, 19) = 73.00, p < 0.001 \), and pair, \( F(4, 76) = 113.07, p < 0.001 \), and a significant orientation x pair interaction, \( F(4, 76), p = 0.029 \). Tukey’s comparisons of the different pairs indicated that while response rates for pairs 1/6 and 5/10 did not differ \( (p = 0.798) \), they were significantly lower than the rates for all other pairs \( (p < 0.001 \text{ for all comparisons}) \). The response rate for pair 2/7 was not different from the rate for pair 3/8 \( (p = 0.277) \), but was significantly higher than the rate for pair 4/9 \( (p < 0.001) \). The comparison between pairs 3/8 and 4/9 was not significant \( (p = 0.089) \). The orientation x pair interaction was analyzed with Tukey’s comparisons of upright and inverted orientations for each pair. The
analyses revealed that the rates for upright stimuli were higher than those for inverted stimuli for pairs 2/7, 3/8, 4/9, and 5/10 (p < 0.001 for all), but not pair 1/6 for which the two orientations did not differ (p = 0.113).

A comparable ANOVA for the runner showed significant main effects of orientation, \( F(1,19) = 75.37, p < 0.001 \), and pair, \( F(4,76) = 35.51, p < 0.001 \), but no interaction. Tukey’s comparisons of the different fragment pairs showed that while the rates for pairs 2/7, 3/8, and 4/9 did not differ (p > 0.800 for all), they were significantly higher than the rates for pairs 1/6 and 5/10 (p < 0.001 for all), which did not differ (p = 0.146).

Similarly, the ANOVA for the cat revealed significant main effects of orientation, \( F(1,19) = 13.81, p = 0.002 \), and fragment pair, \( F(4,76) = 120.79, p < 0.001 \), and no interaction. Here, Tukey’s comparisons of the different fragment pairs revealed that the rates for pairs 2/7, 3/8, and 4/9 did not differ (p > 0.600 for all comparisons), but were significantly higher than the rates of pairs 1/6 and 5/10 (p < 0.001 for all). The rate for pair 1/5 was also significantly higher than the rate for pair 5/10 (p < 0.001).

The ANOVA for the pigeon showed significant main effects of orientation, \( F(1,19) = 48.89, p < 0.001 \), and fragment pair, \( F(4,76) = 61.25, p < 0.001 \), and a significant orientation x pair interaction, \( F(4,76) = 19.68, p < 0.001 \). Tukey’s comparisons of the different fragment pairs indicated that while the rates for pairs 1/6, 2/7, and 3/8 did not differ (p > 0.200 for all), they were significantly higher than the rates of pairs 4/9 and 5/10 (p < 0.001 for all). In addition, the rate for pair 4/9 was significantly higher than the rate for pair 5/10 (p < 0.001). The interaction was analyzed with Tukey’s tests of upright versus inverted orientations per each fragment pair. The analyses revealed that the rates were higher for upright versions than for inverted versions of
pairs 3/8, 4/9, and 5/10 ($p < 0.030$ for all). The two orientations did not differ for pairs 1/6 ($p = 0.092$) and 2/7 ($p = 0.274$).

Finally, in Table 4.1, we calculated the average vertical accelerations both in real world coordinates and screen coordinates, and a measure of the inversion effect for each of the fragment pairs that comprised the stimuli in this experiment. Here, the inversion effect is quantified by a difference score, computed by subtracting the proportion of responses consistent with walking direction for the inverted stimulus from the proportion of responses consistent with walking direction for the upright stimulus.

<table>
<thead>
<tr>
<th>Type</th>
<th>Fragment pair</th>
<th>Real world vertical acceleration (m/s²)</th>
<th>Screen vertical acceleration (deg/s²)</th>
<th>Inversion score</th>
</tr>
</thead>
<tbody>
<tr>
<td>human walker</td>
<td>1 / 6</td>
<td>0.03↑ / -1.35↓</td>
<td>0.10↑ / -4.01↓</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>2 / 7</td>
<td>9.45↓ / 0.21↑</td>
<td>27.95↓ / 0.63↑</td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td>3 / 8</td>
<td>0.06↓ / -0.47↑</td>
<td>0.19↓ / -1.39↑</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>4 / 9</td>
<td>-4.16↓ / -1.06↑</td>
<td>-12.30↓ / -3.13↑</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>5 / 10</td>
<td>0.24↓ / -3.45↑</td>
<td>0.71↓ / -10.20↑</td>
<td>0.14</td>
</tr>
<tr>
<td>human runner</td>
<td>1 / 6</td>
<td>4.66↑ / -1.67↓</td>
<td>13.79↑ / -4.95↓</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>2 / 7</td>
<td>8.56↑ / -3.79↑</td>
<td>25.32↑ / -11.21↑</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td>3 / 8</td>
<td>2.05↓ / -4.17↑</td>
<td>6.06↓ / -12.33↑</td>
<td>0.33</td>
</tr>
<tr>
<td></td>
<td>4 / 9</td>
<td>-4.54↓ / -1.32↑</td>
<td>-13.43↓ / -3.92↑</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td>5 / 10</td>
<td>-1.59↓ / 1.82↑</td>
<td>-4.72↓ / 5.39↑</td>
<td>0.22</td>
</tr>
<tr>
<td>cat</td>
<td>1 / 6</td>
<td>0.77↑ / -2.65↓</td>
<td>2.28↑ / -7.85↓</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>2 / 7</td>
<td>2.34↑ / -0.16↑</td>
<td>6.94↑ / -0.48↑</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>3 / 8</td>
<td>1.09↑ / 0.53↑</td>
<td>3.23↑ / 1.58↑</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>4 / 9</td>
<td>2.26↑ / 0.02↑</td>
<td>6.68↑ / 0.05↑</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>5 / 10</td>
<td>-1.24↓ / -2.66↑</td>
<td>-3.68↓ / -7.88↑</td>
<td>0.11</td>
</tr>
<tr>
<td>pigeon</td>
<td>1 / 6</td>
<td>-4.13↓ / 6.33↑</td>
<td>-12.22↓ / 18.73↑</td>
<td>-0.05</td>
</tr>
<tr>
<td></td>
<td>2 / 7</td>
<td>0.65↑ / 20.41↑</td>
<td>1.92↑ / 60.41↑</td>
<td>-0.04</td>
</tr>
<tr>
<td></td>
<td>3 / 8</td>
<td>-0.95↑ / 5.87↑</td>
<td>-2.81↑ / 17.36↑</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td>4 / 9</td>
<td>-4.54↑ / -6.99↓</td>
<td>-13.43↑ / -20.69↓</td>
<td>0.23</td>
</tr>
<tr>
<td></td>
<td>5 / 10</td>
<td>-8.74↑ / -6.98↓</td>
<td>-25.86↑ / -20.65↓</td>
<td>0.07</td>
</tr>
</tbody>
</table>

Table 4.1. Average vertical accelerations and inversion score for each fragment pair. Sign of acceleration value indicates direction of acceleration. Positive and negative acceleration values indicate downwards and upwards accelerations, respectively. Arrows indicate vertical direction of the motion. Inversion score was computed by subtracting the proportion of responses consistent with the facing direction of the walker from which the stimulus was derived for the inverted orientation from that for the upright orientation.
We subsequently compared the inversion score for each stimulus to the absolute difference in vertical acceleration between its two constituent fragments by means of a linear regression analysis. Figure 4.6 shows the regression analyses for the full stimulus set and for the stimulus set with two outliers (open circles) removed. Although no relationship is evident when the full set is considered ($r^2 = .01; p = 0.659$), a significant positive correlation between the inversion score and absolute difference in vertical acceleration becomes evident once the outliers are removed ($r^2 = 0.35; p = 0.009$).

Figure 4.6. A comparison of the inversion score and the absolute difference in vertical acceleration between the two constituent fragments of each stimulus. The dotted and solid lines represent separate linear regression fits for the full data set and the data set with two outliers (open circles) omitted, respectively.
4.3.3 Discussion

The overall response rates for the walker stimuli in this experiment are very similar to the response rates obtained for the 150 ms fragment pairs in Experiment 1. The results here extend upon the findings of Experiment 1 by showing that direction can also be retrieved from fragments of foot motions of a human runner and other animals. For all stimulus types, this ability depends on the orientation of the stimulus.

Interestingly, although the cat stimuli appear to be the most informative for direction retrieval overall, they carry the smallest inversion effect in comparison to all other stimulus types. As can be seen from Table 4.1, the cat’s foot motion carries the smallest vertical accelerations as compared to all stimulus types.

The inversion effect for the cat stimuli is generally quite small (score of ~0.05 for pairs 1/6, 2/7, 3/8, and 4/9). In contrast, the runner’s foot motion carries very pronounced vertical accelerations, particularly fragments 2 and 3 of pairs 2/7 (score = 0.25) and 3/8 (score = 0.33), respectively. Correspondingly, we found here that the runner stimuli carried the largest inversion effect overall as compared to all other stimulus types. Additionally, as shown in Figure 4.6, a positive relationship is evident between the inversion effect and the absolute difference in vertical acceleration between the two constituent fragments of each stimulus. These findings are consistent with our hypothesis that the local-motion based inversion effect in biological motion perception is carried by the vertical acceleration contained in foot motions.

The finding that different stimulus types are differentially effective for direction retrieval and have differentially-sized inversion effects is not consistent with results reported for full-walker scrambled displays (Troje & Westhoff, 2006; Chang & Troje, 2008). In these studies, direction discrimination ability did not differ among the various animal types. As noted earlier
however, this apparent discrepancy may be due to the fact that the stimuli in the current study display only foot-specific motion information and are thus more sensitive measures that allow the differences between the various animals to be teased out.

Considered together, the results of this experiment suggest strongly a relationship between acceleration and the local-motion based inversion effect and are congruent with previous suggestions that variations in velocity are evaluated in the perception of both animate and inanimate events (Bingham et al., 1995). In Experiment 3, we tested the importance of acceleration for the local inversion effect by manipulating the velocity profile of the foot motion. Specifically, we presented observers with the natural human walker stimuli of the present experiment, and with stimuli that were matched spatiotemporally to the natural stimuli, but moved with constant speeds along the trajectory paths. If in fact, acceleration carries the local inversion effect in biological motion perception, there should be no difference in performances between the upright and inverted versions of the constant speed stimuli which contain no acceleration along the specified paths.

4.4 Experiment 3

4.4.1 Methods

Participants

A new group of 20 naïve observers that ranged in age from 17 – 23 years (mean age = 18.3 years; 12 females, 8 males) participated in this experiment. All observers had normal or corrected-to-normal vision.
Stimuli and Apparatus

Two sets of stimuli were employed in this experiment. The first set comprised of the paired fragments derived from the natural foot motion of the human walker in Experiment 2 (Figure 4.3, human walker). The second set comprised of stimuli derived from fragments that were identical to the natural walker stimuli, but had accelerations removed; that is, each individual dot moved along a path identical to that defined previously but with a constant speed equal to the average speed of the dot with the specified fragment. Note that this manipulation renders only the speed along the path to be constant and does not preclude the possibility that there may be a change in velocity between the beginning and the end of the motion path. That is, in order to achieve the same trajectory paths of the natural stimulus, some acceleration must still be present in the constant speed stimulus due to changes in the direction of the velocity vector. Importantly, this manipulation disrupts the natural velocity profile of the foot motion and in particular, eliminates the vertical asymmetries caused by gravitational acceleration. For each fragment pair then (i.e., 1/6, 2/7, 3/8, 4/9, 5/10), two versions served as the basis for the stimuli in this experiment: one that contained natural acceleration and one that contained no acceleration along the specified paths.

As in Experiments 1 and 2, each display lasted 1000 ms in duration and consisted of 10 sets of a specified fragment pair that were randomly positioned within a 4.8 x 4.8 deg area in the centre of the screen. Each set had its own randomly chosen onset time and was visible only during the 150 ms time window. The stimuli were presented at both upright and inverted orientations. Each display subtended a visual angle of 4.8 x 4.8 deg. The apparatus was identical to that of Experiments 1 and 2.
Procedure

As in Experiments 1 and 2, participants were asked to indicate the facing direction of the perceived entity (left or right). The practice block comprised of 40 trials that consisted of left and right, and upright and inverted versions of the five fragment pairs belonging to each of the two stimulus types (natural and constant).

The experiment proper consisted of 400 trials. In the experimental block, all possible combinations of the two stimulus types, five fragment pairs, two orientations, and two directions were repeated 10 times. Within each repetition, the 40 possible stimuli were presented in random order. All other testing procedures were identical to Experiment 2.

4.4.2 Results

As in Experiments 1 and 2, the data were analyzed both in terms of the proportions of responses consistent with the facing direction of the walker from which the stimuli were derived and in terms of $d'$ sensitivity measures derived from these rates. Again, both analyses yielded identical results. As in Experiment 1 and 2, the results are presented in terms of proportions of responses consistent with walking direction.

The mean proportion of responses consistent with walking direction of the two stimulus types for the five fragment pairs at upright and inverted orientations are presented in Figure 4.7. An examination of the mean response rates for upright versus inverted orientations collapsed across stimulus type and fragment pair revealed that in general, the rates for upright stimuli (0.56 proportion of responses) were higher than the rates for inverted stimuli (0.52 proportion of responses). However, a closer inspection of the means separately for the two stimulus types reveals that the rates for upright displays were higher than the rates for the inverted displays for the natural stimuli only.
Response rates were analyzed with a 2 (type) x 2 (orientation) x 5 (fragment pair) repeated-measures ANOVA that showed a significant main effect of orientation, $F(1, 19) = 7.15$, $p = 0.015$, a significant main effect of fragment pair, $F(4, 76) = 28.33$, $p < 0.001$, and a significant type x orientation interaction, $F(1, 19) = 13.11$, $p = 0.002$. The main effect of type and other interactions were not significant.

Tukey’s post-hoc comparisons of the different fragment pairs, collapsed over the two stimulus types and two orientations showed that while the rates for pairs 2/7, 3/8, and 4/9 did not differ ($p > 0.500$ for all comparisons), they were significantly higher than the rates for pairs 1/6 and 5/10 ($p < 0.001$ for all), which did not differ ($p = 0.219$).

Figure 4.7. The mean proportion of responses consistent with the facing direction of the walker for the natural walker and constant speed stimuli across the five fragment pairs and two orientations. Response rates above and below 0.5 indicate larger proportions of responses that were consistent, $C$, and inconsistent, $I$, with walking direction, respectively. Error bars represent $\pm 1$ standard error of the mean.
The type x orientation interaction was analyzed with Bonferroni-corrected, two-tailed \(t\) tests comparing upright and inverted orientations for each stimulus type. The analyses revealed that the rates were higher for upright displays than for inverted displays for the natural stimuli only (\(p < 0.001\); upright = 0.58 proportion of responses; inverted = 0.49 proportion of responses). Upright and inverted versions of the constant speed stimuli did not differ (\(p = 0.335\); upright = 0.53 proportion of responses; inverted = 0.55 proportion of responses).

### 4.4.3 Discussion

The pattern of results for the natural human walker stimuli in this experiment is similar to that obtained for the same set of stimuli in Experiment 2. Importantly, the results obtained here for constant speed stimuli that are otherwise matched to the natural stimuli reveal a complete absence of an inversion effect. This contrast is critical as the two sets of stimuli are identical in all respects (e.g., identical average speeds), and differ only in terms of the presence (natural stimuli) or absence (constant stimuli) of acceleration along the motion path. This finding suggests that the inversion effect associated with the local foot dot motions is carried by the acceleration contained in these motions.

Consistent with the results of Experiment 2, the fragment pairs that were most informative for the natural stimuli were pairs 2/7, 3/8, and 4/9. Interestingly, an inspection of the means of the same fragment pairs (2/7, 3/8, and 4/9) for the constant speed stimuli reveals that observers were still able to derive the direction that was consistent with the facing direction of the walker from which the stimuli were derived. In fact, the mean response rates for these pairs are very similar to the mean rates for the corresponding pairs of the natural walker stimuli when collapsed across the two orientations. One possibility is that this pattern may be a result of simple
speed effects. That is, for both sets of stimuli, perceived direction may be determined by the higher-speed fragments within each pair. Indeed, in pairs 2/7, 3/8, and 4/9, the facing direction of the display is signaled by the higher-speed fragments. Importantly however, the results show an inversion effect for the natural walker stimuli only. As the two sets of stimuli differ only in terms of the presence or absence of acceleration along the motion path, we conjecture that acceleration accounts for the difference in orientation-specificity observed for the data of the natural and constant speed stimuli.

The results of this experiment show clearly that the visual system exploits the accelerations contained in the natural walker stimuli. Moreover, these accelerations must be assessed in accordance to some framework that renders them anisotropic. These findings suggest that the responsible mechanisms are impressively sensitive to the acceleration – able to retrieve it from our very limited 150 ms trajectories.

