

LOCAL MOTION SIGNALS - PREVALENCE, RESPONSES AND INTERACTIONS

A Dissertation

Presented to the Faculty of the Graduate School
of Cornell University

in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy

by

Eyal Izhak Nitzany

August 2015

© 2015 Eyal Izhak Nitzany

ALL RIGHTS RESERVED

LOCAL MOTION SIGNALS - PREVALENCE, RESPONSES AND INTERACTIONS

Eyal Izhak Nitzany, Ph.D.

Cornell University 2015

Extraction of local motion signals is crucial for our survival. Lack of information from local motion signals will significantly reduce our ability to discriminate objects from background, avoid obstacles, and navigate.

Despite the apparent effortless with which we perceive visual motion, there are indications that the underlying neural computations are complex. Three kinds of local motion signals have been distinguished, based on the kinds of spatiotemporal correlations that generate them: Fourier (F), based on 2-point correlations [1]; non-Fourier (NF), based on 4-point correlations [2]; and glider (G), based on 3-point correlations [3]. G signals have two subtypes, expansion and contraction, associated with objects that are looming and receding, respectively. Detection of isolated G and NF signals cannot be mediated by a purely multiplicative cross-correlator or a purely quadratic motion energy model. G signals have recently attracted substantial attention, following the demonstration that a wide range of species (human [3], macaque [4, 5], zebrafish [6], dragonfly [5], and fruitfly [7]) respond to them in similar ways suggesting that there are advantages to using these signals in visual tasks.

This work expands the above lines of research in several respects. First, our computational work shows that these motion signals appear in natural scenes and characterizes the basic statistical relationships between them [8]. Second, we report neurophysiological recordings in two distinct visual-specialist

species (macaques and dragonflies) that demonstrate that at the neuronal level, cells response in a similar manner to motion signals in many respects, although there are subtle differences in responses between the species. This convergence at the algorithmic and neural-implementation levels indicate the fundamental biological importance of using the many kinds of motion signals to guide behavior. Finally, we carried out a psychophysical experiment to probe human ability to use multiple kinds of local motion signals simultaneously to solve simple directional task. We found that humans can combine different kinds of motion signals to solve this task, and, interestingly, that sensitivity to different kinds of motion signals is context-dependent.

BIOGRAPHICAL SKETCH

I was born and raised in Rishon Le-Zion in Israel. I always loved mathematics and therefore it was natural for me to obtain my B.A. in mathematics and computer science at the Technion. Later on, I became interested in biology and pursued a Masters degree in biostatistics at Tel-Aviv University. In my Masters thesis, under the supervision of Profs. Melijson and Hammel from Tel-Aviv University, I developed a statistical model for the size of granules in different cells [9].

I then worked for several companies in high-tech industries in Israel, mainly as programmer, algorithm designer and project manager, until I decided to follow my passion for research.

My graduate work consisted of a cluster of collaborative multidisciplinary studies of how real-world visual motion signals are analyzed by the visual system. In parallel, I explored how abstract causal relationships are learned in a dynamical, interactive virtual environment.

The thesis research makes use of computational analysis, psychological experiments in humans, and physiological experiments in non-human primate, as well as in other model systems (dragonfly, wasp, jumping spider, and more). I am fortunate to work with leading scientists on my committee and in the Department of Neurobiology and Behavior at Cornell.

I would like to dedicate this work to Michal, my wife, who supports me in many ways. I could not have done this without you.

ACKNOWLEDGEMENTS

I would like to thank my committee members: Profs. Jonathan Victor, Shimon Edelman and James Cutting. In particular, I would like to thank Jonathan for the numerous hours of detailed discussion, for the patience to hear many ideas and the encouragement to pursue a few of them. In addition, Jonathan introduced me to the fascinating world of experimental neuroscience and specifically to neurophysiology in the non-human primate. This was a wonderful experience and I'm grateful for that.

Shimon introduced me to many computational ideas and concepts and allowed me to share his unique uniquely perceptive insights about how things work. I'm grateful for that. Shimon and Jonathan, I feel that you assisted in shaping me as a scientist.

I would also like to thank, the third member of my committee, James Cutting for assisting me with several conceptual issues and for sharing his database. I found the perception lunch very useful and enriching. I always enjoyed hearing (and occasionally contributing to) the diverse ideas around the table.

I was fortunate to work in close collaboration with Prof. Ronald Hoy from the Department of Neurobiology and Behavior at Cornell. He introduced me to the world of insects and arachnids and allowed me to be part of several projects. I spent many hours in his lab and with other graduates students there. In particular, I'm grateful for the many hours of experiments and discussions shared with Gil Menda, Paul Shamble and James Golden.

I would also like to thank my program — CBM (Computational Biology & Medicine) — for supporting me over the years, and in particular to thank the program director, Dr. David Christini, and the coordinators, Mrs. Sue Bishop and Mrs. Margie H. Mendoza.

Finally, I would like to thank my supportive environment. I could not have completed this journey without my wife Michal. Also, I thank my kids who gave me strength and inspiration whenever I needed it, and my friends and colleges who were there for me.

TABLE OF CONTENTS

Biographical Sketch	iii
Dedication	iv
Acknowledgements	v
Table of Contents	vii
List of Tables	ix
List of Figures	x
1 Introduction	1
1.1 Chapter by chapter summary	3
2 The statistics of local motion signals in naturalistic movies	6
2.1 Abstract	6
2.2 Introduction	7
2.3 Materials and methods	9
2.3.1 Data	9
2.3.2 Quantification of motion signals	10
2.3.3 Combining local motion signals within a shot	15
2.4 Results	22
2.4.1 Covariation of motion signals	31
2.5 Discussion	33
2.6 Supplementary Materials	40
2.6.1 Motion scores: supplementary analyses, properties and a note on analysis of synthetic motion stimuli	40
2.6.2 Movies	61
3 Evolutionary convergence of motion computation in monkey and dragonfly	63
3.1 Abstract	63
3.2 Introduction	63
3.3 Results	65
3.4 Discussion	74
3.5 Methods	76
3.5.1 Physiology and Recording Methods	76
3.5.2 Motion Stimuli (both species)	81
3.5.3 Data analysis	84
3.6 Supplemental Information	85
3.6.1 Additional analyses	85
3.6.2 Alternative definition of responsiveness to each kind of motion.	85
3.6.3 Analysis of combination of motion signals, subdivided by brain area.	85
3.6.4 Video. Demonstration of the motion stimuli used in these experiments.	90

4	Perceptual interaction of local motion signals	91
4.1	Introduction	91
4.2	Methods	92
4.2.1	Visual stimuli	93
4.2.2	Data analysis	97
4.2.3	Subjects and display	98
4.3	Results	99
4.4	Discussion	104
4.4.1	Summary	106
5	Discussion and future work	108
5.1	Discussion and future directions	108
5.1.1	Specific predictions and suggested frameworks	113

LIST OF TABLES

2.1	Clips details	62
2.2	Annotation scheme for clip 2F	62
4.1	Examined motion types.	94

LIST OF FIGURES

2.1	A summary of calculation of the basic local motion scores	13
2.2	A summary of calculation of the RMO and PMO local scores in a spatiotemporal ROI, using NF-S motion as an example	20
2.3	Prevalence of different kinds of motion signals is similar across movies.	24
2.4	Prevalence of different kinds of motion signals, analyzed at a coarse spatial scale, is similar across movies.	25
2.5	The effect of binarization on local motion scores	26
2.6	Prevalence of different kinds of motion signals is similar across binarized movies.	28
2.7	Prevalence of different kinds of motion signals, analyzed at a coarse spatial scale, is similar across binarized movies.	29
2.8	Prevalence of different kinds of motion signals is similar across movies, as measured by the PMO score.	30
2.9	Prevalence of different kinds of motion signals is similar across movies, as measured by the RMO score.	31
2.10	Covariance patterns of motion scores (<i>YT</i> ; "The 39 steps").	33
2.11	Dependence of motion signals on spatial discretization, for the different motion scores.	42
2.12	Dependence of motion signals on spatial discretization, for the different motion scores.	43
2.13	Dependence of motion signals on spatial discretization, for the PMO score.	44
2.14	Dependence of motion signals on spatial discretization, for the RMO score.	45
2.15	Prevalence of different kinds of motion signals is similar across movies, as measured by the PMO score in the XT-plane.	46
2.16	Prevalence of different kinds of motion signals is similar across movies, as measured by the RMO score in the XT-plane.	47
2.17	Prevalence of different kinds of motion signals is similar across movies, as measured by the PMO score.	48
2.18	Prevalence of different kinds of motion signals is similar across movies, as measured by the RMO score.	49
2.19	Covariance patterns of motion scores (<i>YT</i> ; "Mr. & Mrs. Smith").	51
2.20	Covariance patterns of motion scores from 500-frame scenes drawn from a random movie.	52
2.21	Covariance patterns of motion scores (<i>YT</i> ; "A Night At The Opera")	52
2.22	Correlations between RMO Fourier score, PMO Fourier score, and output of a Reichardt detector.	54
2.23	analysis of artificial stimuli generated on a grid.	60

3.1	Analysis of single unit responses to elementary motion signals in the brain of macaque monkeys and dragonflies.	66
3.2	In dragonflies (top) and macaques (bottom), similar fractions of central neurons respond to Fourier, glider, and non-Fourier motion signals.	69
3.3	In dragonflies (blue) and macaques (red), combination of motion signals at a cellular level is similar.	70
3.4	In macaques and dragonflies, motion processing unfolds in a similar fashion.	74
3.5	Summary of motion stimuli.	83
3.6	Fraction of neurons with a significant response to one or more kinds of motion for dragonflies (A) and macaques (B).	86
3.7	Comparison of direction selectivity index (DI) for Fourier (standard) motion, and each of the other subtypes, for macaque neurons localized to either V1 or V2.	88
3.8	Comparison of direction selectivity index (DI) for Fourier (standard) motion, and each of the other subtypes, for dragonfly neurons localized to either medulla or lobulla.	88
3.9	Comparison of direction selectivity index (DI) for Fourier (standard) motion, and each of the other subtypes, for dragonfly medulla and macaque V1 neurons.	89
3.10	Comparison of direction selectivity index (DI) for Fourier (standard) motion, and each of the other subtypes, for dragonfly lobula and macaque V2 neurons.	89
4.1	Stimulus examples	96
4.2	Psychophysical performance for direction judgments for stimuli containing mixtures of Fourier (F) and glider (G) signals.	100
4.3	Fits of Weibull functions to performance along each ray.	102
4.4	Isodiscrimination curves for combinations of Fourier (F) and glider (G) contraction.	103
4.5	Comparison of sensitivity to Fourier (F) motion in the context of glider (G) contraction (red) and expansion (green).	103

CHAPTER 1

INTRODUCTION

The ability to detect motion cues in a visual environment is arguably one of the most critical tasks for the visual system and is necessary for navigation, obstacle avoidance, and completing virtually any task that requires targeting. In other words, extraction of local motion cues is essential for our ability to survive.

Motion can be thought of as a form of structured change, and thus occurs in a wide variety of environments and sensory domains, including audition and vision. Motion extraction is a complex task that is traditionally considered to begin with extraction of local motion cues that are followed by later advanced processing such as abstraction of information, and ends with actions. Many species, including humans, seem to perform these actions effortlessly, but it requires a lot of infrastructure and computational power. This task is not a one-way road, but rather a bidirectional path. The more obvious direction is that local motion cues serve as inputs to systems that try to abstract rules to guide motor action. The less obvious direction is that different kinds of rules are expected to be applicable in different context and tasks [10], and an online Bayesian update rule can be used to identify the specific rules - for example, the kinds of local motion signals – that are extracted. Michotte [11] had demonstrated in a series of experiments on the perception of causality that different motions and interactions between moving objects affect the way people perceive the relationships between the objects. In other words, as expected, generating abstract rules is dependent on the types and nature of the inputs.

This work presents three studies that explore various different aspects of

local motion signals. I used several classical neuroscience modalities: computational, ethological, neurophysiological, and psychophysical approaches. The thesis begins with computational analysis of local motion cues in natural stimuli, continues with neurophysiological experiments on neurons in different species to determine their responses to local motion signals, and ends with a study about perceptual integration of two kinds of local motion cues.

These studies are weaved together, as results from one influences and motivates the other. Our first computational study shows that various kinds of motion signals appear in the natural environment and are correlated with each other. This motivates us to explore whether neurons are responding/sensitive to these different kinds of motion. In addition, as we observed that different kinds of motion co-occur spatially, we ask whether organisms that rely on visual motion take advantage of this fact. We ask this at the cellular level in two widely divergent species (macaque and dragonfly), and we ask this at the perceptual level in humans. The basic question is, whether different motion signal types are processed together or separately?

Thus, this work should be viewed in the light of Marr's [12] framework. In his classic book, "Vision", Marr proposed three different levels for the understanding of information processing systems: computational theory; representation and algorithm; and hardware implementation. Note that the first part computational theory includes elucidation of why the computations are done. Thus, all three of Marr's layers are relevant here.

The term "local" in local motion cues deserves special attention. Motion signals are "local" in the concrete sense of the spatiotemporal relationship between the elements that generate the motion signal. Yet this "locality" also ap-

plies in a temporal manner, that becomes apparent when we deal with formation of abstract rules.

We speculate that there is a fundamental reason that visual-specialist species process motion in a local manner (in both of the above senses). Had a global processing strategy been used, the curse of dimensionality would apply. But local computations scale favorably as resolution increases, which is crucial for systems with high spatial or temporal resolution.

1.1 Chapter by chapter summary

The starting point for the analysis of local motion signals is the approach developed by Hu & Victor [3], which showed that humans are sensitive to a wide variety of mathematically-distinct local motion cues. We take this approach, which constructed artificial movies that isolated individual local motion cues, and modify it so that it can dissect and quantify the kinds of motion cues in the natural environment. This extends the "exemplar" approach to analyzing local motion signals introduced by [2, 3].

The rest of this work is in the following order: in chapter 2, we examine the kinds of motion cues in a relatively natural visual environment: Hollywood movies. In chapter 3, we analyze how local motion cues are processed at the cellular level and in chapter 4 we investigate their processing at the perceptual level. We close with discussion and suggestions for future research in chapter 5.

Here I present a short summary of the chapters to follow. Chapter 2 belongs to the first level of Marr framework (i.e. computational theory), and ad-

dresses the question of why visual systems are sensitive to multiple kinds of motion signals. At first glance, this might seem surprising, since it has long been established that the Reichardt model [1], which extracts only a single, simple kind of motion signal, suffices to account for a large number of neurophysiologic and perceptual observations, across a wide range of species. However, it is also well-recognized that there are other kinds of local motion cues that drive visual motion perception: the "non-Fourier" cues identified by Chubb & Sperling [2], and the "glider" motion cues identified by Hu & Victor [3]. Specifically with regards to the glider cues, Fitzgerald et. al. [13] advanced a theoretical argument that incorporation of glider cues could make biological motion processing more efficient, provided that they are present in the natural environment. The work reported in Chapter 2 shows that this is the case. This chapter was published in the Journal of Vision [8].

Chapter 3 belongs to the second and third layers of Marr framework (algorithm and hardware) and addresses questions about the neural representation of local motion signals. We recorded activity of single units (spiking neurons) from two hierarchically sequential visual brain areas in the central nervous system of two drastically divergent visual specialists: dragonfly and macaque monkey. Neural responses were probed with visual stimuli that isolate several different kinds of local motion signals. We found that neurons in visual brain areas of the two species responded to this library of stimuli with striking similarity in several respects. First, similar fractions of neurons in both species responded to each kind of motion signal, and to specific subsets of motion signals. Second, neurons that responded to multiple kinds of motion were found to respond to those signals with similar direction selectivity. Finally, characteristics of motion processing changed similarly in the two species, as processing

unfolded from lower to higher brain areas (V1 to V2 in macaque and medulla to lobula in dragonfly). Notably, in both species, motion opponency, a fundamental characteristic of the classical Reichardt model for detecting motion [1, 14], held in the earlier stage, but the second stage showed systematic deviations from opponency. This chapter is in preparation for submission and parts of it were presented at SFN 2013 [4] and CoSyNe 2014 [5].