4.5 General Discussion

Using novel biological motion stimuli that display solely foot-specific information, we showed that the cues that carry the local inversion effect in biological motion perception cannot be retrieved independently from a single local motion of the foot but can be exploited from displays derived from two foot dots that trace counterphase positions of the gait cycle (Experiment 1). Extending our findings from human walker stimuli, we showed that displays derived from other foot types also provide information about facing direction and carry an inversion effect. Importantly, the inversion effect was largest for the runner stimuli which exhibit large vertical accelerations and smallest for the cat stimuli, which carry comparatively minor vertical accelerations (Experiment 2). We showed further a linear correlation between the
inversion effect and the absolute difference in vertical acceleration between the two constituent fragments of each stimulus. Finally, we showed that if acceleration is removed from stimuli that are otherwise identical to the natural walker stimuli, the inversion effect disappears (Experiment 3). Together, these results suggest that the local inversion effect in biological motion perception relies on accelerations contained in the foot motions. These findings are congruent with previous demonstrations of the important role of velocity profiles for the perception of animate and inanimate events (Bingham et al., 1995).

Although we have provided evidence to suggest that the inversion effect is carried by the velocity profile of the foot motion, it remains unclear as to the specific role of such cues for achieving the facing direction of a display. We conjecture that horizontal (left-right) and vertical (up-down) asymmetries of the trajectories of the foot motion play very different roles in the manner in which they contribute to the results of the present experiments, and in general, to the perception of biological motion. While information regarding the facing direction of the walker must be encoded somehow in horizontal asymmetries, the inversion effect must be carried by vertical asymmetries in the stimulus. As noted earlier, these asymmetries could in principle be due to a number of different attributes, including trajectory shape, the direction of the motion along the trajectory, and many other parameters inherent to the velocity profile of the dot moving along the trajectory. Trajectory shape is likely not a valid cue for the inversion effect as we employed a variety of different shapes in Experiment 2. Furthermore, we noted earlier that the present data is also able to discount the polarity of vertical velocity as a cue for the inversion effect (see discussion to Experiment 1). Thus, although there are still even higher-order motion derivatives that could be considered, the next most parsimonious cue for the inversion effect lies with the polarity of the vertical acceleration carried by the trajectory. How do vertical
asymmetries such as those carried by accelerations then contribute to the observers’ ability to retrieve facing direction from stimuli in the present experiments?

Note that only horizontal asymmetries can contain information about the facing direction of the walker. The amount of such information remains the same if the stimulus is inverted (i.e., flipped about the horizontal axis). However, the visual system's ability to make use of this information seems to depend on the orientation of the stimulus and is thus affected by the inversion of the display.

We suggest that the presence of vertical motion that is consistent with a mass being propelled away from the ground and then subsequently pulled back by gravity first flags an event as being animate. Only once the event is deemed to be animate does our visual system attribute a facing direction to it. Recall that there is no net translation in our stimuli. The intended locomotion in one direction is always countered by the inverse motion of the invisible belt of a treadmill or the panning motion of a camera. Overall the stimulus remains stationary. The remaining intrinsic motion can only reveal the intention to move into a certain direction; that is, the facing direction. We conjecture that the visual system attributes such intentionality only to something that indicates a degree of animacy, and uses the vertical motion profile to validate a trajectory as being animate. Thus, the vertical asymmetries may determine the validity of a trajectory for further analysis as an animate being while the horizontal asymmetries may determine the facing direction of valid trajectories.

In order to appreciate the significance of current findings that demonstrate visual sensitivity to velocity profiles (i.e., accelerations) contained in biological motion, it is worth considering findings in the literature for the perception of acceleration with non-biological stimuli. Human observers seem to be remarkably inept at assessing accelerated motion (e.g.,
Schmerler, 1976; Calderone & Kaiser, 1989; Babler & Dannemiller, 1993). For example, observers have a tendency to perceive uniform acceleration despite sudden changes in the velocity of the stimulus (Burke, 1952). In one direct investigation of the perception of accelerated motion, Schmerler (1976) measured observers’ ability to detect accelerating and decelerating motion as a function of the rate of acceleration/deceleration and the ratio of terminal to initial velocity. On each trial, the observers’ task was to classify the motion of the stimulus (a horizontal-moving ball) as accelerating, decelerating, or moving with constant velocity. Judgement accuracies varied only slightly for the varying acceleration/deceleration rates, and at all rates, dropped below chance when the ratio of terminal to initial velocity was less than approximately 3. From his data, Schmerler concluded that in order for motion change to be detected 50% of the time, 230% – 320% increases or decreases in velocity are necessary. In consideration of our results reported here, it is intriguing that despite human observers’ seeming ineptness at judging accelerated motion explicitly, accelerations are in fact processed by the visual system, as evidenced by the results of Experiment 3.

Recall that in Experiment 3, the natural walker stimuli resulted in an inversion effect but the constant speed stimuli did not. The only difference between the two sets of stimuli was the presence (natural stimuli) or absence (constant speed stimuli) of accelerations suggesting that the accelerations contained in the natural stimuli must be exploited by the visual system. According to Table 4.1, accelerations for the walker stimuli are most pronounced during the swing phase fragments 2 and 3. A closer examination of the motions of fragments 2 and 3 however revealed that their respective ratios of terminal to initial velocities are -1.05 (negative denotes change in direction) and 1.02, respectively. That is, the final and initial velocities for these two fragments are almost identical. Although we cannot conclude if observers explicitly detected
acceleration/deceleration from these fragments, these findings raise the possibility that accelerations contained in the swing phase fragments are interpreted by the visual system independently of whether they can be explicitly detected by the observer psychophysically. Still, in reviewing the apparent conflict between observer sensitivity to visual acceleration and sensitivity to trajectory form, Muchisky and Bingham (2002) noted that the ability to detect variation in velocity along a trajectory does not require sensitivity to acceleration. They argued that changes in velocity over space, rather than time, could be exploited by the observer instead.

In another study of acceleration perception with non-biological stimuli, Calderone and Kaiser (1989) presented a stimulus of a single square that moved either up, down, left, or right while accelerating or decelerating. The observers’ task was to decide if the square did or did not accelerate/decelerate. Interestingly, for horizontal motions, decelerations were easier to detect while for vertical motions, accelerations were easier to detect. Of particular relevance, thresholds did not differ within the principle axes (upwards versus downwards or leftwards versus rightwards). As noted by Calderone and Kaiser, this last finding is surprising given the abundance of objects in the visual environment that often accelerate while moving downwards and decelerate while moving upwards.

The results presented in our current study suggest that the visual system is sensitive to the orientation (direction) of accelerating biological motions. That is, the vertical accelerations exhibited by the local foot motions must be interpreted according to some framework that renders them anisotropic. The leg muscles are almost inactive during the swing phase of the gait cycle (e.g., Crowninshield & Brand, 1981), and it has been suggested that the foot moves entirely under the influence of gravity during this phase (Mochon & McMahon, 1980). We conjecture that the visual system is sensitive to the accelerations that result as the foot interacts with gravitational
force. Specifically, we proposed earlier that the visual system evaluates accelerations according to a gravitational framework in order to validate trajectories as being animate.

The use of gravitational heuristics for visual perception has been reported in the past both for inanimate dynamic events (e.g., Pittenger, 1985; Stappers & Waller, 1993; McConnell, Muchisky, & Bingham, 1998) and for the perception of biological motion (e.g., Runeson & Frykholm, 1981; Jokisch & Troje, 2003; Shipley, 2003). For example, Pittenger (1985) showed that observers could retrieve the length of a swinging pendulum if informed about its period. In another study, Stappers and Waller (1993) presented observers with animations of a fountain on a video background and asked observers to adjust the magnitude of gravity to make the display appear natural. They found that observers were well able to match gravitational acceleration with apparent depth. Gravity-based interpretations have also been found for the perception of biological motion. Runeson and Frykholm (1981) found that observers were able to judge a box’s mass from a point-light sequence of an actor lifting a box. They reasoned that observers exploited the relationship between postural changes and the acceleration of the box using dynamic principles. Jokisch and Troje (2003) provided further evidence to suggest that the visual system exploits a fixed relationship between spatial and temporal parameters that exists in a gravity-based environment by showing that observers, when presented with point-light displays of dogs walking with varying frequencies, judged the size of the dog to be smaller at high stride frequencies than at low frequencies. In another study, Shipley (2003) presented a point-light figure walking on his hands in upright and inverted orientations and found that walker detection was better for the upright display which contained familiar dynamic relations.

Although the studies summarized above clearly implicate the use of assumptions about gravity, they do not reveal the reference systems in which these assumptions are employed.
Conceivably, orientation effects for biological motion perception could be described according to either observer-centered (e.g., retinal) or allocentric (e.g., gravitational) coordinates. To this end, Bingham et al. (1995) found that the recognition of point-light events was stronger for upright displays than for inverted displays regardless of whether observers were oriented upright or rotated by 180 degrees. The authors therefore concluded that point-light events are perceived in relation to the gravitational field rather than the relative orientation of the observer and the stimulus. These results however appear to contrast those reported by Troje (2003) which showed no advantage for detecting a walker in displays that were oriented upright and presented to rotated observers than in displays that were rotated and presented to upright observers. Instead, performance only depended on whether the display was aligned with the observer regardless of their orientation in space suggesting that orientation constraints are in fact egocentric. As such, the exact coordinates according to which orientation effects for biological motion perception should be described (whether based upon global or local cues) requires further clarification.

Finally, the individual fragments that comprised both sets of displays in Experiment 3 extended only 150 ms in length, suggesting that the accelerations can be exploited by the visual system with very limited information. Indeed, we showed in Experiment 1 (paired fragments) that the inversion effect was still present (although comparatively smaller) for displays comprised of fragments extending just 100 ms in length. Interestingly it has been shown that accurate acceleration detection requires integration times of approximately 100 ms (Werkhoven, Snippe, & Toet, 1992). What is the threshold temporal window for retrieving acceleration cues contained in local biological motions? An answer to this question requires further empirical work.

In sum, it has become evident that at least two inversion effects need to be considered in biological motion perception. While the well-demonstrated inversion-related impairment in
perceiving coherently organized displays may be attributable to impaired global processing (e.g., Pavlova, 1989; Bertenthal & Pinto, 1994), there is another inversion effect which relies on the local foot motions (Troje & Westhoff, 2006). Here, we have provided evidence to suggest that this latter inversion effect is carried by vertical accelerations exhibited by foot motions.
References


Chapter 5

Is Acceleration Sensitivity Size Invariant?

5.1 Introduction

Human observers are remarkably inept at judging accelerated motion. For example, observers have a tendency to perceive constant acceleration despite sudden changes in the velocity of the stimulus (e.g., Burke, 1952). When presented with a moving target that subsequently disappears behind a mask, observers are less able to extrapolate (i.e., estimate when the target will pass behind a marker) from accelerated motion as compared to constant velocity motion (Rosenbaum, 1975). Still, it has been demonstrated that acceleration plays a critical role for visual perception of animate motions (e.g., Runeson & Frkyholm, 1981; Jokisch & Troje, 2003; Chang & Troje, 2009). It is of particular interest then to achieve a better understanding of human sensitivity to visual acceleration.

Acceleration is defined as the change in velocity per unit time. Most studies of acceleration perception, however, characterize acceleration in terms of the overall change in velocity. For example, acceleration sensitivity has been measured in terms of the ratio of final to initial velocity (Schmerler, 1976) or the proportion of change in velocity relative to the average velocity (Regan, Kaufman, & Lincoln, 1986; Calderone & Kaiser, 1989). The authors of these studies argue that measuring acceleration as it is defined (the change in velocity per unit time) is inappropriate as sensitivity is likely to depend on factors such as the duration of acceleration, the initial velocity of the object, and the size or distance of the trajectory (e.g., Calderone & Kaiser,
Yet, a systematic investigation of how such factors affect acceleration sensitivity has yet to be undertaken.

Despite the variable methods of parameterizing acceleration, the acceleration perception literature yields many insights into human sensitivity to nonuniform motion. In an early study of acceleration perception, Gottsdanker, Frick, and Lockard (1961) asked observers to discriminate a negatively or positively accelerated stimulus from a constant velocity stimulus that was equated for mean velocity. Various ranges of velocity, acceleration, and presentation times were tested. The results indicated that discrimination performances (expressed in terms of proportion of correct responses) for the various conditions decreased with increases in mean velocity and decreases in presentation time. Moreover, for each presentation time tested, thresholds (expressed in terms of change in velocity per time) were estimated by linear interpolation to find the acceleration yielding 75% correct responses. Interestingly, thresholds obtained in this manner were better for negative acceleration than for positive acceleration. The authors subsequently examined the thresholds relative to the velocity of the motion both in terms of a percentage of change per unit time (threshold acceleration/mean velocity) or as a percentage of change in the viewing period (change in velocity/mean velocity). Critically, the authors showed that for a given mean velocity, the thresholds agreed better across different presentation times when expressed in terms of relative change of velocity rather than in terms of relative acceleration. Thus, they argued that acceleration is detected by comparing earlier and later velocities rather than by direct sensing of acceleration over time.

In a subsequent study by Gottsdanker (1961), an explanation for the apparent difference between identifying positive and negative accelerations was advanced. Specifically, Gottsdanker proposed that this finding may be explained due to different consequences of the observer
missing the start of the motion. Missing the onset of motion means that low velocity segments are lost in the case of positive acceleration and high velocity segments are lost in the case of negative acceleration. Still, the effective difference between the initial and final velocities is reduced equally for both positively and negatively accelerated motion. Importantly, the mean velocity of the remaining motion available for judgment would be higher in the case of the positively accelerated motion. As it has been shown that judgment accuracies are based on the change of final to initial velocities relative to the mean velocity viewed, this ratio would be lower for positively versus negatively accelerated motion. Consequently, lower accuracy would be expected for positively accelerating motion relative to negatively accelerating motion. His findings supported this hypothesis. Target motions were presented either with a sudden onset or with a lead-in motion, and terminated stationary (visible) or disappearing behind a shield. The accuracies were equal for the positive and negative accelerations when a lead-in motion was present, but only when termination was visible.

A later study by Schmerler (1976) also reported comparable findings for acceleration perception showing that negative acceleration was significantly easier to perceive than positive acceleration. In this study, the effects of varying the rate of change in velocity (i.e., acceleration) and the ratio of final to initial velocity on the perception of positive and negative acceleration were investigated. On each trial, a ball emerged from one tunnel and disappeared into a second tunnel 2 ft away across the horizontal axis. Observers were asked to classify the motion of the ball as accelerating, decelerating, or moving with constant velocity. Thresholds, calculated from the frequencies of the three response categories for each velocity ratio, varied only slightly as a function of the rate of change in velocity (positive or negative), but varied with velocity ratio. Specifically, performance decreased with decreases in velocity ratio. Consistent with findings
reported by Gottsdanker et al. (1961), thresholds for negative acceleration were better than those for positive acceleration. Schmerler argued however that missing the onset of motion, as Gottsdanker et al. suggested, is not a satisfactory explanation for this phenomenon. For example, poorer thresholds for positive acceleration as compared to negative acceleration due to missed information at motion onset, as put forth by Gottsdanker et al., also predict that there should be a difference between the ease of perceiving different accelerations. As less motion would be missed at slow (positive) accelerations than at fast accelerations, the mean velocity available for judgment would be higher at fast accelerations than at slow accelerations. The results from Schmerler’s study showed that this was not the case. He argued instead that the perception of nonuniform motion involves assimilatory tendencies, and as a result, favors perceived homogeneity. He suggested that assimilation depends on the preceding and following segments of motion and proceeds in the direction of the sign change. That is, in the case of positive acceleration, the preceding stimulus increases in intensity and seems more similar to the comparison stimulus that follows. By contrast, in the case of velocity decreases, the preceding stimulus decays and appears more similar to the comparison stimulus. Schmerler suggested that the assimilation model can account for differences in perceiving positive and negative acceleration if velocity changes are considered between equidistant, rather than temporally equal segments as velocity changes, when considered between equidistant segments, are larger when acceleration is negative than when acceleration is positive. Thus, the difference in the ease of perceiving positive and negative accelerations may be due to the fact that the change which is assimilated per unit distance is smaller in the case of positive acceleration than in the case of negative acceleration.
Regardless of the explanation for the difference in perceiving positive and negative acceleration reported by Gottsdanker et al. (1961) and Schmerler (1976), both studies involved only the use of horizontal motion. How does human sensitivity to acceleration along the vertical axis compare to sensitivity along the horizontal axis? To this end, Calderone and Kaiser (1989) examined acceleration sensitivity while taking into account both acceleration sign and motion orientation. In this study, a small square served as the target and the beginning and ending of its trajectory were occluded by a mask. On each trial, observers were instructed to indicate whether or not the stimulus accelerated (or decelerated). Observers were tested on two blocks of trials (positive or negative acceleration) where each block consisted of three sets of trials, one for each of three average velocities (0.7 – 1.7 deg/s). Within each set of trials, the four directions of motions were tested (up, down, left, right). The results indicated that thresholds, expressed as the proportion of change in velocity relative to the average velocity, did not differ among the various average velocities. Notably however, there was an interaction between the axis of motion and the sign of acceleration. Specifically, thresholds were better for negative acceleration than positive acceleration for horizontal motion. Conversely, thresholds were better for positive acceleration than negative acceleration for motion along the vertical axis. Although the authors offered no explanation for this phenomenon, they noted that their data question models of acceleration perception that propose a comparison of early and late velocity segments and explanations involving missed initial segments of motion. Such models predict that thresholds should differ across varying velocities since at higher velocities failure to attend to velocity information for a fixed interval would affect detection thresholds more. The consistent thresholds observed across the varying velocities in this study are thus incongruent with these models.
There was no difference for detecting acceleration between upward and downward motions in the study by Calderone and Kaiser (1989). The authors noted that this is a surprising finding given the propensity of objects in a gravity-driven environment to accelerate when moving downward and decelerate when moving upward. It has been shown that when presented with a free falling object, observers report the perception of constant velocity motion (Shanon, 1976). Interestingly however, observers, when shown displays of objects in free-fall, are able to retrieve the object’s size, distance, and landing location (Babler & Dannemiller, 1993; Hecht, Kaiser, & Banks, 1996; McConnell, Muchisky, & Bingham, 1998). Considered together, the literature suggests that the human visual system is in fact sensitive to acceleration, although this sensitivity may be better accounted for in terms of retrieving information such as object size and distance, or deviation from natural motion.

Sensitivity to velocity changes has also been proposed to underlie the recognition of visual events as depicted by point-lights (Bingham, Schmidt, & Rosenblum, 1995). In their study, Bingham, Schmidt, and Rosenblum (1995) presented observers with point-lights depicting the motions of events of varying dynamics (e.g., biodynamics, hydrodynamics, etc.). The task was to describe the events with free responses or by circling properties from a list. Results showed that observers were well able to identify inanimate and animate events from their motions. The authors concluded that underlying dynamics play a role in the perceptual significance of motions and that variations in velocity of the motion along its path, termed \textit{trajectory form}, contribute to event identification (see also Runeson, 1974; Bingham, 1987).

Additional evidence that the visual system is sensitive to gravitational acceleration comes from tasks requiring interpretations of spatial and temporal parameters (e.g., Pittenger, 1985; Stappers & Waller, 1993; Shiple, 2003). For example, Pittenger (1985) showed that observers
could estimate the length of a swinging pendulum if visual information about its period is
provided. In another study, Stappers and Waller (1993) presented observers with animations of a
fountain on a video background and asked observers to adjust the magnitude of gravity to make
the display appear natural. The authors found that observers were well able to match
gravitational acceleration with apparent depth. Gravity-based interpretations have also been
found for the perception of biological motion. Runeson and Frykholm (1981) found that
observers were able to judge a box’s mass from a point-light display of an actor lifting a box and
reasoned that observers exploited the relationship between postural changes and the acceleration
of the box using dynamic principles. Jokisch and Troje (2003) provided further evidence to
suggest that the visual system exploits a fixed relationship between spatial and temporal
parameters that exists in a gravity-based environment by showing that observers, when presented
with point-light displays of dogs walking with varying frequencies, judged the size of the dog to
be smaller at high stride frequencies than at low frequencies. These studies suggest that observers
use knowledge about the direction and effects of gravity when interpreting biological and other
inanimate dynamic events.

Given the numerous demonstrations of the ability of the visual system to sense visual
acceleration and the significance of employing knowledge about acceleration for interpreting
animate and inanimate events, a critical question is whether sensitivity to acceleration is size
invariant. This question has important real-world and empirical implications. For example, if
sensitivity to acceleration is not size invariant, the size of an agent in the visual environment
would affect our ability to retrieve important cues from biological motion.

Here, we investigated whether acceleration sensitivity indeed exhibits size invariance. Early work on acceleration perception has suggested that acceleration sensitivity depends on
average velocity and thus behaves in a Weber-like manner (e.g., Gottsdanker et al., 1961). Size-invariance for acceleration sensitivity predicts that acceleration thresholds should indeed be scaled according to average velocity.

We present an experiment designed to obtain more accurate measurements of human sensitivity to accelerated motion. We parameterized acceleration as it is defined; that is, in terms of the change in velocity per unit time. We tested acceleration sensitivity at two base velocities (3.2 and 9.6 deg/s). We also manipulated the angular size of the stimulus path (1.6 deg or 4.8 deg). Varying the sizes of the stimuli within the two base velocities consequently varied the stimulus duration of the display. For a particular base velocity at a given size, acceleration thresholds were measured for two signs of acceleration (positive and negative). Note that the sign of acceleration is described here in relation to the direction of velocity. Thus, positive and negative acceleration refer to increases and decreases in velocity, respectively. In addition, thresholds were obtained for two axes or orientations of motion (horizontal and vertical) and two directions of motion within each axis (left/right and up/down). As in the study by Calderone and Kaiser (1989), the stimulus emerged from and disappeared behind a mask in order to avoid problems of nonveridical perceptions of velocity with motion initiation (Runeson, 1974).

5.2 Methods

5.2.1 Participants

A group of 18 naïve observers that ranged in age from 18 – 25 years (mean age of 20.1 years; 9 males, 9 females) participated in this experiment. All observers had normal or corrected-to-normal vision.
5.2.2 Stimuli and apparatus

The stimulus consisted of a single dot 0.14 deg in diameter. The stimulus traveled in a straight line along either the vertical or horizontal axis in one of two directions (up or down; left or right) and could accelerate either positively (i.e., in the direction of motion) or negatively (i.e., in a direction opposite to the motion). Path length (size) was either 1.6 deg or 4.8 deg and the average velocity of the motion was either 3.2 deg/s or 9.6 deg/s. Consequently, stimulus duration ranged from 167 – 1500 ms.

The stimulus was generated using MATLAB (Mathworks, Natick, MA) and displayed by a Tektronix Inc. 608 oscilloscope with a 6.6 inch screen via a National Instruments NI-USB-6229 data acquisition device. Depending on the trial block, one of two cardboard masks with a square aperture was superimposed on the screen such that the dot’s trajectory began and ended just outside of the aperture.

5.2.3 Procedure

Using the QUEST staircase procedure (Watson & Pelli, 1983), acceleration thresholds were measured in terms of the minimum acceleration (i.e., change in velocity per unit time) that can be detected at the 82% correct level. On each trial, a fixation dot first appeared at the centre of the oscilloscope screen for 1000 ms, after which an accelerated stimulus and an alternate constant velocity stimulus equated for mean velocity, size, orientation (axis), and direction were presented in two separate intervals in random order. The two intervals were separated by a 500 ms inter-trial interval and participants were asked to indicate the interval containing the accelerated stimulus by pressing one of two buttons on the keyboard. Auditory feedback was given for correct/incorrect answers. Stimuli were viewed binocularly at a distance of 80 cm in a dimly illuminated room.
Thirty-two thresholds were obtained for each participant. Participants completed two blocks of trials, one for each sign of acceleration (positive or negative). For each block, all possible combinations of the two sizes and two average velocities were tested in four separate sub-blocks of 200 trials each. The order of testing for the two blocks was counterbalanced among participants and the order of testing of the four sub-blocks was fixed as follows: 4.8 deg size, 3.2 deg/s average velocity; 1.6 deg size, 3.2 deg/s average velocity; 4.8 deg size, 9.6 deg/s average velocity; 1.6 deg size, 9.6 deg/s. Each sub-block contained four interleaved staircases of 50 trials each for the four directions of motions (up, down, left, right). Prior to each block, participants completed a set of 40 practice trials where the sign of acceleration corresponded to that of the block, stimulus size was 4.8 deg, and average velocity was 3.2 deg/s. Where the sign of acceleration was positive participants were instructed to look for the “accelerating” stimulus and where the sign of acceleration was negative participants were instructed to look for the “decelerating” stimulus.

5.3 Results

The data were first analyzed by means of t tests comparing the two directions of motions separately for horizontal and vertical axes. The analyses indicated that thresholds for the right and left directions did not differ, $t(1,17) = -1.998, p = 0.06$, and the thresholds for the up and down directions did not differ, $t(1,17) = .921, p = 0.369$. The data of the two directions within each axis were consequently collapsed for subsequent analyses.

Figure 5.1 presents the data for the two average velocities and two sizes, separately for the horizontal and vertical axes. The stimulus duration per condition is also presented in this figure. Thresholds were generally worse for the faster average velocity (mean threshold for 9.6 deg/s = 4.87 deg/s$^2$) than for the slower average velocity (mean threshold for 3.2 deg/s = 0.49
deg/s²), and worse for the smaller size (mean threshold for 1.6 deg = 4.14 deg/s²) than for the larger size (mean threshold for 4.8 deg = 1.23 deg/s²). From this figure, it is clear that thresholds increased with decreases in stimulus duration. In addition, thresholds were generally higher for the vertical axis (mean threshold = 2.85 deg/s²) as compared to the horizontal axis (mean threshold = 2.51 deg/s²).

Figure 5.1. Acceleration thresholds for horizontal and vertical axes across the two average velocities and sizes. The resulting variation in stimulus duration across the conditions is also presented. Error bars represent ± 1 standard error of the mean.

The above observations were confirmed by a 2 (sign of acceleration) x 2 (average velocity) x 2 (size) x 2 (axis) repeated-measures ANOVA that showed significant main effects of average velocity, $F(1,17) = 354.82, p < 0.001$, size, $F(1,17) = 207.09, p < 0.001$, and axis, $F(1,17) = 18.51, p < 0.001$. In addition, the analysis revealed significant interactions of average velocity x size, $F(1,17) = 118.10, p < 0.001$, average velocity x axis, $F(1,17) = 13.23, p = 0.002$, and size x axis $F(1,17) = 5.91, p = 0.026$. The three-way interaction of average velocity x size x axis was also significant $F(1,17) = 5.29, p = 0.034$. 
The interactions from the initial analyses indicate that thresholds varied in a non-linear manner. The dependency of acceleration thresholds on average velocity is congruent with the assumption that acceleration sensitivity varies in accordance with Weber’s law (Gottsdanker et al., 1961). Moreover, it is clear from Figure 5.1 that acceleration thresholds decreased systematically with increases in stimulus duration. Consequently, further analyses were performed by fitting the data of each participant to a regression model that assumes acceleration thresholds vary in accordance with Weber’s law and size invariance (i.e., are proportional to the mean velocity) and are inversely proportional to stimulus duration, as given in Equation 5.1, separately for the horizontal and vertical axes.

\[
\Delta a = cv / t
\]

(5.1)

In this equation, \( c \) represents an estimated constant (slope), and \( v \) and \( t \) represent average velocity, and duration, respectively. The fit for the horizontal data yielded an average \( r^2 \) of .96 across all participants. The fit for the vertical data yielded an average \( r^2 \) of .98. The mean value of the constant obtained for the two axes of motion is shown in Figure 5.2. The mean estimated value of the constants was 0.14 for the vertical axis and 0.12 for the horizontal axis. Individual \( t \) tests conducted on the values of the constants obtained for the two axes separately for each individual indicated that both the horizontal \((t(1,17) = 14.96, p < 0.001)\) and vertical constants \((t(1,17) = 21.50, p < 0.001)\) were significantly different from zero. The two sets of values were then compared by means of a paired \( t \) test. The analysis indicated that the constant was larger for vertical than for horizontal motion, \( t(1,17) = -3.667, p = 0.002. \)
Figure 5.2. The mean slope of individual linear regressions applied to the data of each participant separately for the horizontal (filled squares) and vertical (open circles) axes. Error bars represent ± 1 standard error of the mean.

5.4 Discussion

We showed that acceleration thresholds, measured here in terms of velocity change per unit time, are well described by the assumption that acceleration sensitivity is size invariant with thresholds varying in a Weber-like manner – increasing linearly with increases in average velocity. Moreover, we showed that thresholds are dependent on time such that thresholds increase with decreases in stimulus duration. These findings are congruent with those reported earlier by Gottsdanker et al. (1961). In the present study, thresholds did not differ between the two signs of acceleration (positive and negative). Interestingly however, we found that acceleration thresholds varied depending on the axis of motion with thresholds being smaller for motions along the horizontal versus vertical axis. This finding is novel and has not been reported in the previous literature. The difference in acceleration sensitivity according to motion axis is
reflected in the value of the constant term estimated from Equation 5.1 as presented in Figure 5.2. This term was larger for vertical as compared to horizontal motion. This constant may be regarded as some invariant that the visual system employs to achieve size invariance for acceleration perception.

Equation 5.1 can also be applied to the thresholds estimated in the early study of Gottsdanker et al. (1961) who also manipulated average velocity and stimulus duration systematically. From the thresholds provided by the authors across average velocities ranging from 0.96-15.38 deg/s and stimulus durations ranging from 0.45-3.64 s (Gottsdanker et al., 1961, Table 3, p. 35), we computed a value of the invariant of approximately 0.6. This higher number is likely due to the rather crude manner in which thresholds were obtained—group thresholds presented in their table were obtained by linear interpolation from the group-averaged data, estimating the acceleration that would yield 75% correct acceleration discrimination performance for the various stimulus durations.

The slope values estimated from Equation 5.1 can additionally be placed into the context of animate stimuli (e.g., biological motion). For reference, we estimated the values of the constant in Equation 5.1, across the trajectory of the human walker foot (vertical dimension only) (Figure 5.3). These values were estimated by systematically sampling the starting temporal positions of the gait cycle (0 - 1.1 s) and computing constants obtained for a range of durations from these starting positions (0 - 1.1 s). For each starting position of the gait cycle, the duration at which a value of 0.15 is first attained for the constant is presented in Figure 5.3. From this figure, it is evident that estimated values of the constant in a similar order of magnitude as those estimated in the present study can in fact be obtained by considering parts of the foot trajectory. Moreover, from this figure, it is evident that the duration required to obtain an estimate of the
constant similar to that obtained in the current study (~0.15) varies depending on the particular portion of the gait cycle being considered. A value of 0.15 for a trajectory beginning at the start of the gait cycle as plotted in this figure, requires a duration of approximately 0.9 s to be attained. By contrast, a trajectory beginning at about 0.25 s position of the gait cycle can attain a comparable value with a duration of 0.4 s. That is, the duration of a stimulus sampled from the foot motion of a human walker that is sufficient to provide a value comparable to the one obtained in the present study depends on the position along the gait cycle from which it is sampled.

Figure 5.3. Values of the constant were computed for the foot motion of a human walker for trajectories beginning at different temporal positions on the gait cycle and lasting variable durations. The line conveys for each temporal position on the trajectory, the duration at which the computed value of the constant, \( c \), first reaches 0.15. For reference, the position on the foot’s trajectory corresponding to “heel strike” is also marked.

Returning to the present data, the lack of difference between thresholds for positively and negatively accelerated motions in the present study is incongruent with previous findings reported
for horizontal (Gottsdanker et al., 1961; Schmerler, 1976) and vertical motions (Calderone & Kaiser, 1989). This difference may at least in part be explained by variations in the manner in which sensitivity is measured and in task and methodology. For example, in contrast to our two-interval design, both previous studies employed a single-interval design requiring observers to decide whether the stimulus is accelerating/not accelerating (Calderone & Kaiser, 1989), or accelerating/decelerating (Babler & Dannemiller, 1993).

Of particular interest is the lack of difference in acceleration sensitivity between upwards and downwards vertical motion. This mirrors the finding by Calderone and Kaiser (1989) and as noted earlier, may be surprising given the gravity-based environment in which human observers must operate. Notably, we have shown that vertical acceleration carries an orientation effect for perceiving the direction of motion of point-light displays depicting animate motions (Chang & Troje, 2009). In this study, observers were presented with stimuli derived from naturally accelerating foot motions and stimuli that were identical but had accelerations removed along the trajectory paths. When the naturally accelerating stimulus was presented upright (and accelerations were consistent with gravitational direction), observers were well able to discern walking direction; however, perceptual performance was impaired once the stimulus was presented at the inverted orientation (and accelerations were opposed to gravitational direction). Critically, this inversion effect disappeared if the accelerations were removed from the trajectories. These findings suggest that human observers are differentially sensitive to biological motions, depending on the orientation of the vertical accelerations.

One may extrapolate findings reported in Chang & Troje (2009) to suggest that human observers may be differentially sensitive to vertical acceleration depending on its direction. We, like others (Calderone & Kaiser, 1989), have shown that this is not the case for a linearly
accelerating stimulus. It remains possible however, that the lack of (vertical) directional differences in our results may be explained by our choice of the reference stimulus; that is, in all cases, observers were asked to discriminate the accelerated stimulus from one in which acceleration was set to zero. If observers are indeed sensitive to gravitational acceleration, using a reference stimulus that carries the same rate of acceleration as gravitational force may yield different results. In similar vein, we can only speculate as to why sensitivity differs according to axis of motion in the present study (and accordingly, are characterized by different invariants). Perhaps a change in the reference stimulus from zero acceleration to gravitational acceleration – a reference commonly observed of objects falling on earth – would provide a more suitable measure of the invariant used to perceive vertical motion.