Chapter 4 also belongs to the second layer of Marr framework (i.e. representation and algorithm). Since multiple motion signals are typically present in natural stimuli, here we ask how these signals combine perceptually. To do this, we determine thresholds for detecting motion direction in synthetic stimuli that contain controlled combinations of two kinds of motion signals. In a two-alternative force-choice psychophysical experiment, we find that motion signals are integrated prior to perceptual threshold: that is, even if the two component motion signals are below threshold, their combination can be readily detected. We further find that there is a hint of an asymmetry in this integration, related to how Fourier motion signals interact with light and dark glider contraction signals. Interestingly, we report that sensitivity to motion signals can be context-dependent – that is, subjects demonstrated reduced sensitivity to Fourier motion in an environment which included glider expansion signals, compared to an environment containing glider contraction signals. Preliminary results from this work were presented at SFN 2014.

Finally, I close with discussion and a summary chapter (5), which includes several lines of work for future research.

CHAPTER 2
THE STATISTICS OF LOCAL MOTION SIGNALS IN NATURALISTIC
MOVIES

2.1 Abstract

Extraction of motion from visual input plays an important role in many visual tasks, such as separation of figure from ground and navigation through space. Several kinds of local motion signals have been distinguished based on mathematical and computational considerations (for example, motion based on spatiotemporal correlation of luminance, and motion based on spatiotemporal correlation of flicker), but little is known about the prevalence of these different kinds of signals in the real world.

To address this question, we first note that different kinds of local motion signals (e.g., Fourier, non-Fourier, and glider) are characterized by second- and higher-order correlations in slanted spatiotemporal regions. The prevalence of local motion signals in natural scenes can thus be estimated by measuring the extent to which each of these correlations are present in space-time patches and whether they are coherent across spatiotemporal scales. We apply this technique to several popular movies.

The results show that all three kinds of local motion signals are present in natural movies. While the balance of the different kinds of motion signals varies from segment to segment during the course of each movie, the overall pattern of prevalence of the different kinds of motion and their subtypes, and the correlations between them, is strikingly similar across movies (but is absent

from white noise movies). In sum, naturalistic movies contain a diversity of local motion signals that occur with a consistent prevalence and pattern of covariation, indicating a substantial regularity of their high-order spatiotemporal image statistics.

2.2 Introduction

Extraction of motion from visual input is crucial to making use of the visual input for a variety of purposes, including separation of figure from ground [15], navigation through space [16], and collision avoidance. Neural processing of visual motion is usually considered to consist of two stages: first, the extraction of local motion signals, and second, a stage in which these local signals are combined.

Local motion signals are typically classified according to their mathematical properties [1, 2, 17]. This has led to an important insight: there are two kinds of cues with distinct mathematical properties (Fourier and non-Fourier, see below) that can lead to the perception of visual motion. But it is unclear how these mathematical distinctions relate to the kinds of motion signals that are present outside of the laboratory. This is the question we address here: specifically, do naturalistic spatiotemporal stimuli contain different kinds of local motion signals? If so, how do they covary? These questions have important functional implications. For example, if different kinds of local motion signals are strongly correlated, extraction of only one kind of motion signal could suffice from a functional point of view, and sensitivity to the other kinds of local motion might just be a byproduct of neural computations, useful to investigators for uncov-

ering their nature. Alternatively, if the complement of motion signals depends on context (e.g., object motion vs. self-motion), then there might be selective pressure for separate extraction of multiple kinds of local motion signals. This would enable higher-level modules for action, object recognition, etc. to be separately linked to the appropriate kinds of local motion signals.

Approaching this question requires quantifying and characterizing the different kinds of motion signals that are present in natural contexts, and doing this in a way that they can be compared on an equal footing. This is not as straightforward as it might at first seem, because motion types have been defined in very different manners. Specifically, Fourier motion (F) is typically defined by the presence of a pairwise spatiotemporal correlation [1, 14] of luminance. (The reason that the term "Fourier motion" is used is that the set of pairwise correlations – the autocorrelation function – is the Fourier transform of the power spectrum, as is well-known [18].) In contrast, other kinds of motion signals have been defined on the basis of perceptual phenomena that occur in the absence of such correlations. The best-known examples of this are often called non-Fourier (NF) motion [2, 19], in which there is pairwise spatiotemporal correlation of a feature (e.g. a spatial edge or a temporal flicker edge). Moreover, motion perception can also occur in the absence of pairwise correlations of luminance (F motion) or of local features (NF motion), a phenomenon known as glider (G) motion [3, 13]. However, the extent to which these mathematically distinct signals are present in naturalistic inputs is unknown. To address this question, a necessary first step is to formalize the notions of F, NF, and G motion signals (and their subtypes) in terms of specific mathematical transformations, so that they can be compared on equal footing.

Here, we develop such measures and apply them to naturalistic movies (several popular films). The data show that all kinds of motion signals that we analyzed co-exist in moving visual images. The proportions of motions are relatively constant across movies, and there are consistent correlations between the different kinds of motion signals. However, within individual movie segments, one or another kind of motion signal may predominate, indicating that these correlations are only partial: that is, the different motion types provide non-redundant information.

2.3 Materials and methods

2.3.1 Data

The movie database was assembled by J.E. Cutting (www.cinematics.lv) and included hand annotations of the boundaries between continuous camera segments (i.e. "shots"). We used these boundaries (with fades and similar transitions excluded, and a five-frame additional margin) to define the analysis segments. All movies had similar characteristics: 24 frames per second, with each frame provided at a resolution of 256 by 256 pixels to respect copyright concerns. Most analyses made use of this resolution, though some (as indicated below) were carried out after further downsampling these pixels in blocks by averaging. We use the term "check" to represent the analysis unit, i.e., either a single pixel or a block of pixels that have been averaged. Since the original films had a "landscape" aspect ratio, each pixel in the database represented a rectangular region of the original film, larger in the horizontal direction than

the vertical. Specific movies that were selected were (1) "The 39 Steps" (1935) (2) "A Night at the Opera" (1935) (3) "Anna Karenina" (1935) and (4) "Mr. & Mrs. Smith" (2005). In designating check position, we use "matrix" convention, in which the X-coordinate increases from top to bottom, and the Y-coordinate increases from left to right. The analyses in the main text concern the YT plane (i.e., horizontal motion); parallel analyses in the XT plane (vertical motion) are in Supplement 2.6.1.

2.3.2 Quantification of motion signals

Our goal is to quantify different kinds of local motion signals (Fourier (F), standard non-Fourier (NF), and glider (G)) in a segment of a naturalistic movie. We do this by first measuring each kind of motion signal based on the luminance correlations within the appropriate spatiotemporal "template" (Figure 2.1) to obtain "local motion scores", and then, for each kind of motion, we combine these scores across space in different ways. We begin by motivating the definition of each kind of motion signal. Typically, F motion is defined by pairwise spatiotemporal correlation of the luminance values in the image [20]. NF motion denotes the motion of a local feature, such as an edge or flicker, in the absence of pairwise spatiotemporal correlation of luminance. An example of NF motion is an object that is flickering randomly - thus eliminating pairwise correlations - while moving across a background of equal mean luminance [2]. However, although several models for NF motion extraction have been proposed [2, 19], there is no single mathematical quantity (analogous to spatiotemporal correlation used for Fourier motion) that is recognized as defining its strength. Finally, G motion [3] encompasses third- or higher-order correlation in slanted

spatiotemporal regions, and occurs in the absence of pairwise spatiotemporal correlation of luminance (F motion) or simple features (NF motion).

These motion types have a fundamental similarity: they all depend on correlations within a slanted spatiotemporal region (Figure 2.1). For F motion, the correlation is pairwise, and the region consists of two checks, offset in space and time. For NF motion, the region consists of four checks, and the shape of the region depends on the subtype of NF motion: for NF motion of a spatial feature (NF-S), the region is a parallelogram consisting of two pairs of checks, each in adjacent two time-slices. Each pair of checks effectively detects the spatial feature (match vs. mismatch), and the combination of the two pairs detects whether this feature moves. For NF motion of a temporal feature (NF-T), the same region is rotated, to interchange the roles of space and time: each pair of checks detects whether there is local flicker, and the combination of the two pairs detects whether the feature moves. For the G motion types considered here, the region is a triplet of checks. Depending on the orientation of the triangle formed by the three checks, the region corresponds either to expansion or contraction over time. Thus in all cases, the local motion signal corresponds to the correlations among a group of checks in a specific shape, i.e., the "template" (Figure 2.1). The templates shown in Figure 2.1A correspond to motion to the right; flipping them across the Y axis corresponds to motion to the left.

To quantify the correlations within these templates, we calculate the product of the luminance values in their checks (after subtracting the mean luminance of each shot separately). To implement this for color movies, we first convert the color inputs to gray levels by using Matlab's `rgb2gray` function (The numeric range of luminance is irrelevant because we later normalize our calcu-

lations by a parallel computation for a movie with spatial correlations removed (see next section for details).

Following Reichardt [1] and many others, we note that the raw correlation value (i.e., the product of the luminance contrasts) will contain spurious motion signals when a static spatial edge is present. As is standard for F motion, we remove this spurious signal by an opponent process, in which correlations from left-facing and right-facing templates are subtracted (Figure 2.1). This strategy suffices for NF motion as well, but is insufficient for G motion (Figure 2.1): to eliminate this signal for G motion, we add a second opponent stage, in which signals from forward- and backward-facing templates are subtracted. Fundamentally, this second opponent stage is needed because the glider for G motion lacks the symmetry of the gliders for F and NF motion: for F and NF templates, left-vs-right spatial opponency is equivalent to forward-vs-backward temporal opponency. In other words, because of this symmetry for F and NF templates, the "standard" single-opponent calculation (space only) is equivalent to a double-opponent calculation (space and time), but for G templates, these two opponencies must be explicit (Note also that had we only included the forward-vs-backward opponency for G templates, then we also would not have eliminated spurious motion signals due to full-field flicker).

Formally, the calculation of the local motion score is as follows. A glider corresponds to a "template", B , which is a set of spatiotemporal voxels in a specific relative position. We represent a template as a set of triplets $[(x_1, y_1, t_1), (x_2, y_2, t_2), \dots, (x_n, y_n, t_n)]$, in which each of the x_i , y_i , and t_i are integers, and n is the number of elements in the template. Since the template is determined by the relative positions of its voxels, we require that $\min(x_i) = \min(y_i) = \min(t_i) = 0$,

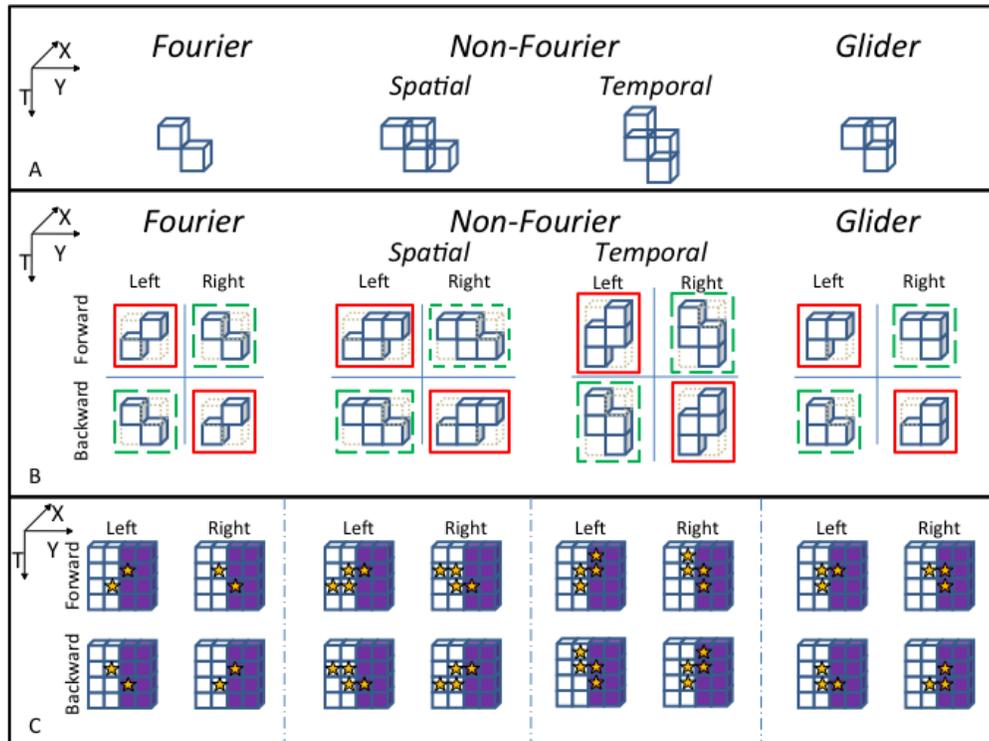


Figure 2.1: A summary of calculation of the basic local motion scores

Panel A: The templates used to quantify each kind of local motion. Panel B: Details of the various motion score calculations. First, Weber contrast values in the solid-bordered checks are multiplied together, and then, these products are summed in an opponent fashion (scores from red-outlined configurations are added, scores from green-outlined configurations are subtracted) to generate a local score motion signal (see text for details). Panel C: How opponency removes spurious signals due to static luminance edges. Each sub-panel diagrams the result of a computation of the local motion score when the template (stars) is positioned near a luminance edge. The four components of each sub-panel correspond to the four components of the sub-panels in B. For F and NF templates, left-oriented and right-oriented placements of the template each include the same number of dark and light checks. Thus, the left and right components of the calculation result in cancellation by their opponency. In contrast, for the G template, the left-oriented placement of the template contains one dark check, while the right-oriented placement contains two dark checks. Thus, the left vs. right opponency does not result in cancellation. However, forward and backward placements of the template are matched in terms of the luminances of the checks that they contain, and therefore the forward-vs.-backward opponency properly cancels the spurious motion signal. Note that for the F and NF templates, this second explicit stage of opponency has no effect. This is because of their symmetry: a left-to-right flip of the template is the same as a forward-to-backward flip (Panel B.)

where $i = 1, \dots, n$.

A template that is reversed along the X -dimension, which we denote as B^X , is the template in which each triplet (x_i, y_i, t_i) of B is replaced by $[L_X(B) - x_i, y_i, t_i]$, where $L_X(B)$ is the length of the template in the X -dimension, namely, $\max(x_i)$. Reversals along the Y and T dimensions are similarly defined. B^{YT} , for example, denotes a template that has been reversed along the Y dimension and then along the T dimension.

The raw correlation value for the glider B at the position (x, y, t) is defined as a product that involves all offsets contained in the glider:

$$RawCorr(x, y, t; B) = \prod_{(x_i, y_i, t_i) \in B} [I(x + x_i, y + y_i, t + t_i) - \bar{I}_{shot}], \quad (2.1)$$

where $I(x, y, t)$ is the luminance of the image at the position (x, y, t) and \bar{I}_{shot} is the median luminance across the shot. Finally, the local motion score at position (x, y, t) for motion type B in direction Z is defined by the double-opponent calculation:

$$LocalMotion(x, y, t; B; Z) = [RawCorr(x, y, t; B) - RawCorr(x, y, t; B^Z)] - [RawCorr(x, y, t; B^T) - RawCorr(x, y, t; B^{ZT})]. \quad (2.2)$$

Note that although our approach aims to capture specific types and kinds of local motion signals (F, NF-S, NF-T, and G), it can be easily modified to capture motion signals carried by correlations in other spatiotemporal configurations (e.g., [3]) by using the appropriate templates.

2.3.3 Combining local motion signals within a shot

Once the local motion scores were calculated as described above, the next step was to quantify motion signals within a movie "shot" (i.e., a sequence of frames that correspond to an individual scene). We used two kinds of strategies: a first kind that simply aggregates the local motion signals, and a second kind that is sensitive to whether these local motion signals are spatially coherent. Each of these strategies was applied separately to the three kinds of templates (F, NF, and G).