Little is known about the neural substrates that subserve human visual sensitivity to acceleration. Of particular interest is the middle temporal area (MT) of the cortex – a motion-processing region that has been shown to be tuned to the direction and speed of visual motion (e.g., Britten, Shadlen, Newsome, & Movshon, 1992; Maunsell & Van Essen, 1983; Petersen, Baker, & Allman, 1985). In the macaque, visual acceleration is not represented by the activity of MT neurons individually, but may be represented when the pooled population activity is considered (Lisberger & Movshon, 1999; Price, Ono, Mustari, & Ibbotson, 2005). Interestingly, area MT projects to the pretectum and the dorsolateral pontine nucleus (DLPN) – areas in which neurons in the macaque encode the acceleration of retinal motion (e.g., Cao, Gu, & Wang, 2004; Das, Economides, Ono, & Mustari, 2001). Such areas are involved in oculomotor control (e.g., smooth pursuit) and are not thought to govern directly the perception of accelerated motion. However, it has been shown that a common motion processing stage may limit acceleration sensitivity in both smooth pursuit and visual discrimination (Watamaniuk & Heinen, 2003). In
this study, the authors tested the sensitivity of the pursuit system to speed and acceleration by measuring eye acceleration. In a second perceptual discrimination experiment, observers were asked to make psychophysical judgments of stimuli that moved with the same speeds and accelerations as the original pursuit targets. The authors showed that the pursuit data, when converted to oculometric functions (analogous to psychometric functions), were markedly similar to performances on the perceptual task. The authors therefore suggested that a common motion-processing stage limits the performance of both systems.

Interestingly, the perception of gravitationally consistent motion seems to implicate involvement of the vestibular network (Indovina, Maffei, Bosco, Zago, Macaluso, & Lacquaniti, 2005). In this study, brain activity was measured using fMRI while observers perceived visual motion that acted consistently or inconsistently with the direction of gravity. The authors demonstrated that visual motion that was consistent with gravity activated the vestibular network including areas such as the insular cortex, temporoparietal junction, premotor and supplementary motor areas, middle cingulate cortex, postcentral gyrus, putamen, and the posterior thalamus, in addition to several visual motion areas. In contrast, motion that was inconsistent with gravity elicited responses in the visual areas only.

Human sensitivity to visual acceleration, as compared to other motion properties such as speed, is relatively poor (e.g., Snowden & Braddick, 1991; Werkhoven, Snippe, & Toet, 1992). Yet the retrieval of acceleration has been shown to be crucial for the perception of animate motions (e.g., Runeson & Frykholm, 1981). It is of particular interest then to achieve a better understanding of visual sensitivity to acceleration, including its variation across changes in stimulus size. The experiment presented here was designed to obtain refined measures of acceleration sensitivity by manipulating acceleration in a staircase procedure to obtain thresholds.
across various mean velocities and sizes. We showed that acceleration thresholds are well described by the assumption that sensitivity is size-invariant, scaling with changes in mean velocity. Moreover, thresholds scaled inversely with stimulus duration. Towards this particular finding, it remains possible that variations of acceleration sensitivity across time may not be so simple: intuitively, sensitivity may be different at extreme stimulus durations (not tested here). Additional empirical work will be needed to clarify this possibility.
References


Chapter 6

Searching for the Adequate Local Biological Motion

6.1 Introduction

A set of moving light-points placed to represent the motions of the major joints of an agent is sufficient to elicit the percept of a walking person (Johansson, 1973). Biological motion as conveyed by such point-light displays carries significant wealth of information and is thought to be subserved by multiple, independent processing levels (Troje, 2008) rather than a single mechanism.

That distinct mechanisms contribute to biological motion perception is highlighted by demonstrations of two different inversion effects in this domain: one associated with the retrieval of the global shape (structure-from-motion) of the walker and a second associated with the retrieval of the local motions of the limbs (Troje & Westhoff, 2006). It has been well-documented that performances on a variety of behavioural tasks decrease if the point-light figure is shown upside-down (e.g., Sumi, 1984). Inversion effects for biological motion perception have traditionally been attributed to impaired processing of the global shape of the walker (Bertenthal & Pinto, 1994). Troje and Westhoff (2006) showed, however, that point-light animations that are spatially perturbed and entirely devoid of structural information still convey information as to the agent’s walking direction. Notably, such displays are also subject to an inversion effect – the ability to retrieve the facing direction of the scrambled walker is impaired if the walker is shown upside down. This effect was shown further to be carried entirely by the local trajectories of the
feet, corroborating earlier findings by Mather, Radford, and West (1992) that emphasize the importance of motions of the distal parts of the body.

The local motion-based inversion effect is a particularly interesting phenomenon as it has been proposed to reflect a mechanism that detects dynamics signifying ballistic motions characteristic of the locomotion of terrestrial, articulated agents (Troje & Westhoff, 2006; Johnson, 2006; Troje, 2008). Indeed the perception of animacy from scrambled point-light displays containing solely local motions signals is also orientation-specific, supporting the notion that the same orientation-specific cue that indicates direction has a more general function (Chang & Troje, 2008). The cue to which the local mechanism is tuned has been proposed to be some gravity-defined invariant contained in the motions of the limbs (Troje & Westhoff, 2006). One candidate cue for such an invariant is the acceleration exhibited by the foot motion of terrestrial animals that results as the foot is being pulled up and released according to gravitational force – a gravitational-kinetic energy transfer that constitutes a large portion of the total energy change during locomotion (e.g., Cavagna, Heglund, & Taylor, 1977). Evidence that acceleration may be a particularly important cue for the perception of biological motion comes from recent findings that the retrieval of walking direction from stimuli consisting of solely foot trajectories is reliable and is subject to an inversion effect only if the acceleration in the stimuli is left intact. If the acceleration is removed from the stimuli the associated inversion effect disappears. Moreover, the magnitude of the inversion effect is correlated with the vertical acceleration in the stimuli (Chang & Troje, 2009). The finding that the visual system exploits acceleration for the perception of biological motion is congruent with previous studies showing that human observers employ knowledge about the dynamics of events and the effects of gravity and employ assumptions about fixed spatial and temporal relations that exist in a gravity-driven environment.
for the perception of inanimate and animate events (Pittenger, 1985; Stappers & Waller, 1993; Jokisch & Troje, 2003; Shipley, 2003). The orientation specificity observed for the perception of (local) biological motion can thus be regarded as being rooted in the orientation specific nature of gravitational acceleration.

Little is known as to the properties of the mechanism purported to underlie the perception of local biological motion. Here, we attempted to characterize the responsible visual filter by searching for the stimulus to which it is optimally tuned. We used a genetic algorithm (GA) to drive an adaptive, psychophysical search through a multi-dimensional space derived from a second-order Fourier series defining the motion of a dot representing the foot’s motion in a two-dimensional image plane (Troje, 2002) (see Procedure). A second-order representation was chosen as preliminary data indicated that the saliency of directional information in local biological motion is not affected by the omission of the higher harmonics (see Appendix C).

Genetic algorithms driven by subjective human input have been employed successfully in clinical and research settings. For example, GAs have been used to fit customized settings for hearing devices and cochlear implants (Baskent, Eiler, & Edwards, 2007; Wakefield, van den Honert, Parkinson, & Lineaweaver, 2005). In the visual domain, GAs have been used to investigate the neural coding of three-dimensional shapes (Yamane, Carlson, Bowman, Wang, & Connor, 2008). The genetic algorithm, first introduced by Holland (1975) models the natural process associated with biological evolution – notably, the Darwinian rule of the selection of the fittest. Beginning with an initial generation of randomly generated “foot” motions, the stimuli evolved through multiple generations with feedback from observers. Parent stimuli were subsequently morphed to produce descendant stimuli that varied the parents’ motion.
characteristics. The “fitness” of each stimulus determined the probability with which it produced descendent children for subsequent generations.

We defined the adequate or “fittest” stimulus as one that satisfies two criteria: (1) the stimulus carries strong cues as to the facing direction, and (2) the stimulus carries a pronounced inversion effect. In light of previous findings regarding the relevance of acceleration for the local biological motion inversion effect (Chang & Troje, 2009), we hypothesized that the optimization of the inversion effect in particular should be matched by, but not necessarily limited to, the evolution of stimuli towards carrying greater vertical acceleration information.

In order to conduct an effective search through the multi-dimensional space, a large number of trials needed to be completed. As such, the experiment was carried out via a web-based interface that permitted a large number of observers around the world to participate. Two runs of searches were carried out with different mean starting points in the search space.

6.2 Methods

The experiment was implemented in Adobe Flash and was connected to a MySQL database via a PHP server. The experiment was uploaded on the laboratory web space for public access.

6.2.1 Participants

Two groups of observers participated in this experiment. Observers were recruited online and completed the experiment via the web-based interface. Group 1 consisted of 185 unique observers with a mean reported age of 30.67 (reported gender of 81 males, 104 females). Group 2 consisted of 167 unique observers with a mean reported age of 27.17 (reported gender 81 males; 86 females). Logged IP addresses revealed further demographics of the groups. Group 1
comprised observers from 33 different countries with the majority of observers originating in the United States (25%), Germany (15%), and Canada (13%). Group 2 was comprised of observers from 34 different countries with again the majority of observers originating in the United States (38%), Germany (10%), and Canada (8%). Groups 1 and 2 participated in the first and second runs of searches, respectively.

6.2.2 The search space

The search space consisted of eight parameters and included frequency, amplitudes, and phases derived from a second-order Fourier series defining the motion of a single dot (a “foot”) in a two-dimensional plane (Troje, 2002) (Appendix C). The position of a single dot representing a “foot” in dimensions $x$ and $y$ at a given time $t$ is given by

$$
\begin{align*}
    j_x(t) &= c_x + a_{x,1} \sin(\omega t + \phi_{x,1}) + a_{x,2} \sin(2\omega t + \phi_{x,2}) \\
    j_y(t) &= c_y + a_{y,1} \sin(\omega t + \phi_{y,1}) + a_{y,2} \sin(2\omega t + \phi_{y,2})
\end{align*}
$$

(6.1)

where $c$ describes a time invariant constant and the gait frequency $f$ is derived as $\omega/2\pi$. The eight parameters constituting the search space therefore consisted of the first and second order amplitudes $a_{i,1}$, $a_{i,2}$, of both dimensions, the second order phase of the $x$ dimension, $\phi_{x,2}$, (as we were not interested in absolute phase, the first order phase along one dimension was set to zero and was not searched), the first and second order phases of the $y$ dimension, $\phi_{y,1}$, $\phi_{y,2}$, and the gait frequency, $f$. 

147
6.2.3 Stimuli

Each stimulus was derived from a particular “foot” motion, defined by a set of eight parameters (outlined above). These motions were shown in a treadmill-walking manner and were shown in sagittal view (facing leftwards or rightwards). The leftwards version of a particular stimulus was defined as the mirror-flip of the rightwards stimulus about the vertical axis. Similarly, an inverted stimulus was created by flipping the upright stimulus about the horizontal axis. The stimulus consisted of five copies (dots) of one particular stimulus displaced randomly at the centre of a display area measuring approximately 9 deg x 9 deg (as it was not possible to control the observer’s environment, we provide measurements hereafter assuming a viewing distance of 50 cm and a screen resolution of 37 pixels/cm. The five dots were given random onset-times and the stimulus was presented for a duration of 1.5 s.

Each generation consisted of 50 stimuli. For the initial generation only, four “real” foot motions were included in the population that comprised of the motion of the human walker, human runner, cat, and pigeon. The human walking and running motions were computed as the average motion from motion-capture data of 50 men and 50 women (Troje, 2002). The cat motion was obtained by sampling points from single frames of a video sequence showing a cat walking on a treadmill, and the pigeon sequence was generated from motion-capture data obtained from a pigeon fitted with markers. The remaining 46 stimuli of the initial generation were randomly generated from a uniform distribution. Distinct sets of stimuli were generated for Runs 1 and 2. Consequently, the mean “foot” of the initial generation of the first and second runs was different.
6.2.4 Procedure

Genetic algorithm

We used a multi-objective evolutionary algorithm based on the popular non-domination based algorithm (NSGA-II) by Deb, Pratap, Agarwal, and Meyarivan (2002). A general overview of the algorithm follows.

After the initial population was evaluated by the observers, it was sorted based on Pareto ranking and a crowding distance measure. Pareto rankings were determined in the objective space. Here, the objective space was two dimensional and was comprised of an axis reflecting a function that optimized behavioural performances for the upright stimulus and a second axis reflecting a function that maximized the inversion effect (see “Selection” below). The set of individuals or “feet” sorted to the first rank was a completely non-dominant set in the population. That is, these were the best feet in terms of both objectives in the population. Subsequent ranks were assigned to the remaining feet based upon the number of feet in the previous ranks that dominated them (i.e., were better in terms of both objective functions). In this manner, each foot in the population was assigned a rank value. In addition to the rank value, a crowding distance parameter was calculated for each foot. This is a measure of how close an individual is to its neighbours and involved computing the Euclidean distance between each individual within a rank in the multi-dimensional parameter space (i.e., the stimulus search space). A large crowding distance promotes greater diversity in the population. This parameter was used to rank individuals within a given Pareto rank.

Parents were selected from the population via binary tournament selection based on Pareto rank and crowding distance. An individual was selected if their rank was higher than the other or if their crowding distance was greater. The selected population generated offspring from
crossover and mutation operators, which are outlined below. The offspring population was then evaluated. In subsequent generations, a pooled population that included the previous parent population and current offspring was sorted based on Pareto ranking and crowding distance and only the best \( N \) individuals were selected to be parents, where \( N \) was the population size (50). These individuals then generated an offspring population for testing in the next generation.

**Selection:** Individuals were selected based on

1. Non-domination rank, \( p_{\text{rank}} \) (i.e., individuals in Pareto front \( F_i \) have \( p_{\text{rank}} = i \)).

   Pareto rankings were assigned based on two objective functions: the first minimized error rates for stimuli presented in the upright orientation, \( \min(E_{\text{upright}}) \), and the second minimized the difference in error rates between upright and inverted orientations, \( \min(\text{abs}(1/2 - (E_{\text{inverted}} - E_{\text{upright}}))) \). This second function ensured that we did not converge at stimuli that exhibited reversals in perceived facing direction (i.e., perceived to be facing rightwards when shown upright and facing leftwards when shown inverted). Note that as “upright” and “inverted” and “left” and “right” were determined arbitrarily, stimuli were mirror-flipped about the horizontal axis when \( E_{\text{upright}} > E_{\text{inverted}} \), and mirror-flipped about the vertical axis when \( E_{\text{upright}} > 0.5 \). Thus, stimuli where \( E_{\text{inverted}} - E_{\text{upright}} \) was 0.5 or -0.5, for example, were treated equally.

2. Crowding distance, \( F_i(d_i) \).

   Individual \( p \) was selected if \( p_{\text{rank}} < q_{\text{rank}} \), or if \( p \) and \( q \) were of the same rank \( F_i \), then \( F_i(d_p) > F_i(d_q) \).

**Genetic operators:**

Genetic operators were used to generate the set of offspring and included methods of crossover and mutation. Selected parents generated offspring by crossover 90% of the time.
Parents selected for crossover generated two children offspring with each parameter of the offsprings determined as follows:

\[
C_{1,k} = \frac{1}{2} \bigg[ (1 - \beta_k) P_{1,k} + (1 + \beta_k) P_{2,k} \bigg]
\]

(6.2)

\[
C_{2,k} = \frac{1}{2} \bigg[ (1 + \beta_k) P_{1,k} + (1 - \beta_k) P_{2,k} \bigg]
\]

(6.3)

where \( C_{i,k} \) indicates the \( k \)th parameter of \( i \)th child and \( P_{i,k} \) indicates the same parameter in the selected parent. \( \beta_k \) (≥ 0) is a number calculated by

\[
\beta(u) = 2u^{1/(\eta+1)}, \text{ if } u < 0.5
\]

(6.4)

\[
\beta(u) = 1/2 (1-u)^{1/(\eta+1)}, \text{ if } u \geq 0.5
\]

(6.5)

where \( u \) is a uniformly distributed number between 0 and 1, and \( \eta_c \) is a distribution index for crossover. This index is a number that defines the shape of the probability distribution of the operator and determines how close (large values) or far (small values) the children will be from their parents. Here, a fixed value of 20 was chosen in order to obtain better precision of the solution.