In the first kind of strategy, we simply computed the sum of the squares of the local motion signals for all placements of a particular kind of template within the shot. We normalized this quantity by dividing it by the results of a parallel computation applied to the same shots but in which the checks within each frame were scrambled. For the computation of the normalizing quantity, local correlations were determined by subtracting the global mean rather than the shot mean to avoid normalizations requiring division by quantities near zero. This allows for a meaningful comparison of the different motion types, independent of the size and shape of their templates. We call these quantities (computed separately for F, NF-S, NF-T, and G motions) the "simple motion" (SM) scores.

In formal terms, to derive the SM score from local motion scores, we proceeded as follows. First, for normalization purposes, we defined the local motion of a random movie:

$$\begin{aligned}
LocalMotionRand(x, y, t; B; Z) = & \\
& [RawCorrRand(x, y, t; B) - \\
& RawCorrRand(x, y, t; B^Z)] - \\
& [RawCorrRand(x, y, t; B^T) - \\
& RawCorrRand(x, y, t; B^{ZT})]. \quad (2.3)
\end{aligned}$$

where

$$\begin{aligned}
RawCorrRand(x, y, t; B) = & \\
& \prod_{(x_i, y_i, t_i) \in B} [I_{rand}(x + x_i, y + y_i, t + t_i) - \bar{I}_{movie}], \quad (2.4)
\end{aligned}$$

Here, I_{rand} is a movie in which checks are randomly permuted within a shot, and \bar{I}_{movie} is the median luminance across the movie. The SM score for a shot, for motion type B in direction Z , is the local motion score, averaged over the shot, normalized by the corresponding quantity for a random movie:

$$SM(B; Z; shot) = \frac{\langle [LocalMotion(x, y, t; B; Z)]^2 \rangle_{(x,y,t) \in shot}}{\langle [LocalMotionRand(x, y, t; B; Z)]^2 \rangle_{(x,y,t) \in shot}}, \quad (2.5)$$

The above average is taken over all positions (x, y, t) of the template within the shot.

We note that there is an important caveat that arises when this approach is applied to synthetic stimuli (in contrast to digitized naturalistic movies, considered here). Specifically, the grid used for motion analysis and the grid used for stimulus synthesis are separate grids, and must be considered as such. That is, when the motion scores are computed, the template must be placed in generic positions on the stimulus and not just in register with the grid used for stimulus generation. This detail is critical. Without it, the present approach might

fail to detect the motion signal in some of the drift-balanced stimuli of Chubb & Sperling [2], but with it, the approach captures the motion in all of them. This is illustrated and further discussed in Supplement 2.3.3 (Figure 2.23).

The second kind of strategy, which is designed to be sensitive to whether the local motion signals are spatially coherent, generalizes the use of a Reichardt model output to quantify the strength of standard F motion signals. In these strategies, the luminances in the "region of interest" (ROI) of 16 checks (either $1 \times 4 \times 4$ or $4 \times 1 \times 4$; $[X, Y, T]$) are considered together. Each ROI is then scored to indicate to what extent there was a coherent F, NF, or G motion signal throughout the patch. To simplify the process of defining and computing these scores, we first binarized the luminance values in each check—we replaced each luminance by +1 (black) or -1 (white), depending on how it compared with the median luminance within the shot. (Parallel analyses in Supplement 2.6.1 show that the results were robust with respect to the threshold used for binarization [Figures 2.17, 2.18, and 2.21] and that similar results were found for analysis in the XT plane [Figures 2.15 and 2.16]. Results in the main text are for the YT plane.) Note that this binarization can be considered as a form of dimension reduction. Prior to binarization, there are 256^{16} possibilities for the ways that a 16-check ROI can be colored; after binarization, there are only 2^{16} such combinations. Thus, binarization dramatically simplifies the process of defining, and then computing, a mapping from all of the possible ROI to a motion score; this is our motivation for it.

Formally, binarization corresponds to replacing each intensity $I(x, y, t)$ by $I^{binarized}(x, y, t)$, where $I^{binarized}(x, y, t)$ is +1 or -1, according to whether $I(x, y, t)$ is above or below a threshold (here, the median luminance within the shot). All

of the above quantities can then be calculated from the binarized movie. We denote such quantities by $RawCorr^{binarized}(x, y, t; B)$, $RawCorrRand^{binarized}(x, y, t; B)$, etc. Once binarization replaces each luminance value with +1 or -1, the product of luminance values within a template reduces to determining whether there is an even or an odd number of checks of each color. All of the colorings that yield a product of +1 contribute positively to a rightward motion signal, and all of the colorings that yield a product of -1 contribute negatively. Thus, the configurations that contribute positively to the motion score can be enumerated in a library. This is shown in Figure 2.1A, using the four-check NF-S template as an example. Since all of the colorings in the library yield a product of 1, they have an even number of white and black checks distributed among its four positions (two checks at one time step and two checks at the next). Thus, if a coloring has a spatial edge at one time step (one black and one white check), it must have a spatial edge at the next; if it lacks a spatial edge at one time step (two blacks or two whites), it must lack a spatial edge at the next. These relationships capture the notion that NF-S corresponds to spatiotemporal correlation of the presence or absence of an edge.

Based on these libraries, each ROI can be analyzed in terms of the configurations it contains to yield a score that quantifies the amount of each kind of coherent motion. We used two complementary approaches (but as we show below, the conclusions are largely similar).

The first approach (Figure 2.2B, top left) considers all of the placements of the template within the ROI and tallies the number that contributes positively to the motion signal. This yields a set of "rule match" (RM) scoresone for each orientation of the template (right forward, left forward, right backward, left

backward). These components are then compared to form a final “rule match opponent” (RMO) score for the ROI.

The second approach, “pattern match” (PM), treats the ROI in a more holistic fashion. To compute the PM measure (Figure 2.2B, bottom left), we determine the minimum number of checks that must be changed so that the ROI is made up entirely of template colorings within the library (i.e., that all have the relevant motion signal). As is the case for the RM approach, the four separate scores for each orientation of the template are then combined to yield a final “pattern match opponent” (PMO) score for the ROI.

The formal definition of the RMO and PMO scores are as follows. These scores are defined for any template B and any slab-like ROI that can contain the template along either the X or Y dimension, and has its other spatial dimension equal to one. Since the RMO score is an opponent score, we first define its components: the RM score $RM(B; Z; ROI)$. This is given by the total number of displacements (x_i, y_i, t_i) of the template within the ROI for which $RawCorr^{binarized}(x + x_i, y + y_i, t + t_i; B) = 1$, and thus is effectively a sum of $RawCorr^{binarized}$ scores within the ROI. The RMO score, $RMO(B; Z; ROI)$, is then

$$\begin{aligned}
 RMO(B; Z; ROI) = & \\
 & [RM(B; Z; ROI) - RM(B^Z; Z; ROI)] - \\
 & [RM(B^T; Z; ROI) - RM(B^{ZT}; Z; ROI)]. \quad (2.6)
 \end{aligned}$$

To define the PM score (PMO), we use the Hamming distance, a standard way of comparing two sets of binary numbers. Specifically, the Hamming distance between an ROI and another region K of the same size is given by

$$d(ROI, K) = \sum_{x, y, t \in ROI} |ROI(x, y, t) - K(x, y, t)|. \quad (2.7)$$

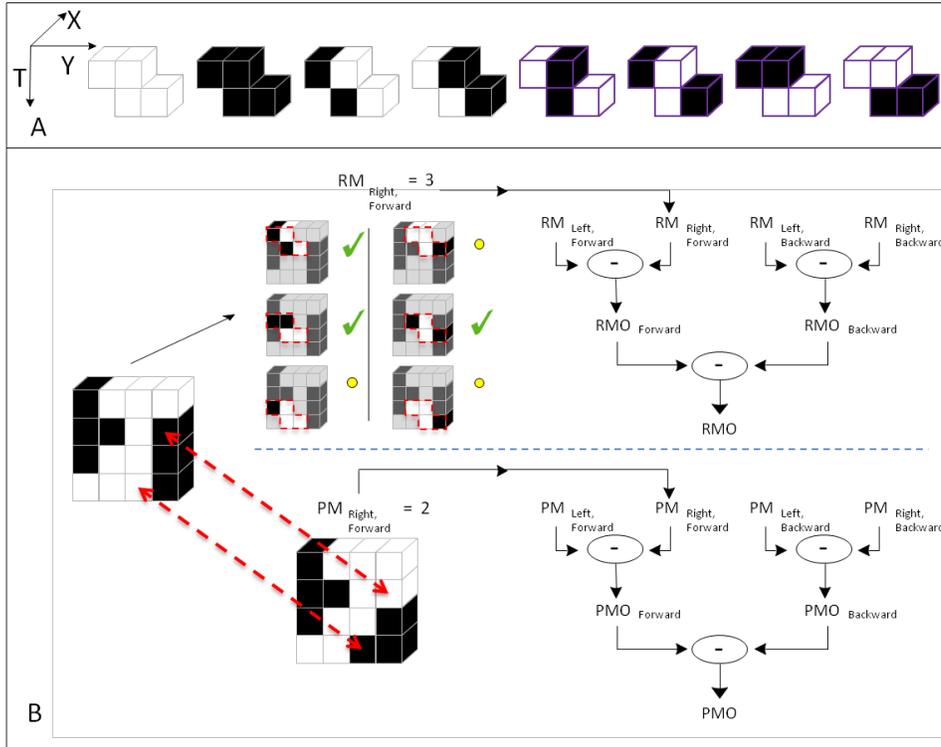


Figure 2.2: A summary of calculation of the RMO and PMO local scores in a spatiotemporal ROI, using NF-S motion as an example

Further details are in the text. (A) The library of eight template colorings consistent with NF-S motion. Note that all colorings have an even number of black checks. (The four rightmost colorings, marked in purple, are the templates used for pure NF-S, as the two-check F templates that they contain are inconsistent with F motion.) (B) Calculation of RMO and PMO scores for a 1x4x4 spatiotemporal ROI. For the RMO method (top), we consider all placements of the template within the ROI. There are six such placements (red dashes), and we tally the placements that yield colorings contained in the library of panel A, as these are the placements in which the black and white checks are consistent with NF-S motion. Checkmarks indicate the placements that result in colorings that are within the library; circles indicate the placements that result in colorings that are not in the library. Tallying the number of placements in the library yields a unidirectional RM motion score (in this case, right forward). Analogous scores are calculated by reversing the NF-S template in space (left forward) and time (right backward, left backward). These four unidirectional RM scores are combined in an opponent calculation to yield the RMO score for the ROI. For the PMO method (bottom) the entire ROI is treated as a whole. We determine the fewest number of checks that must be changed so that every placement of the template within the ROI yields a coloring that is in the library of panel A. In this case, changing two checks suffices: When these two checks are flipped in contrast (dashed arrows), all glider placements are in the library, and the resulting ROI is entirely consistent with NF-S motion. The tally of these changes yields the right, forward unidirectional PM signal. These four unidirectional PM signals are combined by an opponent computation to yield the PMO score for the ROI.

The PM score $PM(B; Z; ROI)$ is the minimum Hamming distance from the ROI to any region K for which every placement of the template in K yields a local motion score of +1. This minimum takes into account all possible colorings of K ; this is one reason why the dimensionality reduction is important. Finally, the PMO score is given by

$$PMO(B; Z; ROI) = [PM(B; Z; ROI) - PM(B^Z; Z; ROI)] - [PM(B^T; Z; ROI) - PM(B^{ZT}; Z; ROI)]. \quad (2.8)$$

Analogous RMO and PMO calculations were carried out for the other motion types (F, NF-T, and G) based on libraries that consisted of all colorings of the corresponding templates (Figure 2.1) that yielded a product of +1. For G motion, the procedure is asymmetric with respect to bright and dark: The resulting library includes three black checks but not three white ones, so it captures expansion and contraction of dark regions but not of light ones. We therefore designated this library G-K (black glider) and, in parallel, carried out computations based on a library G-W (white glider) containing the complementary colorings. Further, we also calculated RMO and PMO scores for the "pure NF-S" and "pure NF-T" signals. These were based on libraries that consisted of the NF-S or NF-T library, from which the libraries that contained F motion signals were removed (purple-outlined templates on the right in Figure 2.2).

Once RMO and PMO scores were calculated within each ROI, they were pooled within each shot by summing their squares. As is the case for the SM scores, we normalized this quantity by dividing it by the results of a parallel computation applied to random movie segments.

Note that for F motion, the RMO score is exactly the output of a special

case of a Reichardt detector (one with spatial inputs that are point-like and closely spaced, and has a pure delay of one frame prior to the multiplication step) operating on a binary image. For the PMO score, the correspondence is close but not exact (see Supplement 2.6.1, Figure 2.22 for further details). These correspondences were expected because the RMO and PMO scores were intended to generalize the Reichardt model in a manner that would be sensitive to local coherence.

2.4 Results

Our results concern the prevalence of different kinds of local motion signals in naturalistic scenes and how they covary. As described above, we considered three basic kinds of motion: F, NF, and G. F motion is equivalent to pairwise spatiotemporal correlation of luminance. NF motion [2, 19] is spatiotemporal correlation of a feature, and we identified two subtypes: spatiotemporal correlation of a spatial feature (NF-S) or a temporal feature (NF-T). The G motion considered here is characterized by expansion or contraction of either white or black patches, which can occur in the absence of pairwise spatiotemporal correlation of luminance, flicker, or edge [3]. We begin with an analysis of the prevalence of each kind of motion in four popular movies, using a simple measure of local motion signals (SM) and two measures that are sensitive to whether these signals are locally coherent (RMO and PMO; see Materials and methods for details on how these measures are defined). Then, using the various subtypes of motion signals that emerge from the RMO and PMO procedure, we consider how the different kinds of motion signals covary across movie segments (shots).

Results for the SM scores are shown in Figure 2.3. Overall, F motion strengths are the weakest, NF motion strengths are approximately 10 times stronger, and G motion strengths are intermediate between those two. There is a slight difference between spatial and temporal subtypes of NF motion: NF-S is slightly stronger than NF-T. Importantly, this pattern of relative strengths of the different kinds of motion signals is preserved across movie. The consistency of motion signals across movies holds across spatial scales. In Figure 2.3, the full available movie resolution was used (each analysis check consisted of a single movie pixel of the 256x256-pixel frame in the database); Figure 2.4 shows the results of an analysis in which each analysis check contains the average across a 16x16 block of pixels. With this coarse-grained analysis, F motion strength remains much weaker than NF and G strengths, and there is a modest change in the behavior of the NF and G signals. Specifically, while the tails for NF motions remain larger than the tails for G motion (as was the case at the fine scale), the median for G motion is now larger than the median for NF motion. This shift, as well as the overall pattern of motion strengths at each scale, is consistent across movies. An extended analysis of motion strengths at intermediate scales is shown in Supplement 2.6.1, Figures 2.11 and 2.11.

The above analyses quantify the strength of each kind of motion signal in each movie segment (shot), but do so via an SM measure that is insensitive to whether the motion signals are locally coherent. We next carry out parallel analyses with two measures (RMO and PMO; see Materials and methods) that are designed to be sensitive to coherence of motion signals within 16- check ROIs. A key step in the construction of these measures is binarization of the movie to reduce the dimensionality of the problem (see Materials and methods). Thus, as a preliminary step, we first examined the effects of binarization itself.

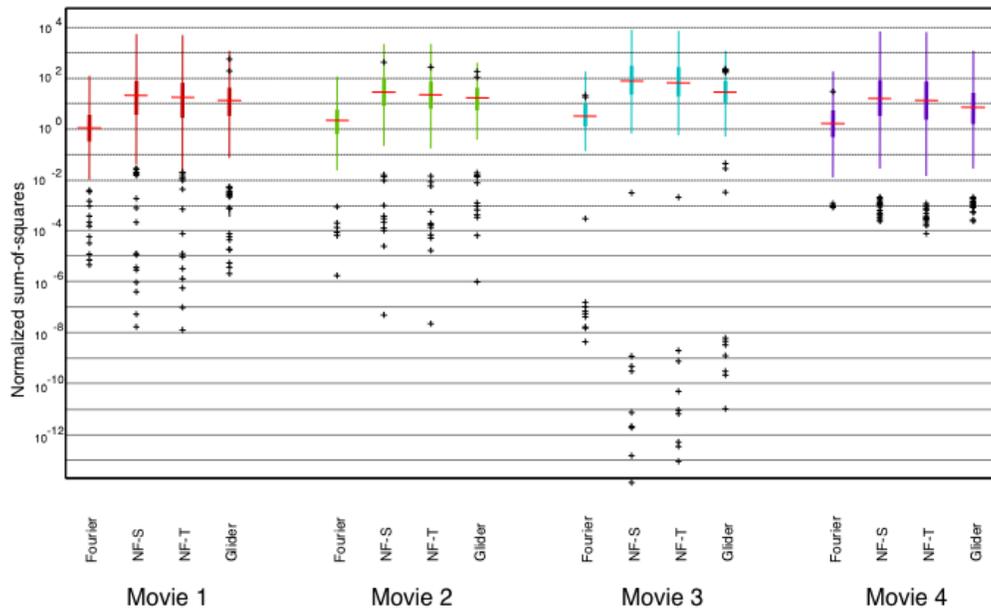


Figure 2.3: Prevalence of different kinds of motion signals is similar across movies.

For each movie, SM scores (see Materials and methods) were calculated for each movie segment, and the distribution is summarized by the median (horizontal line), the interquartile range (heavy vertical line), the “whiskers” (thin vertical line, covering four times the interquartile range), and the outliers (individual symbols, outside the range of the whiskers). Values are normalized by SM motion scores obtained from movies of random pixels of similar segment length. Each motion was calculated with respect to its relevant template shape (see 2.1A) in the YT plane (i.e., horizontal motion); each check corresponded to a single pixel in the discretization of the movie (256x256 pixels per frame, 24 frames per second). Movies were (1) The 39 Steps (1935), (2) Anna Karenina (1935), (3) A Night at the Opera (1935), and (4) Mr. and Mrs. Smith (2005).

Figure 2.5 directly compares measures of motion strength on a shot-by-shot basis calculated with and without binarization. For all motion types, binarization compresses the range of the motion scores from approximately a factor of 10^6 (without binarization) to 10^3 (with binarization to +1 and -1). Most of this compression is due to an increase in the lowest motion scores since binarization eliminates the possibility of multiplication by values near zero. But the upper ends of the distribution are also affected by binarization: Thresholding

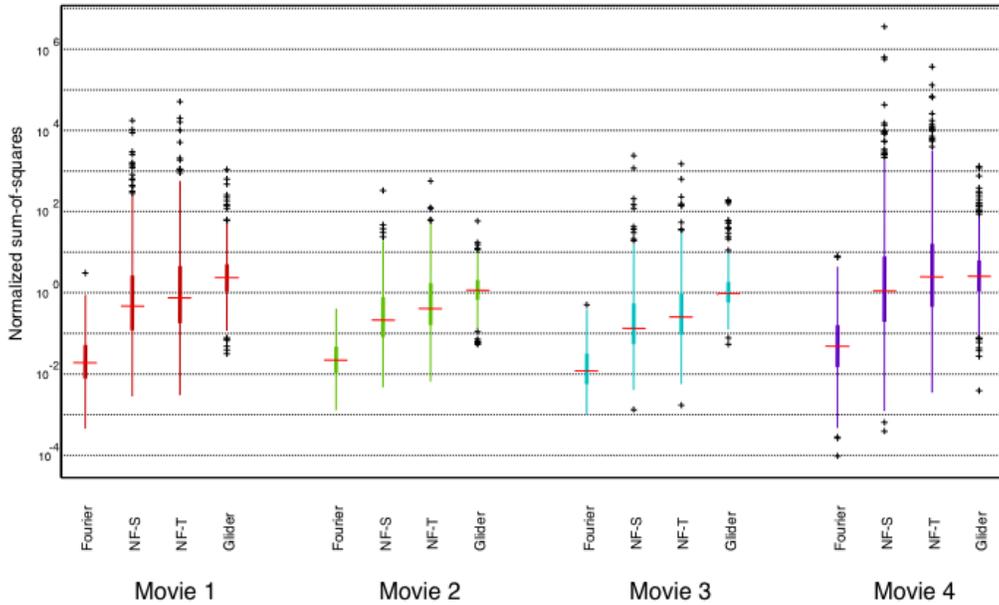


Figure 2.4: Prevalence of different kinds of motion signals, analyzed at a coarse spatial scale, is similar across movies.

For each movie, SM scores were calculated after downsampling each 16×16 block of pixels in the original movie to a single check. For other details see Figure 2.3.

substantially reduces highest values for NF-S and NF-T (Figure 2.5B and C) and slightly reduces the highest values for G (Figure 2.5D). The likely reason for this is that the NF scores reflect products of four values (since the templates have four checks) and the G scores reflect products of three values (since the templates have three checks). Hence, binarization results in a moderate reduction in the extreme high values that result from products of three luminance values (G) and a more severe reduction in the extreme high values that result from products of four values (NF). In line with the increasing range compression as the number of checks in the template increase, correlations of the log-scaled SM scores with and without binarization are largest for F motion (0.79), next-

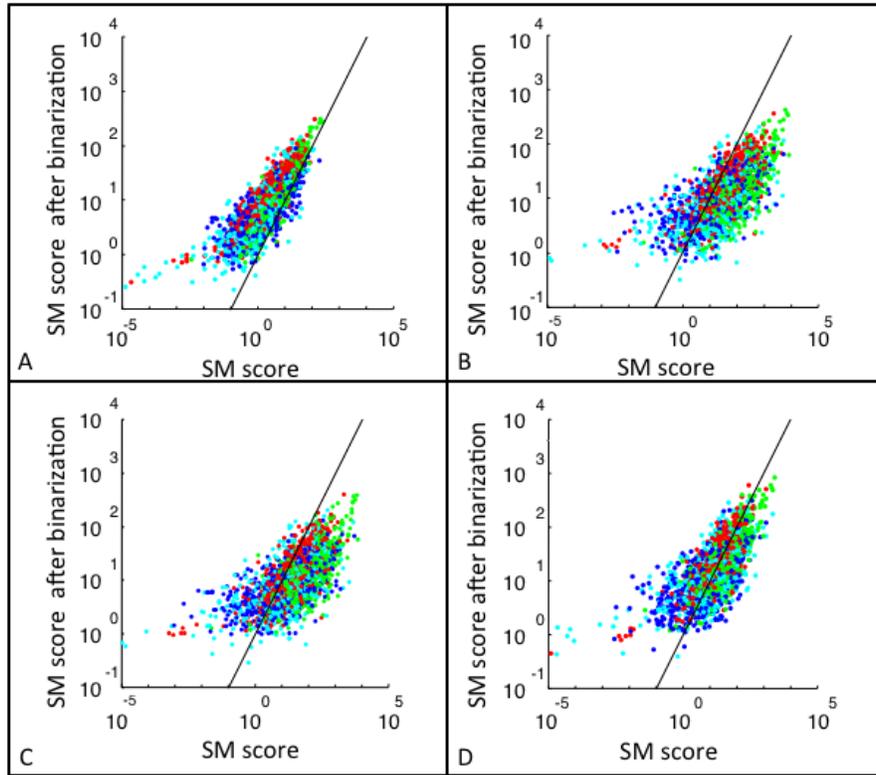


Figure 2.5: The effect of binarization on local motion scores

(A) E, (B) NF-S, (C) NF-T, and (D) G motions. SM scores were calculated for each movie segment based on raw luminance values (abscissa), and also following binarization with the threshold set at the overall shot median luminance (ordinate). No spatial downsampling was applied. A random sample of 500 shots from each movie is presented here. Movies were color coded as follows: (red) "The 39 Steps", (blue) "Anna Karenina", (green) "A Night at the Opera", and (cyan) "Mr. & Mrs. Smith" (2005). The black line is the line of identity.

largest for G motion (0.69), and smallest for NF-S and NF-T motion (0.62 and 0.63, respectively; $p < 0.001$ in all cases).

We note that when applied to binary movies, the local motion score for F motion coincides exactly with the output of a Reichardt detector because the SM score computes exactly the same product as the Reichardt detector, and the binarization step has no effect on a movie that has already been binarized. See Supplement 2.3.3, Figure 2.22 for further details and for the relationship of the RMO and SMO scores to the Reichardt detector for binary and gray-level

movies.

The effects of binarization on the shot-by-shot distributions of SM scores are shown in Figure 2.6 (analyzed at a fine spatial scale) and Figure 2.7 (analyzed after 16x16 downsampling). An extended analysis of motion strengths at intermediate scales is shown in Supplement 2.6.1, Figures 2.11 and 2.12. As expected from Figure 2.5, the distributions are more compact than the corresponding distributions shown in Figures 2.3 and 2.4. F motion remains the smallest signal and G motion is the largest, as the upper tail of the NF motion distributions is most severely affected by the binarization. As is the case for the analysis without binarization, the pattern of motion strengths is similar across movies at each spatial scale.

We now turn to the indices that examine the strength of coherent motion of each type. Briefly (see Materials and methods for further details), the indices were calculated as follows. First, correlations of the binarized movies were calculated within slanted spatiotemporal templates corresponding to each motion type (Figure 2.1). Since the movies are binarized, calculation of the correlations (the products of the luminances in each check) reduces to determining whether each template's coloring is present in a library (shown in Figure 2.2 for NF-S). Second (Figure 2.2B), correlations were combined within a spatiotemporal ROI. Two variants were used for this pooling process: one in which the local correlations were simply summed (RMO) and one in which they were treated holistically (PMO). Third, signals in opposite directions were compared to determine a net motion signal for each ROI. Finally, the sum of the squares of these local signals within each shot were normalized by the results of a similar calculation applied to random movie segments. Results for the PMO index are

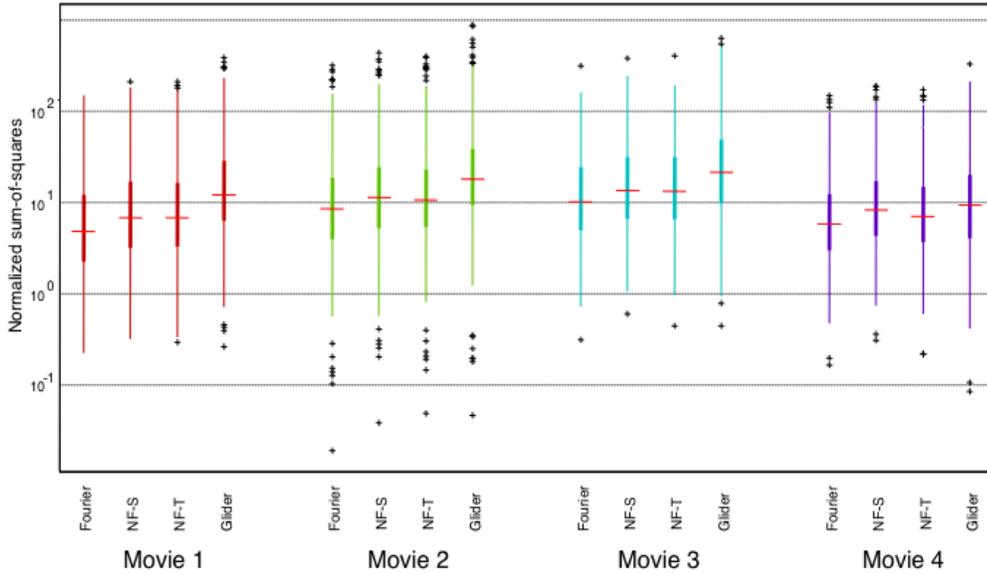


Figure 2.6: Prevalence of different kinds of motion signals is similar across binarized movies.

For each movie, data were first converted to -1 or +1 using a threshold equal to the median overall luminance value within each shot, and SM scores were then calculated. For other details, see Figure 2.3.

shown in Figure 2.8. Overall, F motion strengths were the largest, NF motion strengths were 50% to 70% as large, and the other kinds of motion (pure NF and G) were somewhat smaller. There was virtually no difference between spatial and temporal subtypes of NF motion and virtually no difference between white and black subtypes of G motion. As with all previous analyses, the pattern of relative strengths of the different kinds of motion signals was preserved across movies.

Parallel results for the RMO index are shown in Figure 2.9. Overall values of motion strength are smaller than for the PMO index (Figure 2.8), and the RMO index shows a larger difference between the NF motion strengths and the

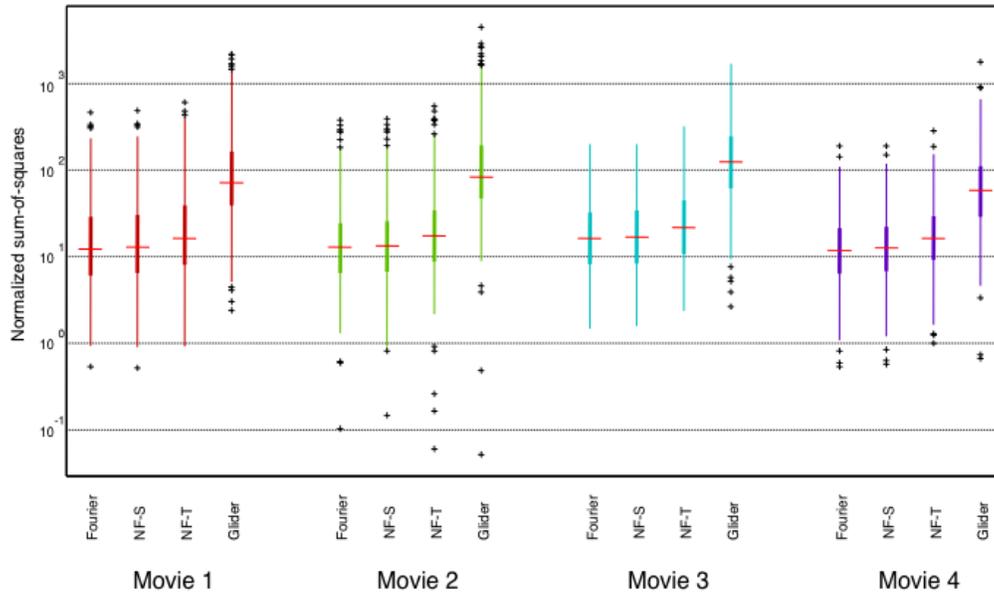


Figure 2.7: Prevalence of different kinds of motion signals, analyzed at a coarse spatial scale, is similar across binarized movies.

For each movie, SM scores were calculated after downsampling each 16x16 block of pixels in the original movie to a single check, and then binarization. For other details see Figure 2.4.

G motion strengths than the PMO index. However, the basic findings obtained with the two kinds of ROI indices are similar: F motion strengths are largest, followed by NF motion, and then by G motion, and the relative sizes of the motion signals are consistent across movies.