Selected parents generated two offsprings by mutation 10% of the time. Each parameter of an offspring was determined as follows:

\[
C_k = P_k + \left( P^*_{k} - P^*_{k} \right) \delta_k
\]

(6.6)

where \( C_k \) indicates the \( k \)th parameter of the child, \( P_k \) denotes the same parameter of the parent and \( P^*_{k} \) and \( P^*_{k} \) as the upper and lower bounds on the parent parameter. \( \delta_k \) is a variable number calculated from a polynomial distribution by using

\[
\delta_k = (2r_k)^{1/(\eta_m+1)} - 1, \text{ if } r_k < 0.5
\]

(6.7)
\[ \delta_k = 1 - \left[ 2(1 - r_k) \right]^{1/(\eta_m + 1)}, \text{ if } r_k \geq 0.5 \]  

where \( r_k \) is a uniformly distributed random number between 0 and 1, and \( \eta_m \) is the distribution index for mutation. Similar to the crossover index, this is a number determining the spread of the children from the parents and was assigned a fixed value of 20.

**Psychophysical direction discrimination task**

Participants were first presented with a set of instructions on the screen and were familiarized with point-light stimuli via demonstrations of full point-light figures (e.g., a walking human, cat, or pigeon) shown in sagittal view. The participants were informed that on each trial a set of dots would appear on the screen that depicted a moving agent. The task was to indicate whether the entity appeared to be facing rightwards or leftwards by pressing one of two buttons that appeared on the screen after stimulus offset.

Upon submitting user information such as age and gender, the experimental trials began. The participants were free to complete as many trials as they wished but only data from users that completed a minimum of 50 trials were stored in the database. Data were saved subsequently in 50 trial blocks. Upon the completion of 200 trials, the observer was provided with a “score” that reflected his/her accuracy rate for all trials and was provided with the option to continue to complete additional trials or to exit.

Each generation consisted of 1000 trials to be completed across different observers. The trials comprised of all possible combinations of the 50 “feet” stimuli that were oriented upright or inverted, facing leftwards or rightwards, repeated five times. Trials were randomized and were divided into 20 subsets of 50 trials.
Upon the completion of all trials for a given generation, the stimuli were evaluated in terms of observer discrimination rates for the two stimulus orientations. The fittest stimuli were then selected to generate the subsequent generation of 1000 trials for testing (see previous section for details on the algorithm). In this manner, Runs 1 and 2 evolved 25 and 27 generations, respectively. The runs were terminated when behavioural performance rates were observed to have reached a plateau level and the evolution of the stimulus generations within the parameter space was observed to be no longer progressing (see Results below).

6.3 Results

6.3.1 Psychophysical discrimination task

Run 1

Mean performances on the direction discrimination task across generations for Run 1, expressed in terms of error rates for the two stimulus orientations are presented in Figure 6.1A. From this figure, it is evident that error rates for stimuli presented in the upright orientation decreased gradually across generations. The mean error rate was approximately 0.45 in the first generation and declined to approximately 0.07 by the last generation. A linear regression applied to the data confirmed the systematic decrease in error rates observed across generations, \( b = -0.02, t(24) = -9.22, p < 0.001, r^2 = 0.78 \). By contrast, performances for the inverted stimuli did not change appreciably. Error rates remained around chance-level across all generations (\( b = 0.03, t(24) = 1.95, p > 0.05, r^2 = 0.10 \)).

We subsequently quantified the inversion effect by using an inversion score, computed by subtracting the error rate for the upright stimulus from that of the inverted stimulus. The
scores for Run 1 are shown in Figure 6.1B. It is clear that the score increased systematically across generations and reached an asymptote of 0.5 (see Procedure for details on the computation of the second objective). This observation is supported by a linear regression applied to the scores, $b = 0.02, t(24) = 10.86, p < 0.001, r^2 = .83$.

Figure 6.1. (A) Mean performances on the direction discrimination task for each generation in Run 1, expressed in terms of error rates for the two orientations. (B) The inversion effect, quantified in terms of an inversion score obtained by subtracting error rates of the upright stimulus from that of the inverted stimulus. Error bars represent ± 1 standard error of the mean.

Run 2

Error rates for the stimuli of Run 2 for the two stimulus orientations are presented in Figure 6.2A. Similar to Run 1, error rates for the upright orientation decreased gradually across generations. The average error rate for the upright stimuli was 0.40 in the first generation and
declined to 0.05 by the final generation. The systematic decline in error rates observed for the upright stimuli across generations was confirmed with a linear regression applied to the data, $b = -0.02, t(26) = -17.77, p < 0.001, r^2 = .92$. The average error rate for the inverted stimuli showed a slight decreasing trend across generations ($b = -0.002, t(26) = -2.62, p = 0.01; r^2 = .18$).

The inversion scores for Run 2 are shown in Figure 6.2B. The score increased systematically across generations, and like Run 1, reached an asymptote of 0.5. The linear regression applied to the scores confirmed this observation, $b = 0.02, t(26) = 18.10, p < 0.001, r^2 = .93$.

Figure 6.2. (A) Mean error rates for each generation in Run 2 for the two orientations. (B) The inversion score across generations. Error bars represent ± 1 standard error of the mean.
6.3.2 Principal components analysis

To examine the evolution of the stimuli across generations, we subsequently applied PCA to the stimulus sets as defined by the eight search parameters. The variables (search parameters) were normalized and the analysis was applied separately to Runs 1 and 2.

*Individual analyses on Runs 1 and 2*

The analysis for Run 1 indicated that approximately 93% of the variance can be explained by considering solely the first two principal components. The progression of the stimuli across generations within these first two principal axes is shown in Figure 6.3. It is evident that the stimuli generally evolved in a coherent direction across generations.

![Principal Component 1 vs. Principal Component 2](image)

Figure 6.3. PCA on the progression of Run 1 through the parameter space. The progression of the stimuli across generations through the space defined by the first and second principal components is shown. The first and last generations are indicated by the open circle and open triangle symbols, respectively. For greater clarity, the locations of these symbols are marked by a large circle. Error bars represent ± 1 standard error of the mean.
An independent analysis was applied for Run 2. As for Run 1, a significant proportion of the variance (90%) can be explained by the first two principal components. The progression of the stimuli across generations within the space defined by the first two principal axes is shown in Figure 6.4.

![Figure 6.4](image.png)

**Figure 6.4.** PCA on the progression of Run 2 through the parameter space. The progression of the generations through the space defined by the first and second principal components is shown. The first and last generations are indicated by the open circle and open triangle symbols, respectively, marked on the figure by larger circles. Error bars represent ± 1 standard error of the mean.

**Runs 1 and 2 pooled in a common space**

We subsequently examined the progression of both Runs 1 and 2 across generations within a common space by applying PCA to the pooled stimuli of both runs. The two runs in a space defined by the first two principal components from this analysis (accounting for 93% of the variance of both sets of data) are presented in Figure 6.5. For reference, the locations of the four
“real” foot motions included in the first generation of both runs within this space are also indicated in the figure. From this figure, it can be seen that the two runs began at neighbouring points in this space and evolved along the first principal axis in a similar direction (to see the foot motions across generations, please visit http://biomotionlab.ca/Demos/BMLevolutionDisplay_Graph.html).

Figure 6.5. The progression of Runs 1 and 2 within a common space defined by the first two principal components from an analysis of stimuli pooled from both runs. The first and last generations are labeled and marked by grey arrows. Locations of the four “real” foot motions included in generation 1 are also indicated in this space (grey symbols). Error bars represent ± 1 standard error of the mean.

The component loadings of the pooled analysis for the original search parameters are presented in Figure 6.6. As can be seen in this figure, the first principal axis is largely contributed to in a positively-weighted manner by the vertical components of the first- and second-order Fourier amplitudes, and the horizontal component of the first-order amplitude. Inspection of the loadings also revealed that the frequency of the gait cycle was the main source
of variation along the second principal dimension. The displacement of the endpoint of Run 2 with respect to Run 1 about the vertical axis in Figure 6.5 can therefore be explained by the additional evolution towards higher gait frequency for Run 2.

![Figure 6.5](image_url)

Figure 6.6. Loadings of the eight search parameters for the first (A) and second (B) principal axes.

### 6.4 Discussion

#### 6.4.1 Performance changes and evolution of stimuli across generations

The results indicate that the GA was effective at optimizing performances for upright stimuli and maximizing the inversion effect on the psychophysical discrimination task. For both runs, error rates for the stimuli presented upright decreased appreciably across generations. The inversion effect, quantified by the difference in error rates between the inverted and upright stimuli also increased across generations, reaching an asymptote at a score near 0.5 in both runs.
(Figures 6.1B, 6.2B). As noted earlier, this asymptote reflects the optimal inversion score that can be obtained with our algorithm (see Procedure). The increase in the inversion score across generations in both runs was attributed mainly to changes in error rates for upright rather than inverted stimuli.

The changes in performance rates in response to the stimuli were accompanied by coherent changes of the stimulus characteristics (defined by the eight search parameters) across generations. Results from the PCA indicated that the two runs evolved stimuli across generations in a manner that can be well described by considering two principal axes (Figure 6.5). Both runs evolved towards the same direction and converged at points that are markedly different from the four “real” foot motions included in the initial generation.

### 6.4.2 Amplitude and acceleration

The vertical component of the 1st and 2nd order amplitude and the horizontal component of the 1st order amplitude were the main components along which the stimuli in both runs evolved. The mean amplitude of each generation are shown separately for the horizontal and vertical components in Figure 6.7A (Run 1) and 6.7B (Run 2). For both runs, horizontal and vertical amplitude increased gradually across generations. The changes observed for the vertical amplitude, in particular, are congruent with previous findings indicating the importance of vertical components of the motion for the inversion effect observed for biological motion defined solely by local motion (Chang & Troje, 2009). In this previous study vertical acceleration was found to be the critical cue. As such, we proceeded to examine the evolution of vertical acceleration across generations.
Figure 6.7. Acceleration and amplitudes for horizontal and vertical components across generations for Run 1 (A) and Run 2 (B). Error bars represent ±1 standard error of the mean. In the second row, the inversion score is shown against both vertical acceleration and vertical amplitude for Run 1 (C) and Run 2 (D).

For each foot, we first quantified acceleration by computing the difference between the maximum and minimum vertical acceleration. The mean vertical acceleration across generations for Runs 1 and 2 are also shown in Figure 6.7A and Figure 6.7B, respectively. For reference,
horizontal acceleration is also shown in these figures. For both runs, vertical acceleration increased across generations. Although the stimuli of the initial generation of both runs contained similar vertical acceleration, the stimuli of Run 2 evolved towards a more significant increase in vertical acceleration by the end of the run. By contrast, the mean vertical amplitude at the end of both runs was very similar for both runs. For Run 2, a marked increase in horizontal acceleration towards the end of the run was also observed.

To illustrate the relationship between the vertical components of motion and the inversion effect, we plotted the inversion score against the vertical amplitude and acceleration in Figure 6.7C and 6.7D for Runs 1 and 2, respectively. A positive correlation between the inversion effect and both vertical amplitude and acceleration is readily demonstrated by these figures (amplitude $r = .94$ and acceleration $r = .84$ for Run 1; amplitude $r = .87$ and acceleration $r = .94$ for Run 2).

From the analyses presented above, it is difficult to disentangle the relevance of vertical amplitude versus vertical acceleration for the evolution of the inversion effect. Certainly for amplitude, there is no asymmetry in this information between upright and inverted versions of the stimuli that could account for orientation dependent perception. Recall that in the previous analysis, acceleration was represented as the mere amplitude of the acceleration curve of each “foot’s” motion. Any information with regards to how acceleration evolved across generations along the trajectory (i.e., spatial distribution) was not available. As such, no asymmetry between upright and inverted stimuli is conveyed by this measure either. It seems worthwhile then to examine further the evolution of vertical acceleration across generations in an alternate manner – more specifically, in terms of the distribution of vertical acceleration for each foot along its trajectory. For real foot motions, the critical acceleration information appears to be contained within the swing phase of the gait cycle, and in particular, when the foot falls from its highest
position to the ground (Chang & Troje, 2009). For each foot, we therefore defined an analogous region within its gait cycle from which the foot moved from position $y_{\text{max}}$ to position $y_{\text{min}}$ (presumably the point at which the “foot” contacts the ground). The distribution of acceleration (vertical only) for this region is shown in Figure 6.8A (Run 1) and 6.8B (Run 2). As is evident in the figure, the “feet” generally evolved towards carrying more pronounced positive acceleration at the beginning of the drop from $y_{\text{max}}$ (the sign of acceleration is defined with respect to gravitational direction) and more negative acceleration before contact with the ground (position 0). From this figure, a shift (decline) in the proportion of the region during which acceleration is positive as compared to negative can also be seen across generations. This proportion is shown in Figure 6.9. Linear regressions applied to both runs confirm a general decline of this proportion across generations, $b = -0.003$, $t(24) = -5.84$, $p < 0.001$, $r^2 = 0.58$ for Run 1, $b = -0.008$, $t(26) = -16.47$, $p < 0.001$, $r^2 = 0.91$ for Run 2. In addition, individual $t$ tests performed on this proportion across generations revealed that for both runs and for all generations, the proportion of the region during which acceleration is positive as compared to negative was significantly above 0.5 ($t(24) = 42.58$, $p < 0.001$ for Run 1, $t(26) = 21.92$, $p < 0.001$ for Run 2).
Figure 6.8. The distribution of vertical acceleration across a region defined analogously to the swing phase of the gait cycle for Run 1 (A) and Run 2 (B). This region corresponds to the portion of the trajectory where the “foot” moves from position $y_{\text{max}}$ to position $y_{\text{min}}$ (0 deg). Positive (grey shading) and negative (blue shading) acceleration is defined with respect to the direction of gravity. The magnitude of acceleration is conveyed by the shading intensity.
These analyses suggest that the change in the amplitude of vertical acceleration across generations as computed earlier (Figure 6.7) was accompanied by changes in the distribution of vertical acceleration within the trajectory. Notably, across generations, the “feet” evolved towards carrying greater negative acceleration prior to contact with the ground. Such a pattern would be a notable contrast to one that may be observed for an inanimate object falling to the ground under the influence of gravity, and may be a cue towards signaling the animacy of a stimulus. Note that although the proportion of the trajectory carrying negative acceleration increased across generations, it reached a plateau level of approximately 0.3 (Figure 6.9). Across generations, a significantly larger proportion of the trajectory remained positively accelerating. This asymmetry in the distribution of positive and negative acceleration during the swing phase of the gait cycle may account for the inversion effect.
6.4.3 The adequate local biological motion reflecting sensitivity to gravity-driven visual invariants

As noted earlier, previous work has shown that the visual system exploits fixed spatial-temporal relationships that are the consequence of gravitational force acting upon events (e.g., Pittenger, 1985; Runeson & Frykholm, 1981; Jokisch & Troje, 2003). Such “knowledge” about the influences of gravity is not limited to visual information but may also involve the integration of information from the vestibular system (Indovina, Maffei, Bosco, Zago, Macaluso, Lacquaniti, 2005; see also, Chapter 7). Nonetheless, for biological motion, an event that is optimized to minimize energy expenditure (Alexander, 2003) and is in part shaped by gravitational force – with the work done by muscles reduced by a pendular transfer between potential and kinetic energy (e.g., Cavagna, Saibene, & Margaria, 1963; Cavagna, Heglund, & Taylor, 1977), it is not surprising that the human visual system has a capacity to detect cues such as gravity-driven acceleration in the distal limbs. The evolution towards the adequate local biological motion in the current experiment supports the notion that the responsible local visual mechanism is tuned to such a gravity-driven invariant with the evolution of the distribution of vertical acceleration along the swing phase of the trajectory across generations.

Sensitivity to the proposed invariant appears to be present very early in life and even at birth. Support for the early presence of mechanisms processing biologically relevant stimuli can be found in research of filial preference behaviour of chicks. Newly hatched chicks respond to point-light animations and prefer to approach displays of intact or scrambled walking hens rather than displays of randomly moving dots (Vallortigara, Regolin, & Marconato, 2005). Furthermore, newly hatched chicks prefer to align with a display of an intact, upright walking hen rather than one in which the intact walker is inverted (Vallortigara & Regolin, 2006). These results suggest that preference behaviours of newborn chicks are guided by visual invariants that
the perceptual systems are tuned to at birth. Similar data can be found with human infants: two-
day-old infants prefer to look at a point-light display depicting a walking hen rather than a display of randomly moving dots. These infants also prefer to look at point-light displays that are oriented upright rather than inverted (Simion, Regolin, & Bulf, 2008).

6.5 Conclusion

Local biological motion that is entirely devoid of configural form information still carries significant information as to the facing direction of an entity (Troje & Westhoff, 2006). Here, we characterized the responsible visual filter by using a genetic algorithm to search for the stimulus to which it is optimally tuned. We showed that the GA was effective at optimizing behavioural discrimination rates of stimuli shown upright and maximizing the inversion effect. These changes were accompanied by the evolution of stimuli across generations towards carrying greater amplitude and acceleration for both the vertical and horizontal components (and the evolution of stimuli towards carrying faster gait frequency). The evolution of vertical acceleration could be described by a change in the distribution of acceleration within a region defined analogously to the swing phase of the gait cycle. Specifically, the stimuli evolved towards carrying more pronounced positive acceleration at the beginning of the swing phase and more negative acceleration before contact with the ground – a characteristic that may signify the animacy of an object. Moreover, across generations, there remained an asymmetry in the proportion of the region carrying positive versus negative acceleration that may account for the inversion effect. The data support a mechanism that is tuned to gravity-defined invariants for the perception of local biological motion.
References


*Perception, 13*, 283–286.


Chapter 7

Reference Frames for the Perception of Biological Motion and Faces

7.1 Introduction

Many objects become unfamiliar or even unrecognizable when they are seen in other than their veridical orientation (e.g., Thompson, 1980; Sumi, 1984). As noted by Rock (1973), this observation contradicts the early Gestalt theorists who considered that the essential information for the perception of form was contained primarily in the geometrical relationship of the features – that is, perceived form would remain unchanged as long as these internal relationships were maintained. Orientation specificity demonstrates that the stimuli are encoded not just in terms of internal relations but also in relation to external reference frames. Such reference frames can be either egocentric or allocentric (see Howard, 1982 for a review). Egocentric reference frames include the retina, head, and body. Allocentric reference frames include gravity and the visual environment. The reference system that is most important in determining the way in which we perceive an object seems to depend on the stimulus class involved. Here, we consider the perception of faces and biological motion.

The recognition of a face is impaired if it is inverted (see Valentine, 1988 for a review). The inversion effect observed for face perception is due primarily to disruptions of configural

---

5 This chapter incorporates the outcome of joint research undertaken in collaboration with Laurence Harris at York University, Canada, under the supervision of Nikolaus Troje. This chapter is reprinted with permission from “Frames of reference for biological motion and face perception” by D. H. F. Chang, L. R. Harris, and N. F. Troje (2010). Journal of Vision, 10(6):22, 1–11. Copyright 2010 by The Association for Research in Vision and Ophthalmology.
rather than featural processing (Freire, Lee, & Symons, 2000). Inversion effects have also been reported for the perception of biological motion (e.g., Sumi, 1984). Biological motion appears to carry at least two distinct inversion effects (Troje & Westhoff, 2006). While there appears to be an inversion effect that is due to impaired processing of the global, configural shape of the walker as conveyed by the display’s spatiotemporal organization (e.g., Bertenthal & Pinto, 1994), there is also a second inversion effect that is associated with local motion signals of the distal limbs (Shipley, 2003; Troje & Westhoff, 2006; Chang & Troje, 2009a). To this end, Troje and Westhoff (2006) presented observers with intact and scrambled walker displays in which the dot’s spatial organization was perturbed. Significantly, observers could retrieve the facing direction of the walker not only for the coherent displays but also for the scrambled displays which retained solely local motion information. Moreover, an inversion effect was observed for both intact and scrambled displays. The authors later showed that the cues to direction in the scrambled displays and the associated inversion effect were carried entirely by the local motion of the feet. Subsequent work has shown that the local inversion effect depends on vertical acceleration contained in the foot motion (Chang & Troje, 2009a). Moreover, the mechanisms underlying the perception of global motion-mediated structure and local motion signals are dissociable according to a variety of behavioural characteristics such as sensitivity to masking or susceptibility to learning (Chang & Troje, 2009b).

Few studies have investigated the reference frames in which animate motions and faces are coded. Observers use implicit knowledge about the direction and effects of gravity when interpreting biological and inanimate events (e.g., Runeson & Frykholm, 1981, Pittenger, 1985; Stappers & Waller, 1993; Jokisch & Troje, 2003; Shipley, 2003). Making assumptions with regards to the direction of gravity, however, does not necessarily implicate direct measurements
of gravitational acceleration (e.g., via input from the vestibular system). The visual system may simply take advantage of the fact that gravity is typically aligned with egocentric coordinates. Still, there is reason to believe that an allocentric system may be involved for the perception of dynamic events (Bingham, Schmidt, & Rosenblum, 1995; Indovina, Maffei, Bosco, Zago, Macaluso, & Lacquaniti, 2005; Lopez, Bachofner, Mercier, & Blanke, 2009). Bingham et al. (1995) found that the recognition of point-light-defined events was stronger for displays that were upright rather than inverted with respect to gravity, regardless of the observer’s orientation in space and concluded that point-light events are perceived in relation to a gravitational rather than an egocentric frame of reference. Later findings, however, appear to be inconsistent with this conclusion. Troje (2003) found that performance on a biological motion task depended only on whether the display was aligned with the observer regardless of the observer’s orientation in space, suggesting egocentric coding. A study with infants also suggested that the egocentric reference frame dominates in the coding of animate motions (Kushiro, Taga, & Watanabe, 2007).

In the domain of face perception, early evidence suggested that the egocentric system is the dominant frame of reference too. Kohler (1940) and Rock (1988) reported that the recognition of faces presented upright with respect to gravity and the visual-environment was impaired for observers with their heads held upside down. The finding of Troje (2003) that performance on a face recognition task depended only on stimulus alignment with the observer corroborates these early reports. However, recent findings by Lobmaier and Mast (2007) seem to suggest a role for gravity as a reference frame for the coding of faces.

To our knowledge, no studies thus far have provided a clear experimental distinction between the roles of gravitational and visual-environmental reference frames for the perception of biological motion and faces. Moreover, whether the global and local aspects of biological motion
are coded in the same reference systems is unknown. In the present study, we teased apart the contributions of three reference frames (egocentric, visual-environment, and gravity) by placing observers inside the York University “tumbling room” – a room furnished with strong directional, visual cues (e.g., table, chair, drapery), that can be rotated about a horizontal axis by 360 degrees (Figure 7.1A).

Figure 7.1. (A) Schematic depiction of the “tumbling room”. The room and observer can be rotated independently by 360 degrees about a horizontal axis. (B) The interior of the tumbling room. The display, located across from the observer’s chair is mounted at the axis of rotation.

Inside the room (Figure 7.1B), an observer can also be rotated (rolled) independently around the same axis. Using this facility, we investigated the perception of biological motion and faces by creating configurations in which two reference frames were put into conflict (aligned with or opposed to the stimulus) while the third was rendered uninformative by arranging it to be orthogonal to the stimulus. An entirely balanced design resulted in 12 experimental
configurations illustrated in Figure 7.2. If biological motion and faces are largely coded by egocentric (we don’t distinguish between the retina, head, or rest of the body here), gravitational, or visual-environmental coordinates, performance should be best when the stimulus is aligned with the respective reference frame.

We investigated the perception of both global motion-mediated structure in biological motion and local motion by manipulating the organization of the walker and the type of mask during a biological motion direction discrimination task. Specifically, the perception of global motion-mediated structure was addressed by placing veridical walkers inside a mask of additional walker dots moving in the opposing direction. This manipulation equated the local motion of the display. Consequently, the task could only be solved by retrieving the global form of the walker. The local aspect of biological motion was addressed by placing walkers that had their individual motion trajectories spatially perturbed (thereby destroying global structure) inside a mask of stationary flickering dots. These displays could thus only be solved based upon local motion cues. Additionally, we investigated reference frames for face perception via a same-different face recognition task previously employed by Troje (2003).
Figure 7.2. Schematic depictions of the 12 room/observer/stimulus configurations. The white arrow on the display represents the orientation of the stimulus where an upwards arrow is upright and downwards arrow is inverted with respect to gravity. For each configuration, the third reference frame was rendered uninformative by orienting it orthogonal to the stimulus orientation. The large arrows describe the reference frames in relation to the stimulus. Upwards and downwards arrows indicate that the stimulus was upright or inverted with respect to the observer (grey), gravity (black), or the room (white), respectively.
7.2 Methods

7.2.1 Participants

Twelve naïve observers 19 – 63 years of age (mean age of 34 years; 7 males, 5 females), recruited from the Centre for Vision Research at York University, participated in this experiment. All observers received monetary compensation for their time and had normal or corrected-to-normal vision. The procedures of this experiment were approved by both the York University Ethics Review Board and the Queen’s University General Research Ethics Board.

7.2.2 Stimuli and apparatus

The orientation of egocentric and visual-environmental cues with respect to gravity was manipulated using the York tumbling room (Figure 7.1). Both the room (2.4 m x 2.4 m x 2.4 m) and the observer’s chair could be rotated independently about a common axis which was normal to the fronto-parallel plane of the observer. Stimuli were generated using MATLAB (Mathworks, Natick, MA) with extensions from the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997) and displayed on a laptop equipped with a 12 inch screen mounted on the wall of the tumbling room that faced the observer and was centered on the axis of rotation. The observer was strapped securely to the chair and had a wireless mouse strapped to one leg. Stimuli were viewed binocularly at a distance of 147 cm.

Biological motion stimuli

The biological motion stimuli were derived from point-light sequences of a stationary-walking human, cat, and pigeon (Figure 7.3). The human walker was computed as the average walker from motion-captured data of 50 men and 50 women (Troje, 2002) and was represented
by a set of 11 dots. The cat sequence was created by sampling 14 points from single frames of a
video sequence showing a cat walking on a treadmill. The pigeon sequence was created from
motion-captured data obtained from a pigeon fitted with 11 markers. All walkers were presented
in sagittal view (facing either rightwards or leftwards). All sequences were shown at their
veridical gait frequencies of 0.93 Hz, 1.7 Hz, and 1.6 Hz for the human, cat, and pigeon,
respectively. On each trial, the starting position of the walker within its gait cycle was selected
randomly.

Figure 7.3. Static frames taken from point-light sequences used in the biological motion task
depicting a coherent human walker (A), and scrambled human walker in which the constituent
dots are spatially displaced (B). For the face recognition task, faces were shown turned either 5
deg or 10 deg in profile to the left or right of frontal view (C).

Walkers were presented coherently (Figure 7.3A) or spatially scrambled (Figure 7.3B).
Scrambled walkers were created by displacing each point’s trajectory to a randomly selected
position within the display area. Coherent walkers were embedded in a scrambled walker mask.
This mask was comprised of dots carrying veridical trajectories of the walker displaced randomly
on the screen and moving in a direction opposite to that of the target walker. The number of
mask dots was set to be identical to the number of dots comprising the target walker in order to
equate the directionality of the individual trajectories of all dots contained in the display. These
coherent-walker-and-scrambled-mask displays thus retained solely global, structure-from-motion information. Scrambled walkers were embedded in a flicker mask comprised of randomly positioned stationary dots with a limited lifetime of 125 ms. Twice the number of dots comprising the walker was used for the mask. These scrambled-walker-and-flicker-mask displays retained solely local motion information.

The walkers subtended visual angles of 2.8 x 5.2 deg, 5.2 x 2.8 deg, and 4.0 x 4.0 deg for the human, cat, and pigeon, respectively. The masking dots were contained in an area measuring 5.9 x 5.9 deg at the centre of the screen.

_Face stimuli_

The face stimuli were images of human faces (n = 30) derived from 3D models of laser-scanned heads (Troje & Bülthoff, 1996). The faces were shown turned either 5 deg or 10 deg to the left or right of frontal view. Hair was removed digitally and the images were rendered such that only the head and the upper section of the neck were visible (Figure 7.3C). Half of the faces used were female. The faces subtended visual angles of 4.4 x 5.6 deg on average.

_7.2.3 Procedure_

Two sequences of the twelve configurations shown in Figure 7.2 were run in two sessions separated by at least one day. During one session, observers completed a face recognition task and during the other session they completed a biological motion direction discrimination task. The order of the two tasks was counterbalanced among participants.
For each configuration subjects completed a block of trials that lasted approximately 2 minutes. The next configuration was then selected and the room and observer (with eyes shut) were rotated at constant velocity to the new orientation. When the chair was moved, the route that minimized time spent with the observer upside down was chosen. The order in which the configurations were run was randomized with the constraint that no two configurations involving an observer held upside down were run consecutively.

**Biological motion task**

On each trial, a single walker was presented at the centre of the screen and the observer was required to indicate whether the walker was facing to the right or left (relative to them). When the stimulus was rendered sideways with respect to the observer, the task was to indicate whether the walker was facing upwards or downwards. Instructions were given for each of the 12 configurations before each block of trials commenced. Each stimulus was presented on the screen for 1000 ms after which it was removed and replaced with a prompt (left/right, up/down arrows) on the screen until a response was provided.

Participants completed a practice block of trials with the orientation of the room, chair, and gravity all aligned in the upright position. The practice block consisted of 24 trials and was comprised of all possible combinations of left and right facing, coherent and scrambled versions of the human, cat, or pigeon presented upright or inverted with respect to each reference frame. For the practice trials only, the walkers were shown unmasked.

After the practice trials had been completed, the participant and room were rotated into the first configuration. The participants were instructed to initiate each test block by using the
mouse and to call out to the experimenter who was standing outside the tumbling room when the block was over. For the test trials, coherent walkers were embedded in a scrambled-walker mask (global cues) while scrambled walkers were embedded in the flicker mask (local cues). Each test block consisted of 36 trials and was comprised of all possible combinations of left and right facing, coherent and scrambled versions of the human, cat, and pigeon (each repeated three times) presented in random order. A total of 432 trials were completed across all configurations. Feedback was not given.

Face task

On each trial in this task, two faces were presented sequentially and the observer was required to indicate whether the faces were of the same or different person. Each face was presented at the centre of the screen for 100 ms and the two faces were separated by a 500 ms inter-stimulus interval. Following the disappearance of the face in the second interval, a prompt appeared and remained until a response was provided.

The face session began with a practice block comprised of 16 trials, consisting of male and female faces presented in upright or inverted orientations (with respect to each reference frame). The test blocks then followed. On all trials, the gender of the two faces was always identical. Half of the trials featured male faces while the other half featured female faces. On each trial, one face was presented facing 5 deg in one direction from a frontal view and the other was presented facing 10 deg in the opposite direction. Half of the trials presented the same individuals while the other half presented different individuals. Each test block consisted of 48 trials. A total of 576 trials were completed across all configurations. Feedback was not given.
7.3 Results

7.3.1 Global structure from biological motion

General performance and stimulus specificity

This section of the results refers to all trials showing coherent walkers in a scrambled walker mask requiring discriminations based upon the retrieval of global structure-from-motion. The data, collapsed across all 12 configurations with reference frames aligned, opposed, or orthogonal to the stimulus are presented in Figure 7.4A. Overall performance is relatively poor as some of the configurations resulted in chance-level performance (Figure 7.5). A one-way ANOVA revealed a significant main effect of animal type, $F(2,22) = 14.34, p = 0.003$. Post-hoc Tukey’s comparisons of the three animal types revealed that performance was better for the human walker than for the cat ($p = 0.002$) and pigeon ($p = 0.004$), which did not differ ($p > 0.500$).

This observation confirms earlier observations (Chang & Troje, 2009b), but is not directly relevant to understanding the contributions of reference systems. The different walkers were used to show that our results are not specific to a human structure and to introduce some variability in order to render it less likely that observers would use stimulus-specific artifacts to solve the task. For this reason, and to facilitate later comparisons between the different tasks, we subsequently pooled the data over walker types and examined them across the separate subsets of four blocks, constituting comparisons between pairs of reference frames.
Figure 7.4. Direction discrimination accuracies for the biological motion task, expressed in terms of the proportions of correct responses. The data for displays involving global structure (A), and local motion (B) cues for the three walker types and collapsed across all configurations (blocks) in which the reference frames are aligned, opposed, or orthogonal to the stimulus as shown in Figure 2. Error bars represent ± 1 standard error of the mean.

_Egocentric vs. gravitational reference frames_

The data for the blocks corresponding to a comparison between egocentric and gravitational frames of reference for the global-structure task are presented in the first panel of
Figure 7.5A. A 2 (stimulus upright/inverted with respect to observer) x 2 (stimulus upright/inverted with respect to gravity) ANOVA indicated a significant main effect of stimulus alignment with the observer, $F(1,11) = 27.34, p < 0.001$, reflecting the fact that accuracies were highest when the stimulus was aligned with the observer. There was no effect of stimulus alignment with gravity. Additional $t$ tests indicated accuracy rates for the two conditions in which the stimulus was inverted with respect to the egocentric system were not significantly different from chance-level ($p > 0.06$ for both).

*Gravitational vs. visual-environmental reference frames*

The second panel of Figure 7.5A shows the data corresponding to a comparison between the gravitational and visual-environmental reference frames for the global-structure sub-task. A two-way ANOVA indicated a significant main effect of gravity, $F(1,11) = 13.17, p = 0.004$. Performance was best when the stimulus was upright with respect to gravity. There was no effect of stimulus alignment with the room. Additional $t$ tests confirmed that accuracy rates for the two blocks in which the stimulus was upright with respect to gravity were in fact above chance-level ($p < 0.003$ for both). By contrast, accuracy rates for the two blocks in which the stimulus was inverted with respect to gravity were not significantly different from chance-level ($p > 0.5$ for both).
Figure 7.5. Proportions of correct responses for global biological motion discrimination (A), local biological motion discrimination (B), and face recognition (C) across the four blocks corresponding to comparisons between the egocentric and gravitational coordinates (left column), gravitational and visual-environmental coordinates (middle column), and the egocentric and visual-environmental coordinates (right column). The arrows below the x axes depict stimulus alignment with the relevant reference frames. Error bars represent ± standard error of the mean.
Egocentric vs. visual-environmental reference frames

The data corresponding to the blocks comparing egocentric and visual-environmental frames of reference for the global-structure sub-task are presented in the third panel of Figure 7.5A. A two-way ANOVA showed a significant main effect of stimulus alignment with the observer, $F(1, 11) = 48.46, p < 0.001$. Performance was highest when the stimulus was aligned with the observer. There was no effect of stimulus alignment with the room. Additional $t$ tests indicated that accuracy rates for the two blocks in which the stimulus was inverted with respect to the egocentric system were not significantly different from chance-level ($p > 0.15$ for both).