The analyses in Figures 2.8 and 2.9 were performed with ROI oriented parallel to the YT plane and for a single resolution (each check used in the analysis corresponded to one pixel in the database's digitization of the movie); the results hold for other resolutions (Supplement 2.6.1, Figures 2.13 and 2.14) and orientations (Supplement 2.6.1, Figures 2.15 and 2.16). In addition, Supplement 2.6.1, Figures 2.17 and 2.18 show parallel results for binarization at

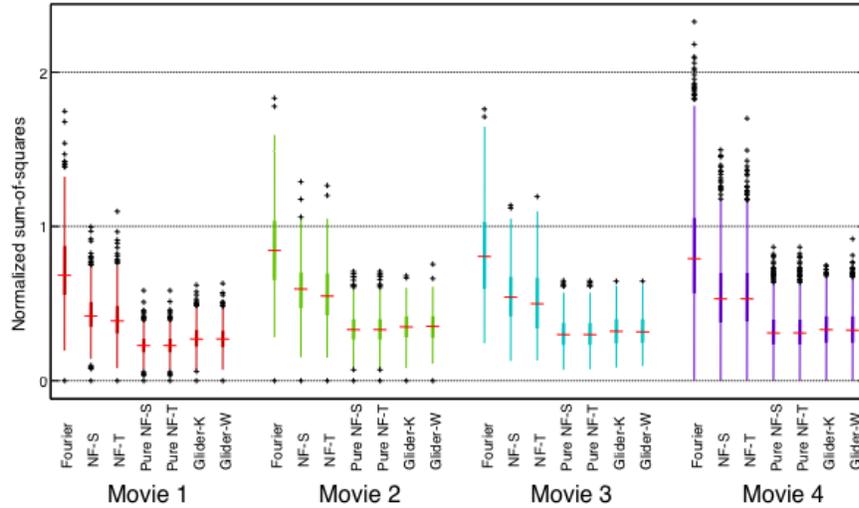


Figure 2.8: Prevalence of different kinds of motion signals is similar across movies, as measured by the PMO score.

The ROI consisted of a 4x4 block of checks in the YT plane (i.e., horizontal motion); each check corresponded to a single pixel in the discretization of the movie (256x256 pixels per frame, 24 frames per second). For other details see Figure 2.3.

the global midgray level rather than the median for each shot.

In sum, when motion signals are measured in a purely local manner (SM score), F signals are weaker than NF or G signals (Figures 2.3, 2.4, 2.6, and 2.7). But when spatial coherence is taken into account (via either the PMO index [Figure 2.8] or the RMO index [Figure 2.9]), F signals dominate. This shift, as well as the pattern of motion strengths captured by each index, is similar across analysis scales and movies.

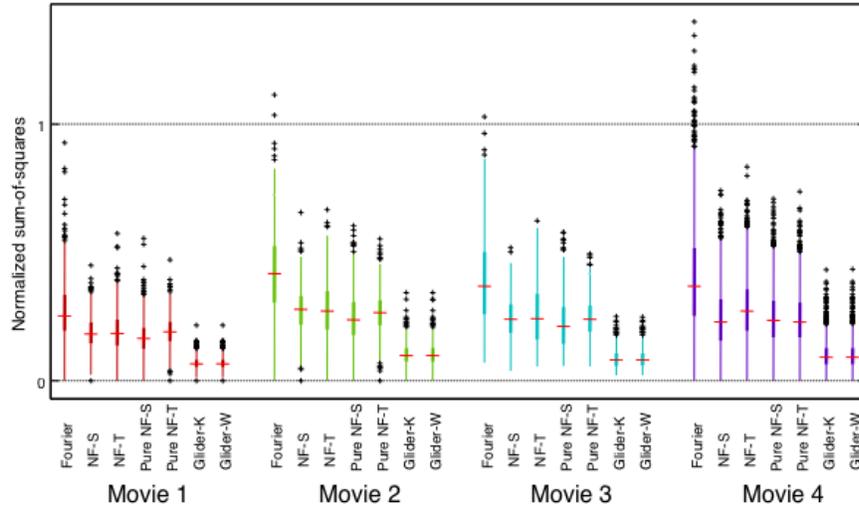


Figure 2.9: Prevalence of different kinds of motion signals is similar across movies, as measured by the RMO score.

The ROI consisted of a 44 block of checks in the YT (horizontal motion) plane. For other details see Figure 2.8.

2.4.1 Covariation of motion signals

The distributions shown in Figures 2.8 and 2.9 indicate a substantial variation in the amount of motion signals present in each shot. We now focus on this variability and examine how the different kinds of motion signals covary with one another. One possibility is that the different kinds of motion signals are tightly correlated that some segments have low levels of all motion signals and others have high levels, with the amount of one signal determining the amount of the others. Alternatively, the motion signals may be somewhat independent, present in ratios that depend on the characteristics of the individual shots. With this motivation in mind, we determined the pattern of covariation of the several

kinds of motion signals.

Results are shown in Figure 2.10 (panel A for PMO and panel B for RMO). While it is clear that there are strong correlations between motion signals of each type with every other type, it is also clear that they are not completely redundant (except for the specific pairs of motion scores that are guaranteed to be identical; see Supplement 2.6.1). That is, given the size of one kind of motion signal, the size of another kind can vary by a factor of two or more. The extent to which one motion signal determines the other depends on the specific pair of signals. For example, considering the PMO indices (Figure 2.10 A), an F signal is only a weak determinant of a pure NF-T signal (their ratios can vary by more than a factor of two), but the G and the NF-S signals are strongly correlated (their ratios vary by less than 40%). With the RMO index (Figure 2.10B), the patterns of covariation are in general similar, though there is somewhat less correlation overall between the motion indices. Supplement 2.6.1, Figure 2.19 shows a corresponding analysis for a second movie; the pattern of covariation for each kind of index (PMO and RMO) is very similar to that of Figure 2.10. This shared pattern of covariation across movies is a consequence of the statistical structure of the movies themselves, not of the way that the indices are calculated (e.g., that they are determined from overlapping sets of templates). This is shown in Supplement 2.6.1, Figure 2.20, when a similar analysis is applied to random movies. In this case, the different kinds of motion signals are largely uncorrelated. Supplement 2.6.1, Figure 2.17 shows parallel results for binarization at the global midgray level rather than the median for each shot for one movie.

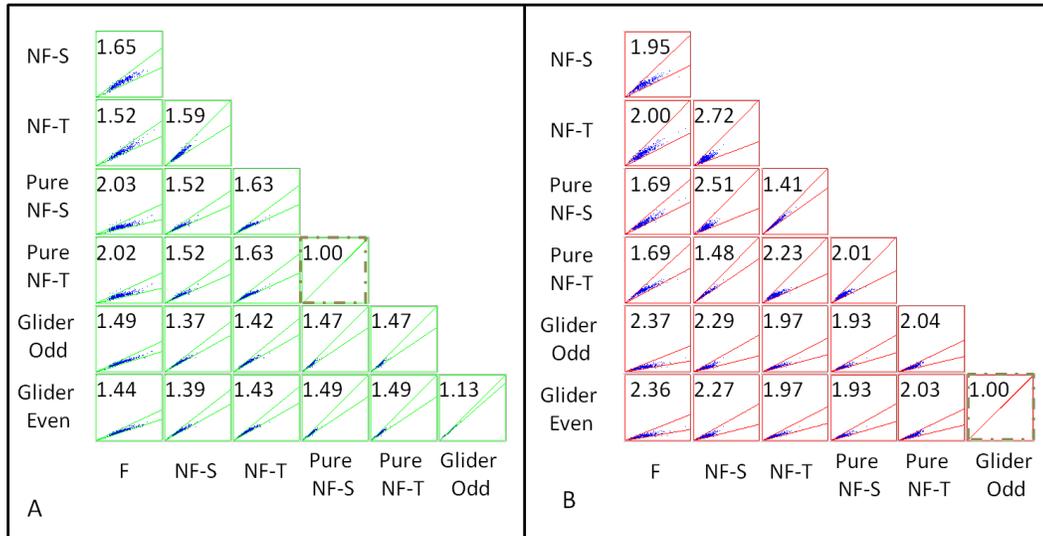


Figure 2.10: Covariance patterns of motion scores (*YT*; "The 39 steps").

(A) PMO and (B) RMO. Within each scattergram, each point represents a pair of normalized motion scores determined from a single movie segment ("shot"). Axes range from 0 to 2 (PMO) and 0 to 1 (RMO). The number in each plot indicates the average ratio between the pair of motion scores; the two sloping lines in each plot indicate the wedge that contains 95% of the values. Large values of one motion score typically occur with large values of the other scores, but the ratios between a pair of scores can vary by up to a factor of two. (Pure NF-S and NF-T PMO scores are identical, and the two G motion RMO scores are identical; see Supplement 2.6.1.) Analysis was carried out in the *YT* plane (horizontal motion) at the maximum resolution of the database for The 39 Steps.

2.5 Discussion

Identifying the presence of moving objects and determining their velocity begins with neural computations that analyze restricted patches of the visual input in order to extract local motion signals.

Based on their mathematical properties, several types of local motion signals have been recognized. The simplest is pairwise spatiotemporal correlation of the luminance pattern [1, 14]; this is known as Fourier (F) motion since the presence of pairwise spatiotemporal correlation can be identified from the Fourier amplitudes of the stimulus. Subsequently, it was recognized that pairwise spatiotemporal correlation is not necessary to produce a percept of visual

motion [2, 3, 19]. A percept of motion can be produced by spatiotemporal correlation of a local feature rather than of the luminance pattern itself a phenomenon typically called Non-Fourier (NF) motion [2] to emphasize that the correlations cannot be identified from the Fourier amplitudes. A percept of motion can also be produced by spatiotemporal correlations among three points, a phenomenon known as Glider (G) motion [3], even when spatiotemporal correlations of simple features are not present. Note that both NF motion and G motion each encompass multiple distinct subtypes of local motion signals—for NF motion, the subtype is determined by the choice of local feature (e.g., edge or flicker), and for G motion the subtype is determined by the geometry of the three spatiotemporal points that are correlated (although here we consider only one specific configuration: a right triangle whose legs are aligned with space and time axes).

Although each of these kinds of motion signals is mathematically distinct and separately available to perception, their occurrence in the natural environment is poorly characterized. We therefore developed several ways to quantify the strengths of different kinds of local motion signals so that they could be compared on an equal footing.

This step was necessary because of the way that the different kinds of motion signals are usually defined: F motion signals are defined in terms of a computational model that is applicable to any stimulus [1, 14], while NF motion and G motion are defined in terms of specific exemplars, along with the absence of an F signal. The key consideration that enabled a comparable measure that is applicable to the different kinds of motion signals is that each kind of motion corresponds to a correlation in a spatiotemporal region with a specific geometry (i.e., within a specific template; see Materials and methods). For standard

(F) motion, this template is a pair of checks on a space-time diagonal. For NF motion, the template is a set of four checks, forming a parallelogram in space-time. For G motion, the template is a set of three checks in a triangle. As expected, when applied to standard (F) motion, this method yields results that are consistent with computations based on standard spatiotemporal correlation (i.e., the Reichardt detector; [14]; [1, 21];[20]: For gray-level movies, results are strongly correlated, and for binary movies, they coincide (Supplement 2.6.1, Figure 2.22).

For NF motion, the development of a motion score is less straightforward, as different exemplars of NF motion can have qualitatively different natures. For example, NF motion stimuli can be constructed based on beats, contrast modulation, transparency, or occlusion. Fleet and Langley [19] observed that the motion signals in all of these stimuli have a common aspect that is manifest in the power spectrum: Power is concentrated in spatiotemporal planes that do not include the origin. Since calculation of the power spectrum requires inspection of a wide region of space, this observation does not directly translate into a measure of a local NF motion signal. However, it is closely linked to the rationale for our approach. As Fleet and Langley [19] observe, power in a spatiotemporal plane removed from the origin corresponds to pairwise spatiotemporal correlation of a feature; this pairwise correlation, in turn, can be detected by a local nonlinearity. This is exactly the approach taken here. Each of the two kinds of features considered-spatial and temporal edges-are identified based on whether the values within a pair of checks match or mismatch (a local nonlinearity). Then, multiplication (or the parity rule) within the four-check template computes the pairwise spatiotemporal correlation of these two-check features. However, the actual computations used by the visual system to extract

NF motion are unknown. Thus, despite the grounding of the approach in a common mathematical feature of NF motion [19], we used multiple variants of the basic correlation measure to ensure the robustness of our results: measures with and without a binarizing nonlinearity, and measures that used different kinds of spatial pooling (SM, RMO, PMO).

An advantage of this approach to measuring NF motion strength is that it extends to G motion simply by changing the shape of the template. Moreover, by changing the template and the rules for scoring its colorings, this approach can be extended to deal with further types of features, such as temporal correlation of orientation [19, 22, 23]. Since many studies have been devoted to understanding the correlation structure of natural scenes [24–26] –including their temporal aspects [22, 23, 27] – it may appear surprising that relatively little is known about the local motion signals that they contain. The basic reason is that the focus of most studies has been on the second-order statistics of natural scenes. While F motion signals can be determined from second-order statistics, G motion and NF motion require, respectively, knowledge of third- and fourth-order statistics. Thus, studies of the spatiotemporal power spectrum [22] cannot characterize NF motion and G motion completely, as the spectrum is a characterization only of pairwise correlations. On the other hand, it is difficult to carry out an exhaustive characterization of high-order statistics of natural scenes because of the dimensional explosion that results. Thus, in order to carry out an analysis that suffices to identify G motion and NF motion, we are necessarily selective about the high— order image statistics that are analyzed.

Once the motion-related high-order statistics were identified, we used two methods to pool them within each ROI (see Materials and methods): RMO and

PMO. The RMO score simply compares the number of template positions in which the local motion signal is present and the number of positions in which it is absent; it is thus linear and is as local as possible, as it adds no further spatial interactions. In contrast, the PMO score determines the closest match of the ROI to an exemplar ROI in which the local motion signal is present in every template position. Therefore, the PMO score is nonlinear—it is strongly sensitive to whether the motion is coherent throughout the ROI—and, consequently, is somewhat less local than the RMO score. Neither score is necessarily larger than the other: For specific ROI colorings, the RMO score may be higher than, equal to, or lower than the PMO score. Thus, the two scores provide different ways to measure the strength of each kind of motion signal. Nevertheless, as emphasized above, our basic conclusion holds in either case (Figures 2.8, 2.9, and 2.10) and across scales of analysis and binarization strategies (Supplement S1). That is, each kind of motion signal (F, two varieties of NF, and two varieties of G) is present in natural movies. Across movies they are present in approximately similar amounts, and at the level of individual movie segments there is substantial variation in the proportion of each kind of motion signal but a similar pattern of covariation.

The level of consistency across movies is perhaps surprising, given the finding of Cutting et al. [27, 28] that the general amount of visual change in movie shots increased significantly over the period that the movies span (1935-2000). Our approach, though, is different, as it is specifically sensitive to different kinds of local motion rather than overall amount of visual change. Further, the similarity of the strengths of local motion signals appears to hold over a range of shot lengths, which vary over a twofold range (8.96 s, 9.6 s, and 9.16 s in the 1935 films but 4.15 s in the 2005 film, also reflecting a trend in movie

making [27, 28]).

The segment-to-segment variation in the relative strength of the different kinds of motion signals (Figure 2.10) also deserves comment. As is evident from these scattergrams, different kinds of motion signals tend to occur in combination. This contrasts with most artificial stimuli used to study motion in the laboratory, as such stimuli are typically designed with the goal of isolating a single kind of cue. Moreover, the fact that there are correlations between different kinds of motion signals in the natural environment has implications for understanding the design of neural circuits that detect motion. If high-order motion cues coexist with F motion cues, then they can be exploited [13] to improve on the performance of a standard Reichardt detector [1]. As we show here, this coexistence is characteristic of motion signals in naturalistic movies (Figure 2.10 and Supplement 2.6.1, Figure 2.19), so these theoretical considerations [13] are relevant to natural vision.

Although there are correlations between different kinds of motion signals, they are not redundant. Specifically, given the level of one motion signal (e.g., F motion) in a movie segment, one can estimate the level of another (e.g., G motion), but the estimate holds only within a factor of two.

This diversity in the complement of motion cues that are present in a given movie segment may have implications for how motion is analyzed at later processing stages. In central visual processing, motion is used for many different purposes, such as navigation, collision avoidance, extraction of object structure [29, 30], and the analysis of biological motion [15, 29, 31–38], each of which is carried out in distinct networks of brain areas [30, 35, 39, 40].

Some aspects of the different kinds of low-level motion signals suggest that they may be selectively important in these different contexts, or for different purposes. One example of this potential for selectivity is that, as Fleet and Langley [19] point out, NF motion signals can arise from occlusion. When an untextured object moves across the visual field in front of a textured background, it progressively occludes and then reveals spatial features of the background, generating a spatial NF motion signal. A similar NF motion phenomenon occurs if the foreground object is semitransparent: The features are not eliminated, but their contrast is modulated. Another way in which a specific kind of motion signal may arise in a specific context is that G motion signals can arise from looming—that is, the motion of an object toward the observer. This is because the basic element of G motion is correlation in a three-point spatiotemporal configuration of checks [3]. For three points arranged in a right triangle with one side aligned with the temporal axis—the case considered here—this corresponds to an expanding or contracting region.

Since different kinds of local motion signals can arise in different contexts, it is reasonable to speculate that brain areas that make use of motion for different purposes (e.g., segmentation vs. navigation) receive inputs from local motion detectors with appropriately matched properties. This would enable the parallel high-level analyses of motion in central visual areas to focus on the kinds of low-level signals that are the most relevant to their separate functions.

2.6 Supplementary Materials

2.6.1 Motion scores: supplementary analyses, properties and a note on analysis of synthetic motion stimuli

Overview The supplementary materials contain three sections: the first is devoted to demonstrating robustness of the main results, the second provides additional information on the local motion scores, and the third focuses on a caveat relevant to applying the present method to movies that are generated in the laboratory, such as the classic examples of non-Fourier motion introduced by Chubb & Sperling [2].

The first section of the Supplement expands on the analyses carried out in the main text. Figures 2.11 and 2.12 (SM score), 2.13 (PMO scores) and 2.14 (RMO scores) show that the results in the main text are robust with respect to the size of the region used for discretization. Figures 2.15 and 2.16 show that the results in the main text, which analyzed motion in the YT plane, also hold for motion in the XT plane. Figures 2.17 and 2.18 show that the results are robust with respect to the binarization strategy: in the main text, the binarization cutpoint was the median for each movie segment; here we show that similar re-sults are obtained when the threshold is set at the mid-level luminance value of the entire range.

Figures 2.19- 2.21 show supplementary analyses related to covariation of motion signals. Specifically, Figure 2.19 shows that the pattern of covariation of motion signals, presented for Movie 1 in the main text, holds for other movies. Figure 2.20 presents a parallel analysis of a random spatiotemporal noise, show-

ing that the correlations among the motion signals in popular movies arise from the structure of the movies themselves, not the analysis procedure. Finally, Figure 2.21 shows that the pattern of covariation of motion signals is robust with respect to the binarization strategy when the threshold is set at the mid-level luminance value of the entire range.

The second section of the Supplement provides mathematical background on the motion scores. In Figure 2.22, we compare the RMO and PMO scores to the output of the Reichardt model for Fourier (F) motion, and show that for RMO, the correspondence is precise. We then demonstrate some mathematical relationships between the scores: (a) that for the subtypes of glider (G) motion, the RMO scores must be identical (but the PMO scores need not be), and (b) that for the two kinds of pure non-Fourier (NF) motion, the PMO scores must be identical (but the RMO scores need not be).

The third section of the Supplement describes a technical point that becomes important in applying our analysis strategy to movies that are generated in the laboratory (but is not relevant analyzing digitized naturalistic movies). In this context, we show that it is crucial to recognize that the grid used to generate the stimulus and the grid used to analyze it are separate entities. All glider placements - not only those that are in register with the stimulus grid, but also the generic placements that are out of register - are relevant to the motion computation. If only the special, in-register placements are considered, the motion signal could be overlooked.

Supplementary analyses of motion scores

Levels of discretization: SM score In most analyses in the main text, each check of the 1x4x4 (or 4x1x4) ROI used to calculate motion scores corresponded to a single pixel of the movie in the original database. Here, we repeat the analysis with using other resolutions, downsampling an $n \times n$ block of pixels

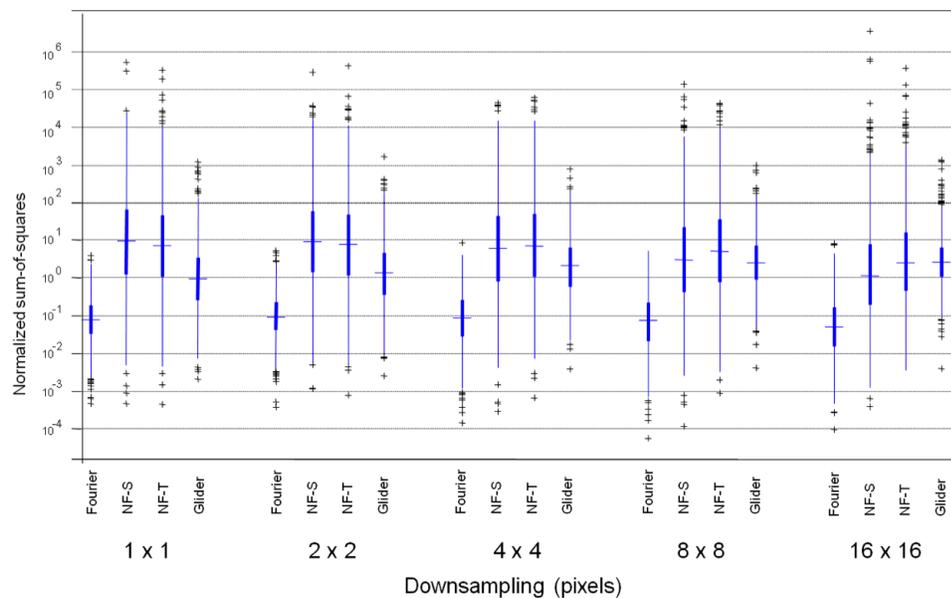


Figure 2.11: Dependence of motion signals on spatial discretization, for the different motion scores.

Motion scores were calculated for each scene, and the distribution is summarized by the median (horizontal line), the interquartile range (heavy vertical line), the "whiskers" (thin vertical line, covering 4x the interquartile range, up to the closest data point) and the outliers (individual symbols, outside the range of the whiskers). Values are normalized by scores obtained from a movie of random pixels. A single movie (movie 4 the main text, "Mr. & Mrs. Smith") was re-analyzed at the original resolution and at four levels of downsampling, in which a single check in the ROI corresponded to an $n \times n$ block of pixels in the original movie. The level of downsampling ranged from 1x1 (no downsampling, corresponding to Figure 2.3 of the main text) to 16x16 (corresponding to Figure 2.4 of the main text). Similar grayscale conversion function was used for all clips. The ROI was oriented in the YT plane. Original movies' inputs were presented at 256 x 256 pixels per frame and 24 frames per second.

($n=2, 4, 8,$ and 16) to a single check. Downsampling was carried out by averag-

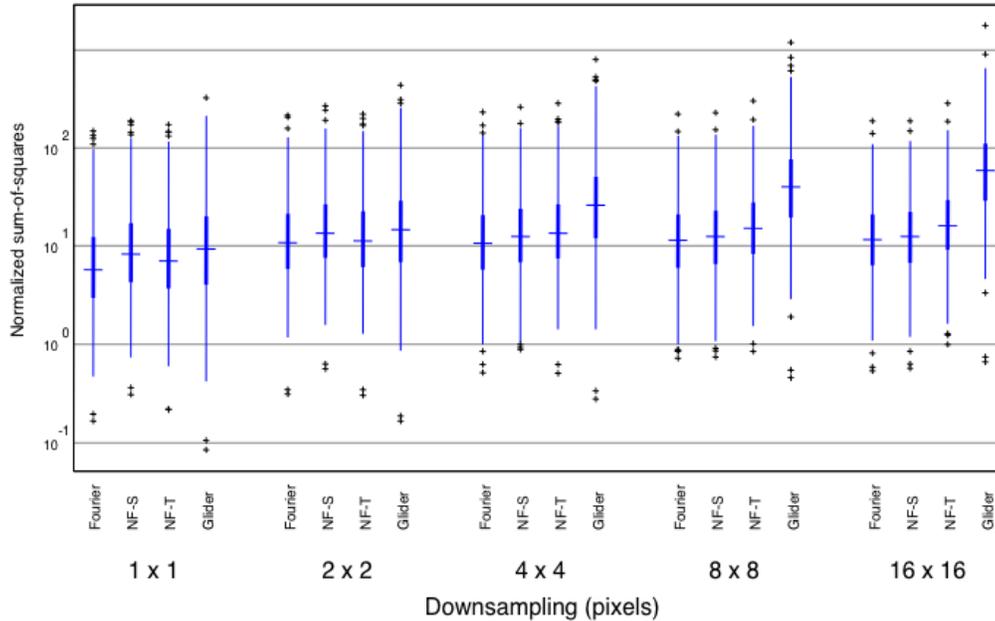


Figure 2.12: Dependence of motion signals on spatial discretization, for the different motion scores.

The level of downsampling ranged from 1x1 (no downsampling, corresponding to Figure 2.6 of the main text) to 16x16 (corresponding to Figure 2.7 of the main text). Following downsampling, luminance values were binarized at the median for each shot. Other details are as presented in Figure 2.11.

ing the original movie over each $n \times n$ block.

Figures 2.11 and 2.12 show the SM score for these intermediate levels of discretization. Figure 2.11 shows the analysis based on raw luminance values, while Figure 2.12 shows the SM score following binarization at the median. In all cases, F scores are the lowest. G scores increasingly dominate as the region of discretization enlarges.

Intermediate levels of discretization: PMO/RMO score Figures 2.13 (PMO) and 2.14 (RMO) examine effects of intermediate levels of downsampling. Over-

all, the magnitude of each motion score decreases with progressively greater downsampling (i.e., with larger checks used for the analysis), and the ratios between motion scores become more restricted. Nevertheless, the results show that orderings between the scores for the different kinds of motions are preserved.

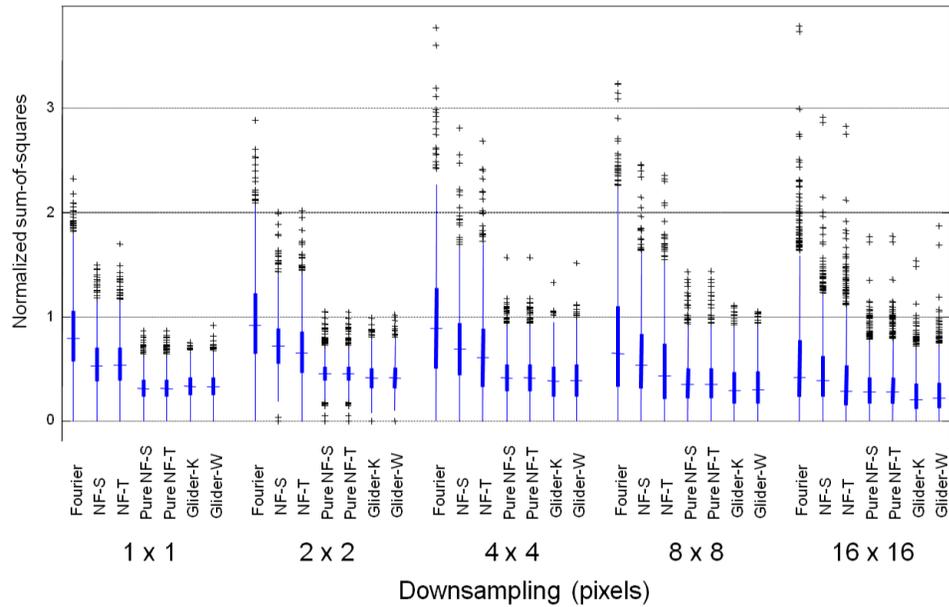


Figure 2.13: Dependence of motion signals on spatial discretization, for the PMO score.

A single movie (movie 4 of the main text, similar movie as in Figure 2.11, “Mr. & Mrs. Smith”) was re-analyzed at the original resolution and at four levels of downsampling, in which a single check in the ROI corresponded to an $n \times n$ block of pixels in the original movie. The level of downsampling ranged from 1×1 (no downsampling, corresponding to Figure 2.8 of the main text) to 16×16 . The ROI was oriented in the YT plane. For other details see Figure 2.11.

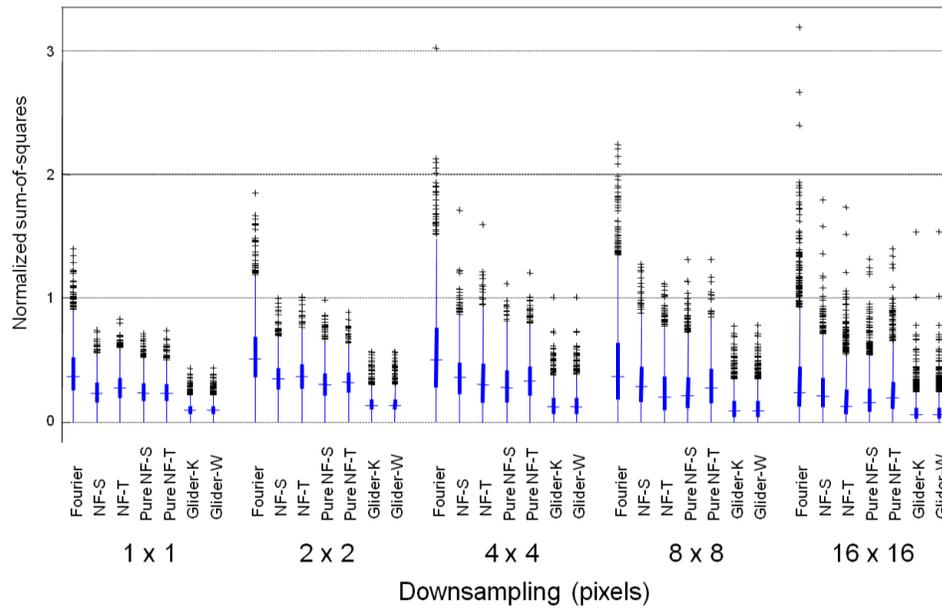


Figure 2.14: Dependence of motion signals on spatial discretization, for the RMO score.

A single movie (movie 4 of the main text, similar movie as in Figure 2.11, “Mr. & Mrs. Smith”) was re-analyzed at the original resolution and at four levels of downsampling, in which a single check in the ROI corresponded to an $n \times n$ block of pixels in the original movie. The level of downsampling ranged from 1x1 (no downsampling, corresponding to Figure 2.9 of the main text) to 16x16. The ROI was oriented in the YT plane. For other details see Figure 2.11.

Orientation of motion direction In the main text, we showed that there was a common pattern of motion scores for four movies; our analysis used an ROI (region of interest) set in the YT -plane and was therefore sensitive to horizontal motion (Figure 2.8 for PMO score, Figure 2.9 for RMO score). Figures 2.15 and 2.16 show when the ROI is set in the XT -plane (vertical motion), the pattern of motion scores is similar.

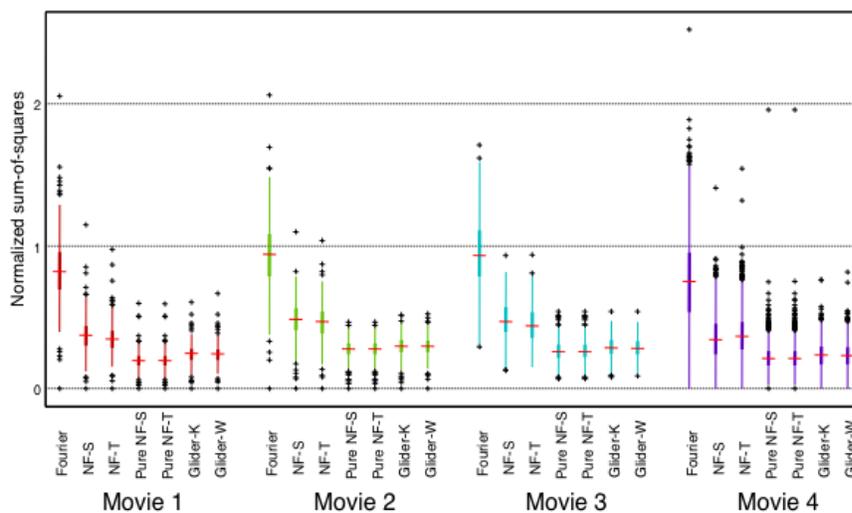


Figure 2.15: Prevalence of different kinds of motion signals is similar across movies, as measured by the PMO score in the XT -plane.