7.3.2 Local biological motion

General performance and stimulus specificity

This section refers to all trials showing scrambled walkers masked with flickering dots – stimuli containing solely local cues to indicate facing direction. The data for the three walker types, collapsed across all blocks are presented in Figure 7.4B. A one-way ANOVA showed no difference in performance rates across the various animals.

Egocentric vs. gravitational reference frames

The first panel of Figure 7.5B presents the data for the four blocks corresponding to a comparison between egocentric and gravitational frames of reference for the local sub-task. A two-way ANOVA indicated a significant main effect of stimulus alignment with the observer, $F(1,11) = 88.01, p < 0.001$. Accuracy was highest when the stimulus was aligned with the
observer. Stimulus alignment with gravity did not affect performance. Additional \( t \) tests indicated accuracy rates for the two conditions in which the stimulus was inverted with respect to the egocentric system were not significantly different from chance-level (\( p > 0.05 \) for both).

**Gravitational vs. visual-environmental reference frames**

The data corresponding to a comparison between the gravitational and visual-environmental reference frames for the local sub-task are shown in the second panel of Figure 7.5B. A two-way ANOVA for these blocks revealed a significant main effect of gravity, \( F(1,11) = 18.58, \ p = 0.001 \). Performance was best when the stimulus was upright with respect to gravity. There was no effect of stimulus alignment with the room. Additional \( t \) tests confirmed that accuracy rates for the two blocks in which the stimulus was upright with respect to gravity were in fact above chance-level (\( p < 0.004 \) for both). By contrast, accuracy rates for the two blocks in which the stimulus was inverted with respect to gravity were not significantly different from chance-level (\( p > 0.5 \) for both).

**Egocentric vs. visual-environmental reference frames**

The third panel of Figure 7.5B presents the data corresponding to the blocks comparing egocentric and visual-environmental frames of reference for the local sub-task. Here, a two-way ANOVA showed a significant main effect of stimulus alignment with the observer, \( F(1, 11) = 33.64, \ p < 0.001 \), reflecting the fact that performance was best when the stimulus was aligned with the observer. Again, there was no effect of stimulus orientation with respect to the room. Additional \( t \) tests indicated that accuracy rates for the two blocks in which the stimulus was
inverted with respect to the egocentric system were not significantly different from chance-level
\((p > 0.20\) for both).

### 7.3.3 Face recognition

**Overall performance**

The data for the face task, expressed in terms of proportions of correct responses, were
first analyzed by means of a paired \(t\) test comparing the two genders of the face stimuli. The
analysis indicated that performances for the male and female faces did not differ. Therefore, the
data were collapsed for further analyses.

**Egocentric vs. gravitational reference frames**

The first panel of Figure 7.5C shows comparisons between the egocentric and
gravitational frames of reference. A two-way ANOVA indicated a significant main effect of
stimulus alignment with the observer, \(F(1,11) = 160.33, p < 0.001\). Performance on the face task
was best in conditions where the stimulus was aligned with the observer rather than with gravity.

**Gravitational vs. visual-environmental reference frames**

The data corresponding to comparisons between the gravitational and visual-
environmental frames of reference for the face task are presented in the second panel of Figure
7.5C. Here, a two-way ANOVA showed no effects of gravity or the room.
Egocentric vs. visual-environmental reference frames

The data corresponding to comparisons between the egocentric and visual-environmental frames of reference are presented in third panel of Figure 7.5C. A two-way ANOVA indicated a significant main effect of stimulus orientation with respect to the observer, $F(1,11) = 107.41, p < 0.001$, reflecting the fact that accuracies were highest when the stimulus was aligned with the observer rather than the room.

7.4 Discussion

7.4.1 Biological motion and faces are primarily egocentrically coded

We found no significant differences in terms of the relevance of the various reference frames for discriminating the direction of the global and local biological motion stimuli. These two types of displays were designed such that different cues must be exploited to retrieve the facing direction of the walker. The “global” displays could be solved only by retrieving the coherent structure of the figure. Orientation effects observed for these particular displays must be due to impaired processing of this structure-from-motion information. By contrast, the “local” displays were devoid of structure and contained solely local cues to direction. The inversion effect observed for the perception of local biological motion has been shown to be carried by the acceleration contained in the foot motion (Chang & Troje, 2009a). The similarity of the data for the two types of displays suggests that both types of information are coded in a similar manner.

In line with earlier work (Troje, 2003), our results confirm that both biological motion and face perception are dominated by an egocentric frame of reference. For all types of stimuli, performance was best when the stimulus was aligned with the egocentric reference frame, regardless of the orientation of the observer with respect to gravity. Performance dropped
significantly when the stimulus was inverted with respect to the observer whether or not it was aligned with gravity or with the room.

Of particular interest are the results for the conditions in which the egocentric reference frame was rendered uninformative. For these conditions, the results obtained for both biological motion sub-tasks were different from those obtained for the face recognition task. Performance on the biological motion tasks (both for conditions requiring global structure-from-motion or local discriminations) was higher when the stimulus was aligned with gravity rather than the room suggesting contribution from a gravity-based reference frame. This was not the case for the face recognition task.

We note that although the biological motion and face recognition tasks used in this study differed in terms of their overall difficulty, it is unlikely that this difference can account for the differential effects of the gravity reference frame observed here. An inspection of the relevant comparisons of interest revealed that overall performance for the face recognition task (Figure 7.5C, middle panel) was significantly worse than for conditions in which the stimuli were aligned with the egocentric frame of reference (Figure 7.5C, left and right panels, first two data bars). This suggests that the lack of an effect of the gravity reference frame for the face recognition task is not due to mere performance saturation. Nonetheless, it should be acknowledged that the generalizability of these findings to other tasks, and more specifically, whether an effect of the gravity reference frame may be found for face perception using other tasks is unknown.

The data for both tasks are summarized in Figure 7.6A. For the face recognition task, it is clear that the data best fit the prediction made by the assumption that egocentric coordinates dominate perception. However, for the two biological motion tasks, the data show an additional
influence of stimulus orientation with respect to gravity. We therefore modeled the contribution of all three reference systems.

7.4.2 Modeling the data with a linear model

We fit the data with a simple linear model by modeling performance rate, \( r = x + eE + gG + vV \), where \( e \), \( g \), and \( v \) were weightings for contributions from the egocentric, gravitational, and visual-environmental reference frames, respectively, added to base performance \( x \). The variables \( E \), \( G \), and \( V \) assume values of \( 1 \) - \( -1 \), or \( 0 \) if the stimulus were aligned, opposed, or orthogonal to the reference frame, respectively.

The model was fit separately to the global biological motion, local biological motion, and face recognition data. The relative weightings obtained are summarized as percentages in Table 7.1. The weightings quantify the role of gravity for each task. For the local and global biological motion tasks, gravity contributes 21-25% of the effect but for faces the contribution is around 2%. The weightings allow us to predict performance for any combination of orientations of visual cues, the body, and gravity.

<table>
<thead>
<tr>
<th></th>
<th>Egocentric</th>
<th>Gravitational</th>
<th>Visual-environmental</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global BM</td>
<td>69%</td>
<td>25%</td>
<td>6%</td>
</tr>
<tr>
<td>Local BM</td>
<td>71%</td>
<td>21%</td>
<td>8%</td>
</tr>
<tr>
<td>Faces</td>
<td>89%</td>
<td>2%</td>
<td>9%</td>
</tr>
</tbody>
</table>

Table 7.1. Relative weightings of the three reference systems derived from the linear models fit to the biological motion discrimination and face recognition data. These are expressed as a percentage of the total contribution from all three reference frames tested.

The predictions for each condition are illustrated in Figure 7.6B. The variations in performance with reference frame alignment is well matched by the model for all three stimulus types (global BM: \( r^2 = .20 \); local BM: \( r^2 = .26 \); faces: \( r^2 = .37 \)).
Figure 7.6. (A) The experimental data are illustrated by the depth of shading for each condition. Cells with no shading represent lowest performance rates (50%) and shading gets darker as performance approaches 100%. (B) The output from the model plotted in the same format. Each column of the matrices correspond to comparisons between the egocentric and gravitational reference frames (E/G), gravitational and visual-environmental reference frames (G/V), and egocentric and visual-environmental reference frames (E/V), respectively.
7.4.3 The role of gravity for biological motion perception

Our data suggest that biological motion is predominantly egocentric-coded. For the retrieval of global structure of an articulated figure, there is no reason to expect involvement of additional allocentric reference frames. However, processing of local motion as a cue to the facing direction of a walker has been shown to be connected to expectations which the visual system has about the dynamics of the body in response to gravitational acceleration (Shipley, 2003; Troje & Westhoff, 2006; Chang & Troje, 2009a).

Of particular interest then, is the question of whether such “knowledge” about the effects of gravity is implemented in terms of a heuristic that assumes gravity is aligned with retinal coordinates, or if direct measures of the direction of gravity (taken either by the vestibular system or kinaesthetic sensors) feed into the visual system. Our data suggest that there is in fact a small but significant contribution from direct measurements of gravity not only for local motion processing but also for the retrieval of global structure from biological motion.

The contribution of gravity for the perception of biological motion is congruent with findings by Lopez et al. (2009) that judgments of stability of tilted human body postures also show a small influence of a gravitational reference frame. Direct involvement of the vestibular system for the perception of visual motion – in particular for motion influenced by gravity, has been demonstrated previously using fMRI (Indovina et al., 2005). In Indovina et al.’s study, motion that was coherent with gravity activated the vestibular network, including areas such as the insular cortex, temporoparietal junction, premotor and supplementary motor areas, middle cingulate cortex, postcentral gyrus, putamen, and the posterior thalamus, in addition to several visual motion areas. In contrast, motion that was inconsistent with gravity elicited responses mainly in the visual areas. These results suggest that the constraints of gravity have been
internalized and represented in the brain not just in terms of prior probabilities about the relations between egocentric, visual-environmental, and gravitational orientations, but also by the integration of direct measurements of the direction of gravity.

7.4.4 Conclusion

Inversion effects for the perception of faces and biological motion have been described pervasively in the literature. The inversion of a stimulus however can be described relative to a variety of reference frames including egocentric components (e.g., retina, head, body), gravity and the visual environment – all of which are usually aligned. Here, we teased apart the roles of these reference frames for the perception of faces and biological motion by systematically misaligning them. We showed that both biological motion and faces are largely coded in an egocentric frame of reference. The relative weightings of the gravitational and visual-environmental frames of reference however seem to depend on the relevance of the stimulus class to these reference frames. Unlike faces, dynamic events occurring on Earth are affected by gravitational force. Non-visual information about body orientation (reflecting this gravitational component) is integrated into the neural coding of these stimuli.
References


Chapter 8
General Discussion

8.1 On the importance of local motion

The projects described in this thesis were designed to investigate information retrieval from local biological motion. At the core of these studies was a hypothesis that the visual system is tuned to some invariant that is present in local biological motion, and in particular, in the motions of the feet. This invariant is suggested to be a cue that results as the foot interacts with gravity. Findings documented here of the importance of vertical acceleration for the perception of local biological motion are congruent with this hypothesis (Chapters 4 and 6).

It should be noted that apart from findings reported by Mather et al. (1991) and Troje and Westhoff (2006), very few studies of the perception of biological motion have focused on the importance of local motion information. Why might this be the case? Saunders, Suchan, and Troje (2009) have provided data to suggest that the relatively little emphasis made by the previous literature on local motion is due to the fact that in many instances, the stimuli employed by these studies simply did not contain the local invariant so discussed in the previous chapters. A significant portion of the literature involves the use of an algorithm established by Cutting (1978) for the synthetic generation of point-light displays depicting biological motion (e.g., Hiris, 2007; Lange & Lappe, 2007; Neri et al., 1998). Saunders et al. showed that stimuli generated in this manner in fact lack the local invariant inherent to the motion of the feet of a real walker. This invariant is present in natural walking motion and can be displayed in stimuli derived from motion-captured data of human walkers (Troje, 2002). In the experiment reported by Saunders et
al., observers were required to complete a walker detection task designed to address global structure-from-motion processing, in which walkers were embedded in a mask of additional walker dots. Additionally, observers were required to complete a second direction discrimination task designed to address local motion processing, in which coherent and scrambled walkers were presented in a mask of stationary, limited-lifetime dots. For both tasks, performances for stimuli derived from Cutting’s algorithm were compared with those for stimuli derived from motion-captured data. Results for the walker detection task indicated that performances did not differ according to stimulus type. By contrast, results for the direction discrimination task revealed that while observers could successfully discriminate the facing direction of scrambled walkers derived from motion-captured data, they could not do so for stimuli derived from Cutting’s algorithm, resulting chance-level performance rates. These data suggest that synthetic stimuli generated by Cutting’s algorithm lack the critical invariant exploited by the visual system to retrieve directional information from local biological motion.

8.2 Other emerging knowledge: Attention and peripheral vision

Several bodies of work are now beginning to emerge from other laboratories that place greater efforts towards understanding the perception of local biological motion. One line of work concerns the role of attention for biological motion perception (e.g., Thornton, Rensink, & Shiffrar, 2002; Jiang & He, 2007). This is a particularly interesting question in the context of initial propositions of Troje and Westhoff (2006) and Johnson (2006) on the role of the local motion mechanism to guide attention to conspecifics (or predators and prey). Such a role for the local motion mechanism was also described within a framework discussed in Chapter 3 of this thesis that suggested further once attention is guided to a relevant stimulus class, a second system
(the global, structure-from-motion mechanism) would then be responsible for acquiring detailed characteristics of the agent. Critically, this view predicts that the mechanism underlying the perception of local biological motion is pre-attentive in nature and can be contrasted with an attention-requiring global structure-from-motion mechanism. These predictions are supported by several behavioural findings. For example, Thornton, Rensink, and Shiffrar (2002) used a dual task paradigm to explore the role of attention for biological motion perception. Observers were required to complete a direction discrimination task from displays that contained, in addition to the walker, a series of rectangle arrays. On half of the trials, these rectangles changed orientation between success views. Observers were tested in two conditions: in the baseline condition, observers were instructed to monitor only the walker and to complete the direction discrimination task. In the dual-task condition, observers were required to monitor both the walker and the rectangles and to complete an additional change-detection task (i.e., determine whether any of the rectangles changed orientation). In one experiment, the walkers were embedded in two types of masks: a mask of additional scrambled walker dots that masked the local motion of the walker, and a mask of stationary dots with limited lifetime that retained access to both global structure-from-motion information and local motion information. The results revealed that under the condition in which observers were instructed to ignore the rectangles, performances were similar for both types of masks. More important however are the results for the condition in which observers were required to simultaneously complete the secondary change-detection task. Here, performance rates dropped to chance-level for walkers masked with additional scrambled walker dots, in contrast to above-chance-level performance for walkers masked with stationary dots. These results suggest that displays requiring the extraction of global structure-from motion
information require more attentional resources as compared to the displays containing local motion information.

Other studies also support a pre-attentive nature for local biological motion processing (Bosbach, Prinz, & Kerzel, 2004; Jiang & He, 2007). Jiang and He (2007) used an interocular suppression paradigm that exploits the fact that different stimuli require different lengths of time to emerge from noise. The length of time for the percept to emerge can be interpreted as an indicator of the salience of the stimulus. For example, it has been shown that upright faces emerge from suppression earlier as compared to inverted faces. In similar vein, familiar words emerge from suppression earlier as compared to unrecognizable words (Jiang, Costello, & He, 2007). In the study by Jiang and He, observers were presented with random noise in one eye and a scrambled walker in the alternate eye. They contrasted the time required for the walker to emerge from noise for conditions where the scrambled walker was shown either upright or inverted. The results revealed that the upright scrambled walker required significantly less time to emerge from suppression as compared to the inverted scrambled walker. As the authors further noted that the observers could not explicitly distinguish between upright and inverted scrambled walkers on a 2AFC task, they interpreted their findings to support automatic processing of local biological motion.