Values are normalized by the PMO scores obtained from movie of random pixels. The region of interest (ROI) consisted of a 4×4 block of checks in the XT plane; each check corresponded to a single pixel in the discretization of the movie. Movies were (1) "The 39 Steps" (1935), (2) "Anna Karenina" (1935), (3) "A Night at the Opera" (1935) and (4) "Mr. & Mrs. Smith" (2005). For other details see Figure 2.11.

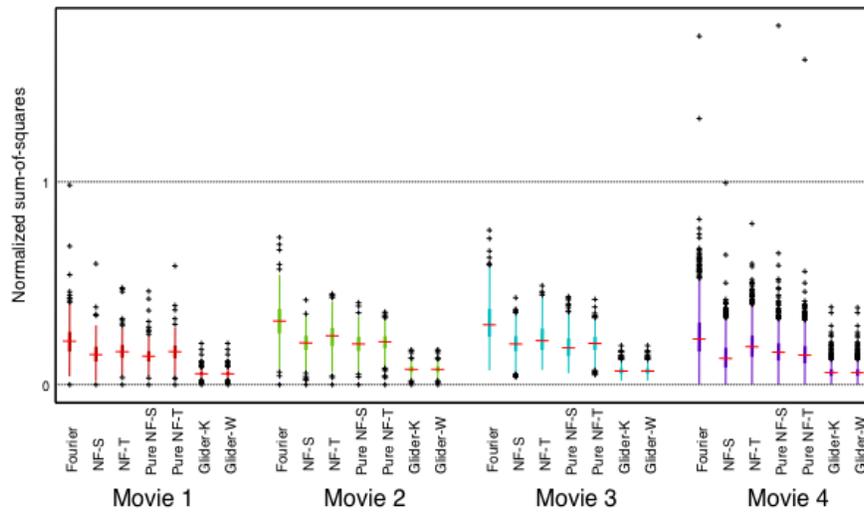


Figure 2.16: Prevalence of different kinds of motion signals is similar across movies, as measured by the RMO score in the XT-plane.

For other details, see Figure 2.15.

Alternative binarization: PMO/RMO score Here we show that the results in the main text hold when we use a binarization threshold equal to the midpoint of the global luminance range, rather than the median for each shot. Figures 2.17 and 2.18 shows that the results for the PMO and RMO scores respectively. The pattern of motion strengths is very similar to the results in the main text (Figures 2.8 and 2.9), obtained with binarization at the median of each shot.

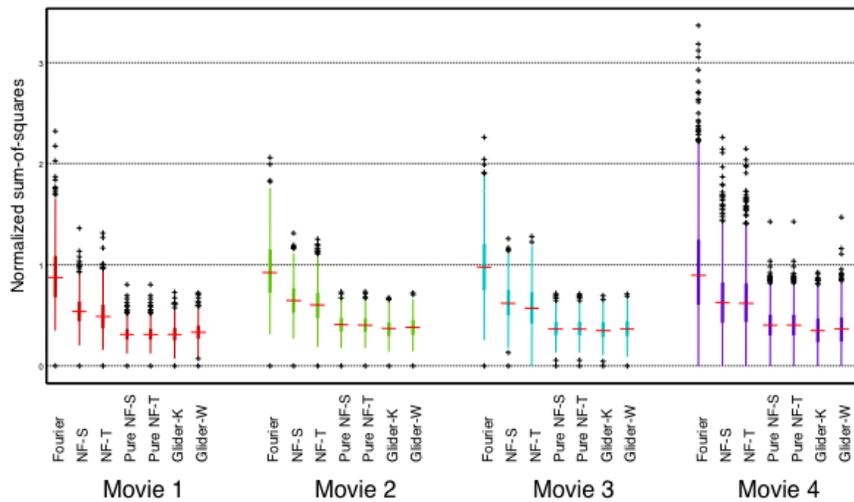


Figure 2.17: Prevalence of different kinds of motion signals is similar across movies, as measured by the PMO score.

Pixels luminance values were thresholded at the midpoint of the luminance range (i.e 127 out of range of 0-255)]. For other details, see Figure 2.8.

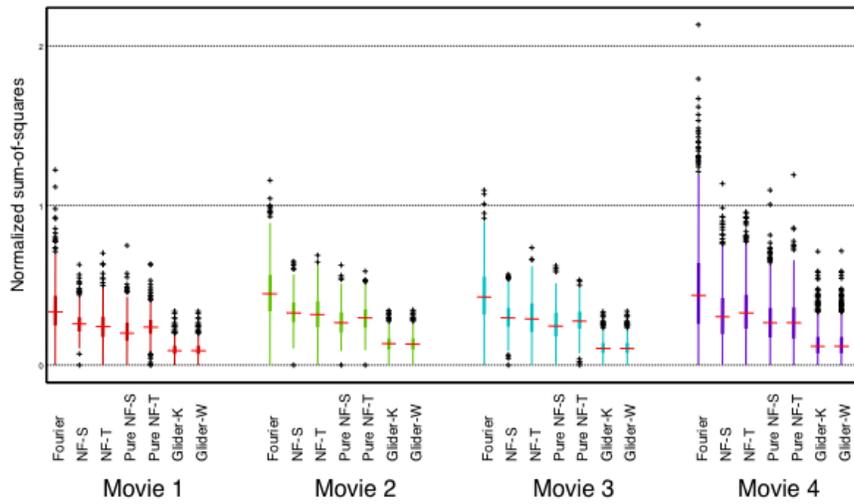


Figure 2.18: Prevalence of different kinds of motion signals is similar across movies, as measured by the RMO score.

For other details, see Figure 2.17.

Covariance of motion signals: auxiliary analyses Figure 2.10 of the main text showed how motion scores covary from scene to scene for movie 2 ("A Night at the Opera"); Figure 2.19 shows the parallel analysis of movie 4 ("Mr. & Mrs. Smith"). The main features of the plots are quite similar. For example, large values of the F motion RMO score tend to occur with large values of the G motion RMO scores, but the ratios between them (the bounding slopes of the scattergram) vary by approximately a factor of two. In Figure 2.20 we show that the pattern of covariance of motion scores arises from the structure of the movies themselves, not from the way that the scores are calculated. That is, we carry out the same analysis on random movies in which each pixel is independently assigned to black or white. As is evident, they are largely uncorrelated among most pairs of motion scores. (There are, however, correlations between F and pure-NF scores, and, necessarily, between the pairs of scores that are mathematically identical, as described below.) Notably, the scatter of the motion scores are much smaller in random movies. This is because for a random movie, non-overlapping ROI's are independent so that when local scores from a large number of ROI's are combined (approximately 256×256 ROI's per frame, and 500 frames per scene), the resulting sum has a very narrow fractional spread. Scatter is much larger for a naturalistic movie, since there are strong spatial and temporal correlations that correlate the ROI's within scenes. Figure 2.21 shows that the covariation of the subtypes of the PMO and RMO scores is similar when we use a binarization threshold equal to the midpoint of the global luminance range, rather than the median luminance of each shot (as in Figure 2.10 and 2.19).

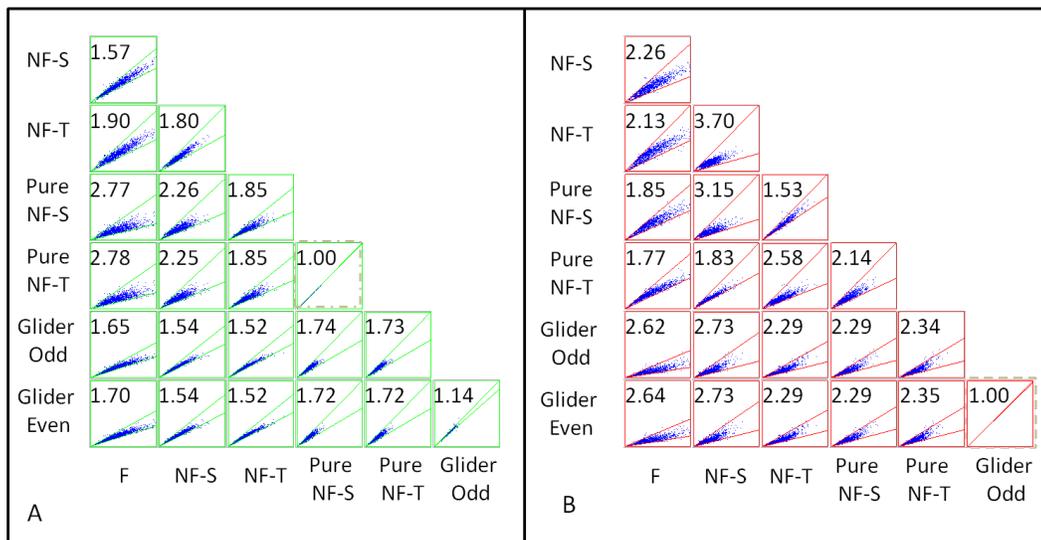


Figure 2.19: Covariance patterns of motion scores (*YT*; "Mr. & Mrs. Smith").

(Panel A: PMO; panel B: RMO. Within each scattergram, each point represents a pair of normalized motion scores determined from a single scene. Axes range from 0 to 2 (PMO) and 0 to 1 (RMO). The number in each plot indicates the average ratio between the pair of motion scores; the lines indicate the wedge that contains 95% of the values. Large values of one motion score typically occur with large values of the other scores, but the ratios between a pair of scores can vary by up to a factor of 2. (Pure NF-S and NF-T PMO scores are identical, and the two G motion RMO scores are identical; see below.) Analysis was carried out in the *YT* plane, for movie 4 ("Mr. & Mrs. Smith").

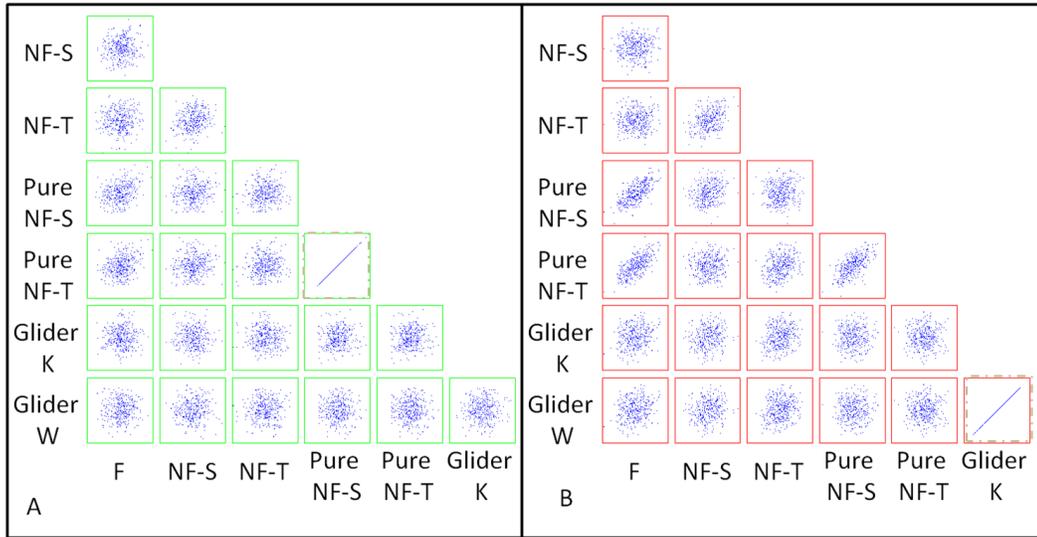


Figure 2.20: Covariance patterns of motion scores from 500-frame scenes drawn from a random movie.

Panel A: PMO; panel B: RMO. Within each scattergram, each point represents a pair of normalized motion scores determined from a single scene. Axes range from 0.995 to 1.005 for RMO and PMO; the small scatter is due to the very large number of uncorrelated ROI's that make up each scene. Analysis was carried out in the *YT* plane.

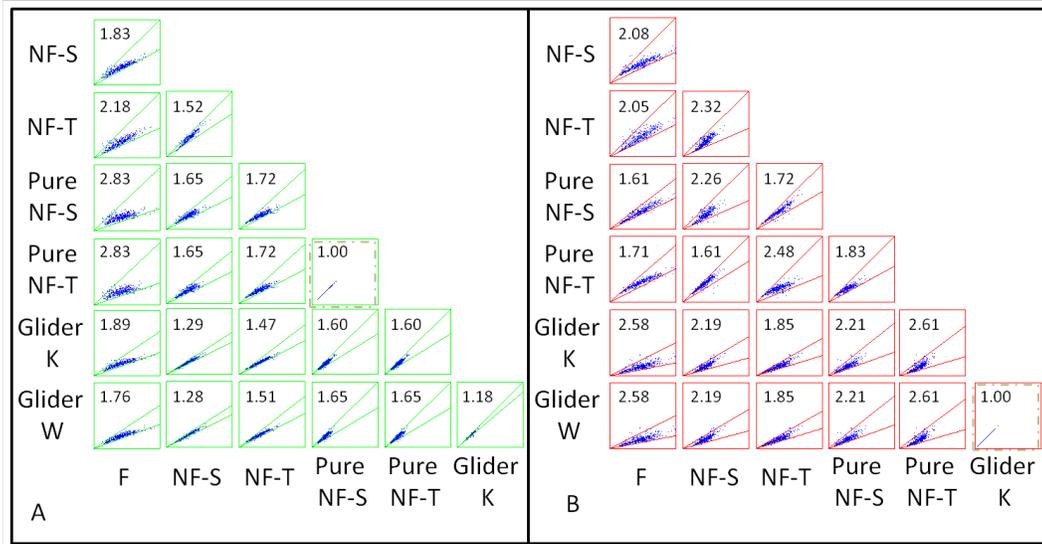


Figure 2.21: Covariance patterns of motion scores (*YT*; "A Night At The Opera")

Panel A: PMO; panel B: RMO. Pixels luminance values were thresholded at the midpoint of the luminance range (i.e. 127 out of range of 0-255). Analysis was carried out in the *YT* plane, for movie 2 ("A Night At The Opera"). For other details, see Figure 2.19.

Properties of the motion scores

Relationship to the Reichardt model As mentioned in the main text, the motion scores developed here are intended to compare several kinds of local motion signals on an equal footing. We chose a computational structure for the scores that would enable a comparison between classical (F) motion signals and nonstandard ones (NF and G motion), for which there is no standard computational model. Here we make contact between this approach and the standard model-based approach for F motion, i.e., the Reichardt model. Specifically, we show that the RMO score for F motion coincides with the output of a Reichardt model for binary spatiotemporal movies. For non-binary movies and for the PMO score, the F motion scores correlate strongly with the output of the Reichardt model, but the correspondence is not exact. To see why this correspondence is exact for binary movies and the RMO score, we note that for binary images, pairwise correlation, which amounts to multiplication, is equivalent to calculating parity. That is, taking black=1 and white=-1, the correlation across two checks is 1 if the parity of the number of black checks is even (i.e. two black checks or two white checks), and -1 if the parity is odd (i.e., one black check and one white check.) As a consequence, the initial stage of the Reichardt model [1], which calculates correlation by multiplication, corresponds precisely to determining whether the coloring of a two-check template follows the parity rule. Similarly, the second stage of the Reichardt model, at which opposing spatiotemporal correlations are subtracted, corresponds to comparing the number of times that the parity rule is followed, for templates in the two opposing directions. The latter is precisely what is computed by the opponent stage of the RMO procedure. Thus, the RMO score and the Reichardt detector output is perfectly correlated (Figure 2.22A). The exact correspondence breaks down

for the PMO procedure (Figure 2.22), since the PMO score in each direction is not simply the sum of the number of local rule violations, but rather, the number of checks whose state must be changed so that the entire ROI contains no parity-rule violations. While these quantities are distinct, they are nevertheless strongly correlated (Figure 2.22B). The precise correspondence also breaks down for grayscale images (Figure 2.22C, D), since multiplication of grayscale values is not equivalent to a parity count on their binarized levels.