A second line of work that has emerged concerns the functioning of the local motion mechanism in the visual periphery. This line of work is also of particular relevance in light of the propositions set forth by Troje and Westhoff (2006) and in this thesis. If local information is a cue to the presence of an animate agent in the environment, then it would be advantageous for the visual system to possess a local motion mechanism that functions well in the visual periphery. The global structure-from-motion mechanism, in contrast, may function poorly in extrafoveal
regions. Indeed, there are empirical data to support such a dissociation between the global and local mechanisms in the visual periphery (e.g., Ikeda, Blake, & Watanabe, 2005).

A poorly functioning global structure-from-motion mechanism in the visual periphery is suggested by findings of Ikeda et al. (2005). In this study, observers were required to complete a two-interval walker detection task. In one interval, a coherent walker was presented and in the alternate interval, a scrambled walker was presented. In both intervals, the walkers were masked by additional walker point-lights that masked local motion and the task was to indicate the interval containing the coherent walker. Thus, this task required the retrieval of global structure-from-motion information. Across trials, the density of the mask was varied according to a staircase procedure and thresholds were obtained both for centrally and peripherally viewed stimuli. In addition, across conditions, the size of the walkers was varied to determine whether performance in the periphery could be matched to foveal performance by a simple magnification factor. The results showed that thresholds deteriorated at large eccentricities. Further, for both centrally and peripherally presented stimuli, thresholds reached asymptotes with stimulus size and peripheral thresholds could not be equated with foveal thresholds by size scaling. Interestingly, the authors also manipulated the orientation of the stimulus and found that the inversion effect (as reflected by poorer thresholds at the inverted as compared to upright orientation) was present only in foveal vision. These data suggest that the ability to retrieve global structure from biological motion is impaired in the visual periphery.

Still, additional data suggest that the perception of biological motion in the periphery depends very much on stimulus parameters. For example, observers can easily perceive coherent point-light walkers in the visual periphery if the stimuli are shown unmasked (Thompson, Hansen, Hess, & Troje, 2007). In this experiment, observers were required to discriminate the
facing direction of coherent and scrambled walkers (unmasked) presented in the periphery. The results indicated that in the absence of noise, the visual system is well able to extract directional information from both coherent and scrambled displays. In a subsequent experiment reported in this study, the authors investigated the effects of various masks on the perception of coherent and scrambled walkers in the visual periphery. Here, observers were required to complete a detection task (i.e., indicate which of two presented intervals contained a coherent walker). Three types of masks were used: a mask of additional scrambled walker dots, a mask of linearly moving dots, and a mask of stationary, limited-lifetime dots. The results indicated that performances for all mask types were generally very different for peripheral versus central vision. Specifically, performances for stimuli presented in the periphery deteriorated more substantially with increases in mask density as compared to performances for centrally presented stimuli. Of particular interest is their finding that peripheral vision was also particularly poor at detecting the walker when masked by additional scrambled walker dots that rendered local motion uninformative, as compared to other masks, suggesting that peripheral vision is more sensitive to biological motion when local cues are retained. The authors argued that the visual system is in fact well equipped to perceive biological motion in the periphery (particularly, local information in biological motion) but peripheral vision, as compared to foveal vision suffers from a problem of poor figure-ground segregation ability.

Yet another study provides further support for the functioning of biological motion mechanisms in the periphery (Gurnsey, Roddy, Ouhnana, & Troje, 2008). Observers in this study were presented with stimuli that were walking just several degrees from the line of sight and were required to complete two separate tasks: a direction discrimination task and a walker identification task that required observers to indicate which of five different walkers were
presented. The results revealed that for both tasks, performances for peripherally presented stimuli could be equated to those for centrally presented stimuli by a simple size-scaling of the stimulus. That is, contrary to initial claims of Ikeda et al. (2005), stimulus magnification is in fact sufficient to equate the perception of biological motion in peripheral and central vision.

8.3 Further considerations: Neural concomitants

The many behavioural studies outlined above have contributed to a better understanding of the nature of the mechanisms underlying the perception of biological motion. Noticeably lacking, however, is a clear definition of the neural substrates underlying the perception of biological motion and in particular, local biological motion.

The posterior superior temporal sulcus (STS) has often been implicated in neuroimaging studies of biological motion perception (e.g., Bonda, Petrides, Ostry, & Evans, 1996; Grossman & Blake, 2001; Grossman et al., 2000; Vaina, Solomon, Chowdhury, Sinha, & Belliveau, 2001). For example, using functional magnetic resonance imaging (fMRI), Grossman et al. (2000) showed that the STS is responsive to coherent point-light displays but is not activated by scrambled point-light displays, coherent motion, or kinetic boundaries. Using positron emission tomography (PET), Bonda et al. (1996) also showed this area to be active when observers viewed point-light animations. More recently, the neural networks underlying the perception of biological motion containing solely global structure-from-motion information were more explicitly probed (Michels, Lappe, & Vaina, 2005). In this study, Michels et al. used fMRI to compare areas activated in response to sequential point-light stimuli that did not contain local motion (as introduced by Beintema and Lappe, 2002) and point-light stimuli with intact local motion. They found that STS activation was similar for both types of stimuli, suggesting that STS activation is not dependent on local motion signals. Importantly, stimuli without local
motion also resulted in increased activation in the fusiform gyrus and the extrastriate body area (EBA) (Downing, Jiang, Shuman, & Kanwisher, 2001).

Responses in regions outside of the posterior STS for biological motion have also been demonstrated by others. In addition to the fusiform gyrus (Grossman et al., 2004; Santi et al., 2003), point-light stimuli also activate the posterior inferior temporal sulcus and middle temporal gyrus (Michels et al., 2005; Peuskens et al., 2005; Saygin et al., 2004). The involvement of such areas for biological motion perception has been argued to reflect engagement of areas that are selective for static images of the human body (Peelen, Wiggett, & Downing, 2006). In their study, Peelen et al. (2006) first presented observers with point-light animations of simple whole-body actions and scrambled versions of the same sequences in order to identify gross regions that respond to biological motion. Functional regions of interest were also identified in each participant with additional experiments that compared responses to oscillating and static low-contrast rings (to localize motion-sensitive middle temporal cortex, hMT+), and responses to bodies, faces, scenes, and tools (to localize body-selective extrastriate body area, EBA, fusiform body area, FBA, and face-selective fusiform face area, FFA). Consistent with previous studies, Peelen et al. found activation in the right posterior superior temporal sulcus in response to biological motion. Responses were also found in the bilateral posterior inferior temporal sulcus (pITS), close to typical coordinates for EBA and hMT+, and the posterior fusiform gyrus (pFG), close to typical coordinates of the FFA and FBA. Further analyses within each region of interest confirmed that the EBA, hMT+, FBA, and FFA were all selective for biological motion. The authors subsequently determined voxelwise correlations between biological motion responses and responses for the localizer stimuli for each region of interest, reasoning that if responses in hMT+ were due to responses of motion-selective neurons for example, there should be a positive
correlation between responses for the motion localizer and biological motion stimuli. Critically, the analysis revealed that in all regions of interest, biological motion responses were significantly correlated with body-selectivity. They suggested that while posterior STS may be sensitive to particular patterns of posture changes, involvement of additional areas reported in the literature are due to the presence of body (form) selectivity in these regions.

Additional complications for attempts to understand the neural substrates of biological motion perception are findings that regions of activation are significantly affected by stimulus and task demands (Jastorff & Orban, 2009). In one experiment, Jastorff and Orban (2009) attempted to separate processing of form information from the processing of kinematics from biological motion. Global structure-from-motion was manipulated by spatially scrambling the walker and kinematics were manipulated by changing the trajectory of each dot to simple translation. In this initial experiment, observers were instructed to maintain passive fixation. The results indicated a separation of regions for form and motion processing. Form processing was reflected in the ventral region (lateral occipital sulcus to posterior inferior temporal gyrus). The processing of kinematics was reflected in a dorsal region including the posterior inferior temporal sulcus and the posterior superior temporal sulcus. In a subsequent experiment, the authors introduced more complex actions such as boxing, throwing, and kicking. In addition, an attentionally demanding one-back task was introduced. The results revealed that the body areas were not affected by task demands. By contrast, areas such as the posterior STS were significantly affected by the complexity of the stimuli and task demands.

In contrast to the many studies that have focused on the retrieval of global form from biological motion, very little is known as to the neural components underlying the perception of local cues contained in biological motion. Electrophysiological data have shown that scrambled
point-light walkers are reflected in a different event-related component as that for coherent animations (Hirai & Kakigi, 2008). In this study, Hirai and Kakigi (2008) used an adaptation paradigm to investigate how local and global aspects of biological motion modulate ERP components. Observers were first allowed to adapt to a coherent point-light stimulus, its spatially scrambled version, or a static control stimulus. Following adaptation, observers were presented with coherent or scrambled walkers during which event-related potentials were measured. The authors reasoned that attenuation of the signal following adaptation would be indicative of a component’s sensitivity to the stimulus. They identified two negative components that responded to the stimuli and modulated differently: an N1 component (around 200 ms after stimulus onset), and an N2 component (240-400 ms after stimulus onset) in the bilateral occipitotemporal region. Notably, the N1 signal was attenuated by both coherent and scrambled stimuli, both when the adaptor was a coherent or scrambled walker (but not the static control). By contrast, the N2 signal was attenuated only when the adaptor and test stimuli both involved coherent walkers. These data were interpreted to indicate that the N1 component is sensitive to local motion while the N2 component is sensitive to the presence of global structure-from-motion information in biological motion. Still, while these data provide insight into the neural dynamics underlying local motion processing, the exact substrates underlying these dynamics remain to be elucidated.

8.4 Closing: Returning to the big picture

In closing, one should be reminded that the local motion mechanism at the focus of this thesis certainly constitutes only a fundamental piece of a highly complex perceptual process. Troje (2008) suggested that at least four processing levels should be differentiated for the perception of biological motion. Forming the early levels are two mechanisms that have already
been discussed here: a mechanism that is sensitive to local invariants in biological motion and is hypothesized to enable attention to be drawn to conspecifics, prey, or predators, and a second mechanism involving global structure-from-motion retrieval that is suggested to enable the recognition of the particular class of the agent (as discussed in Chapter 3). Beyond these stages, Troje suggested the existence of at least two other levels: one dedicated to classifying categories of action and a second dedicated to retrieving further detailed information about the agent such as identity, gender, and emotion. The experiments documented in this thesis have focused on just one of these fundamental stages. Achieving a full understanding of the phenomenon of biological motion perception will require further empirical efforts towards characterizing the other individual stages and understanding how they interact.
References


Summary and Conclusions

Biological motion perception has traditionally been considered to be governed by a single mechanism. The finding of distinct sources for the inversion effect in biological motion perception, one that is based upon the retrieval of global structure-from-motion information and a second that is based entirely on local motion signals has rendered it necessary to consider the contributions of multiple mechanisms. The experiments described in this thesis were designed to investigate specifically information retrieval from local biological motion signals. These experiments have shown that local biological motion carries significant information as to the facing direction of the agent and as to the animacy of a stimulus (Chapter 2). The mechanism underlying the perception of local biological motion can also be dissociated from that underlying the perception of global structure-from-motion information according to characteristics such as susceptibility to learning and noise (Chapter 3). Further, the orientation-specificity of perceiving local motion has been shown to be associated with vertical acceleration contained in the stimulus – a finding congruent with suggestions that the visual system exploits knowledge about the effects of gravity on the biomechanics of animate motions for perception (Chapters 4-6). The use of gravitational heuristics, however, does not necessarily implicate direct measurements of gravity for the perception of biological motion. To this end, data reported in Chapter 7 suggest that non-visual information about gravity, perhaps via vestibular input is in fact integrated for the perception of biological motion. These findings contribute to a greater understanding of a mechanism proposed to serve as a fundamental first stage towards interpreting animate motion patterns.
Appendix A
Permission of Co-Authors

I/we, the undersigned, hereby grant permission to microfilm any material designated as being co-authored by me/us in the thesis copyrighted to the person name below:

<table>
<thead>
<tr>
<th>Name of copyright author:</th>
<th>Signature of copyright author:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Donta H. F. Chang</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Name(s) of co-author(s):</th>
<th>Signatures of co-author(s):</th>
</tr>
</thead>
<tbody>
<tr>
<td>Laurence R. Harris</td>
<td></td>
</tr>
<tr>
<td>Nikolaus F. Troje</td>
<td></td>
</tr>
</tbody>
</table>

DATE: 4th Jul 2010
Appendix B

Copyright Waiver

20-April-2010

Dear Dorita,

Permission is granted to use the following materials in your thesis, provided that Journal of Vision is cited as the source, and ARVO as the copyright holder.


--
Wishing you every success!
Sincerely,
Alice

Alice O'Donnell
Asst Director of Journals,
ARVO
aodonnell@arvo.org
240-221-2930
Appendix C

Supplementary material for Chapter 6: Determining the Fourier series

As the search space for the adequate local biological motion described in Chapter 6 was defined by parameters derived from a Fourier series (Troje, 2002), a preliminary experiment was first conducted to determine whether the number of harmonics represented in the series would significantly affect the saliency of directional information conveyed by the walkers. Here, stimuli generated using second-, third-, and fourth-order Fourier series were compared using a psychophysical direction discrimination task. The preliminary data reported here then determined the magnitude of the search space in the experiment proper.

Methods

Participants

Eighteen observers (10 males; mean age = 19.3 years) with normal or corrected-to-normal vision participated in this experiment.

Stimuli

Stimuli were generated using second-order, third-order, and fourth-order Fourier series. For each series, stimuli were derived from point-light sequences of a walking human, cat, and pigeon. All walkers were presented in sagittal view (i.e., facing rightwards or leftwards) and displayed stationary walking as if walking on a treadmill. Point-light sequences were shown at their veridical speeds with gait frequencies of 0.93 Hz, 1.7 Hz, and 1.6 Hz for the human, cat, and pigeon, respectively.
The point-light walkers were presented upright or inverted about the horizontal axis and were coherently organized or spatially scrambled. All walkers were embedded in a mask of 125 randomly-positioned stationary dots with a limited lifetime of 125 ms.

The stimuli were generated using MATLAB (Mathworks, Natick, MA) with extensions from the Psychophysics Toolbox, and were displayed on a 22 inch ViewSonic P220f CRT color monitor with 0.25 mm dot pitch, 1280 x 1024 pixels spatial resolution, and 100 Hz frame rate. All stimuli appeared as white dots on a black background and the point-light figures subtended visual angles of 2.1 x 4.6 deg, 4.6 x 2.4 deg, and 3.6 x 3.6 deg for the human, cat, and pigeon, respectively.

**Procedure**

A direction discrimination task in which observers were required to indicate the facing direction of the walker (left or right) was used in this experiment. Participants began by completing a practice block of trials to familiarize themselves with the task. This block was comprised of 24 trials that included all possible combinations of the three walker types (human, cat, pigeon), two orientations (upright or inverted), two stimulus organizations (coherent or scrambled), and two facing directions (left or right). All stimuli were masked as described in the previous section and presented in random order. The stimuli for the practice block were defined by a fourth-order series.

After the practice block, participants completed the experiment proper. The experiment proper consisted of 360 trials and was comprised of all possible combinations of the three walker types, two orientations, two stimulus organizations, and two facing directions repeated five times for each Fourier series (second-, third-, and fourth-order). All stimuli were embedded in the
limited-lifetime mask and presented in random order. Stimuli were presented for a duration of 500 ms.

**Results**

Performances on the direction discrimination task, expressed in terms of the proportion of correct responses for the different series and walker types are presented in Figure C.1. From this figure, it is evident that performances did not change according to the number of harmonics represented in the series. This observation was confirmed by statistical analyses.

![Figure C.1](image)

Figure C.1. Performances on the direction discrimination task, expressed in terms of the proportions of correct responses across the different number of harmonics represented and walker types. Error bars represent standard ± 1 standard error of the mean.

A 3 (series) x 3 (walker type) x 2 (organization) x 2 (orientation) ANOVA revealed significant main effects of organization, $F(1,17) = 74.62, p < 0.001$, reflecting the fact that performances were generally worse for scrambled as compared to coherent displays, orientation, $F(1,17) = 101.42, p < 0.001$, reflecting the fact that performances were worse when the stimuli were presented upside-down rather than upright, and a significant organization by orientation
interaction, $F(1,17) = 27.61, p < 0.001$. There was no effect of series and no interactions involving this factor.

In Figure C.2, the data are plotted for the two stimulus organizations and orientations. The interaction was investigated by means of a paired $t$ test comparing differences in accuracy rates between the upright and inverted conditions for coherent and scrambled displays. The analysis revealed that the interaction was due to the fact that the difference in accuracy rates between the upright and inverted conditions was larger for the coherent displays as compared to the scrambled displays, $p < 0.001$.

Figure C.2. Performances on the direction discrimination task for the two stimulus organizations and two orientations, collapsed across walker type and series. Error bars represent ± 1 standard error of the mean.

**Conclusion**

The data from this experiment suggest that there is no significant difference in the saliency of directional information conveyed by biological motion (including displays conveying solely local cues) among second-, third-, and fourth-order representations of the motions.