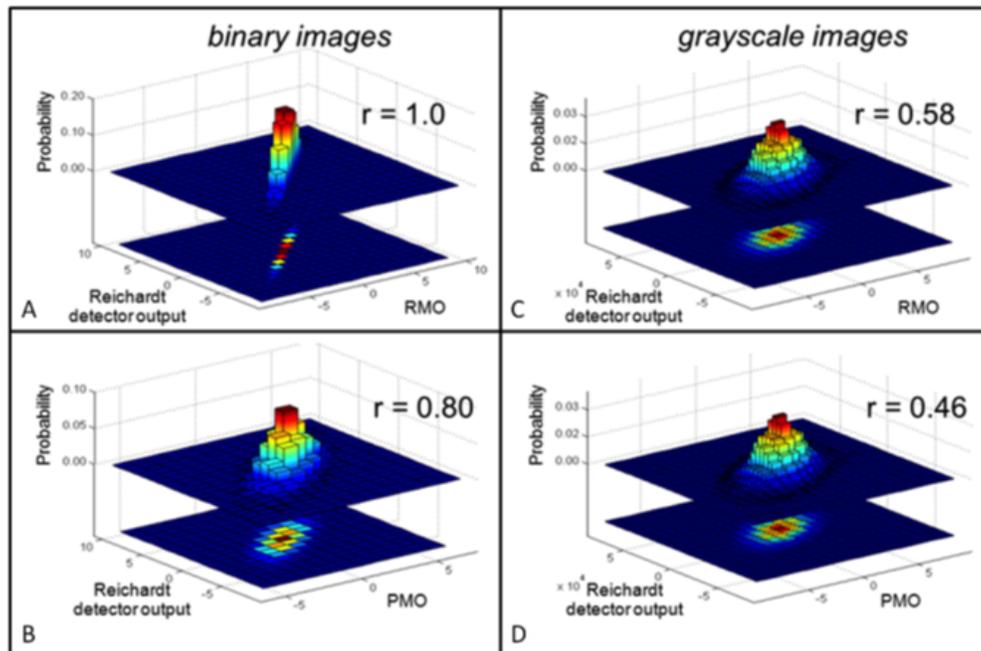


Figure 2.22: Correlations between RMO Fourier score, PMO Fourier score, and output of a Reichardt detector.

Left, an exhaustive analysis of all binary images in a 4×4 ROI. The RMO score (A) correlates perfectly with the output of the output of a Reichardt detector; the PMO score (B) is strongly correlated to the Reichardt detector output, but not identical to it. Right, a parallel analysis of grayscale images. Gray levels are drawn independently from a uniform distribution. For each random example of a 4×4 grayscale image, the summed outputs of Reichardt detector are compared with the RMO (C) and PMO (D) Fourier scores of the corresponding binarized images. The scores are highly correlated with the Reichardt detector output; the correlation is stronger for the RMO score.

Exact relationships among some kinds of motion scores Here we demonstrate some mathematical details concerning the relationships of the scores: (a) that for the two varieties of G motion, the RMO scores must be identical (but the PMO scores need not be), and (b) that for the two kinds of pure NF motion, the PMO scores must be identical (but the RMO scores need not be). These relationships are suggested by the data of Figures 2.10, 2.9, and 2.21; here we show that they are guaranteed mathematically.

(a) First, we consider the relationship of the G-W and G-K motion RMO scores. In both cases, the initial step of the calculation depends on counting the number of placements of a three-check template for which the colorings either follow (G-K) or violate (G-W) the parity rule. Denote the former by N_K and the latter by N_W . Since every coloring either follows or violates the parity rule, it follows that $N_K + N_W = P$, where P is the number of placements of a triangular glider template within the ROI. (For a $1 \times 4 \times 4$ ROI, $P=9$). It therefore follows that $N_W = P - N_K$. Consequently, at the spatial opponent stage (Figure 2.4C of the main text), the value for the leftward-rightward difference signals will be opposite:

$$N_W^{L,forward} - N_W^{R,forward} = -(N_K^{L,forward} - N_K^{R,forward}) \quad (2.9)$$

Beyond this stage, there is one more subtraction (the forward vs. backward contrast), and the resulting values are then squared and summed to obtain a final RMO score. The forward vs. backward contrast maintains the equal-and-opposite relationship between the G-W and G-K scores, so that when values are squared and summed, the resulting scores are identical.

Note that this relationship does not hold for the PMO score; the latter is sensitive to which placements violate the rule, as switching the state of one

check can affect the parity of more than one glider placement. Thus, the Hamming distances from the ROI to a perfect-match coloring need not be identical for G-W and G-K.

(b) Next, we consider the relationship between the PMO scores for spatial and temporal pure non-Fourier motion. The key step is to characterize the perfect-match ROI's, i.e., the ROI's in which every template placement follows the non-Fourier (NF) rule but does not follow the Fourier (F) rule. As we will show, the perfect-match ROI's are identical for the spatial and temporal cases. Since the PMO scores are determined by the distance from the ROI to the closest perfect match, it follows that PMO scores must be identical too.

To characterize the perfect-match ROI's for the spatial case, we start with Figure 2.2A, which enumerates all of the template colorings that follow the spatial non-Fourier rule (NF-S). Of these, the four on the left follow the Fourier rule; the pure NF-S library includes only the template colorings on the right. These have a simple description: in each case, the color assigned to a check at (y,t) is the opposite of the color assigned to the check at $(y+1,t+1)$. Conversely, all such colorings are present in these four templates. This description is symmetric in space and time, and therefore, also applies to the pure temporal non-Fourier rule (pure NF-T). Thus, perfect-match ROI's for the spatial rule are also perfect-match ROI's for the temporal rule. Since PMO scores depend only on Hamming distances from these perfect matches, the pure NF-S and pure NF-T PMO scores must be identical.

A caveat concerning analysis of synthetic motion stimuli

Perhaps the best-known examples of non-Fourier motion stimuli are the “drift-balanced” stimuli of Chubb & Sperling [2]. These are synthetic stimuli in which the moving object is distinguished from the background by the presence of random spatial or temporal contrast changes. The simplest example of this category of stimuli consists of a randomly flickering check moving across a gray background ([2], Figure 2.4A.). Fourier signals are eliminated by ensuring that the mean of the randomly flickering check is equal to that of the background. We focus on this stimulus here because it highlights a technical point that is crucial in analyzing synthetic spatiotemporal sequences.

Figure 2.23 shows how our analysis with the NF-S (panel A) and NF-T (panel B) templates applies to this stimulus. In the left column, the templates are placed in a way that is aligned with the stimulus grid. Note that each placement of the template always includes a check that is entirely contained within a gray region of the background (contrast=0). Hence, the products of the contrast values within the four checks is always 0, and no local motion signal is generated. In the second column, the template is displaced by a fraction of the discretization unit of the stimulus: in A (the spatial case), the displacement is a fraction of the check width; similarly, in B (the temporal case), the displacement is a fraction of the frame duration. The contrasts within these displaced templates are shown in the right column. In the spatial case (A), all of the checks of the NF-S template see a nonzero local contrast, and the products of these contrasts is always positive, producing a net NF-S local motion signals. Similarly, in the temporal case (B), an NF-T local motion signal is produced. Thus, generically positioned templates detect the NF motion, while templates that are

aligned to the stimulus grid do not.

The reason for this is as follows. Formally, this drift-balanced stimulus - like the other NF stimuli of Chubb & Sperling [2] - carry a motion signal because they contain many coherent correlations among four points in a slanted region of space-time. Consider movement in the $+Y$ direction, rendered in a stimulus of grid size Δy and frame duration Δt . For any two nearby points (y_1, t_1) and (y_2, t_2) that are inside the moving stimulus check within the same frame, then $(y_1 + \Delta y, t_1 + \Delta t)$ and $(y_2 + \Delta y, t_2 + \Delta t)$ will always be inside the same moving stimulus check on the next frame. This in turn means that for any two such points (y_1, t_1) and (y_2, t_2) , the product $I(y_1, t_1)I(y_2, t_2)I(y_1 + \Delta y, t_1 + \Delta t)I(y_2 + \Delta y, t_2 + \Delta t)$ will always be positive, whether or not the check flickers as it moves. In contrast, the opposing-direction product will always be zero. Each of these imbalances of fourth-order correlations represents a motion signal. Since the NF-S template computes such a product, the NF-S template will detect this signal provided that two of its template checks are inside the same stimulus check (as shown in the middle and right columns of Figure 2.23A). This will generically be the case unless the template is always positioned to have its spatial boundary coincident with an edge of the stimulus grid (as shown in the left column of Figure 2.23A). Put another way, if the template is placed randomly on the movie (without regard to stimulus-check boundaries), then typical placements will yield a proper motion signal. Figure 2.23B shows that the same kind of phenomenon occurs in the time domain: the NF-T template will also detect the correlations that carry the motion, provided that its temporal boundaries occasionally occur during a frame. (Another way of looking at this is that the motion of the original Chubb & Sperling stimulus is defined by correlations of unsigned contrasts across space and time. This is captured by the fourth-order

product $I(y, t)^2 I(y_1 + \Delta y, t + \Delta y)^2$, which is in turn equivalent to the generic product $I(y_1, t_1) I(y_2, t_2) I(y_1 + \Delta y, t_1 + \Delta t) I(y_2 + \Delta y, t_2 + \Delta t)$ provided that (y_1, t_1) and (y_2, t_2) are within the same stimulus check.)

Note that this issue - the need to consider generic placement of the template with respect to the stimulus grid - does not arise in the analysis of digitized naturalistic movies, since the check and frame boundaries used for digitization bear no fixed relationship to the features in the original image. Thus, even if the grid used for the templates is the same as the grid of the movie digitization, the relationship of the templates to the objects in the original image will still be generic.

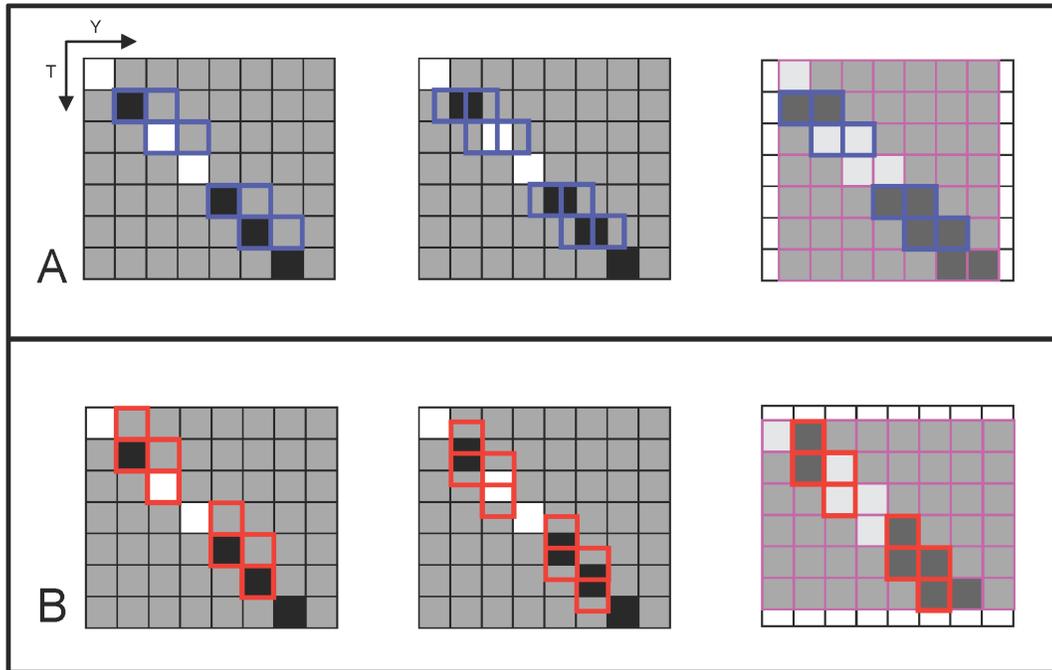


Figure 2.23: analysis of artificial stimuli generated on a grid.

For analysis of artificial stimuli generated on a grid, consideration of generic placements of the template is crucial. This is shown for analysis via the NF-S template (panel A) and the NF-T template (panel B), as applied to the classic NF stimuli presented by Chubb & Sperling [2]. When the analysis template is aligned to the stimulus grid (left columns), at least one of its checks is entirely within the background. Thus, the product of the contrast values in the four template checks is zero, and no motion signal is detected. However, any misalignment (middle columns) allows all four template checks to be influenced by the stimulus, so the product of the four contrast values captures the motion (right columns).

2.6.2 Movies

Description of the supplementary movie clips The first set of clips (1A to 1C in Table 1 below) are synthetic binary movies that show examples of three kinds of motion: Fourier (F), non-Fourier spatial (NF-S), and glider (here, glider expansion for black, G-K). Each movie consists of fully coherent, unidirectional motion to the right, and is synthesized by the method of Hu et al. [3]. Clips are 10 sec long (10 frames/sec) and each frame is 16 x 16 pixels.

The second set of clips (2A to 2F in Table 1 below) consist of a short snowboarding clip annotated to illustrate where different motion types occur. Clips are 256 x 256 pixels and 15 frames/sec. Original clip was captured by one of the authors in upstate NY. Clip 2A is the original scene at the above-specified resolution. In clip 2B, the scene is binarized around its median. Clips 2C, 2D, and 2E annotate the clip based on three kinds of motions; the pixels are colored green where the RMO score for horizontal motion of each type is nonzero. Finally, clip 2F annotates the three motion types together, superimposing red for F, green for NF-S, and blue for G-K (Table 2), to make the differences in the locations of the motion types more evident.

Table 2.1: Clips details

Index	Source	Description	File name¹
1A	Synthetic	F motion	synthetic_bin_F
1B	Synthetic	NF-S motion	synthetic_bin_NFS
1C	Synthetic	G-K motion	synthetic_bin_GL_K
2A	Snowboarding	original clip	snowboard_orig
2B	Snowboarding	binarized clip	snowboard_bw
2C	Snowboarding	scene annotated for F	snowboard_F
2D	Snowboarding	scene annotated for NF-S	snowboard_NFS
2E	Snowboarding	scene annotated for G-K	snowboard_G_K
2F	Snowboarding	scene annotated for F	snowboard_all

Table 2.2: Annotation scheme for clip 2F

#	Annotation	F	NF-S	G-K
1	Black	N	N	N
2	Red	Y	N	N
3	Green	N	Y	N
4	Blue	N	N	Y
5	Purple	Y	N	Y
6	Yellow	Y	Y	N
7	Cyan	N	Y	Y
8	White	Y	Y	Y

¹All clips are in mov format that is played best using quicktime player (<http://www.apple.com/quicktime/>)

CHAPTER 3
EVOLUTIONARY CONVERGENCE OF MOTION COMPUTATION IN
MONKEY AND DRAGONFLY

3.1 Abstract

The ability to detect moving objects is as important as correctly identifying them. Neurophysiologic evidence in primates indicates that in mammalian cortex, high-order, nonlinear mechanisms are used to extract and process motion cues. Motion sensitivity in insects rivals that of mammals. Dragonflies are swift, acrobatic fliers whose acutely motion-sensitive vision enables them to capture aerial prey. We present neurophysiologic findings that mammalian visual cortex and dragonfly brain process motion using similar algorithms: they respond to multiple types of motion cues in very similar ways. The commonality of motion computations at the neuronal level indicates that two extremely dissimilar animals possessing non-homologous visual brain centers have converged on a common computational scheme for detecting visual motion.

3.2 Introduction

Objects come and go in the flux of any visual scene. Animals possess an overwhelming diversity of visual organs – eyes – for capturing features of a visual image. The extraction of motion cues, however, is mostly done in the brain. Primates are near the apex of visual performance: 25% to 40% of their brain is involved in processing vision [41]. Humans and monkeys detect movement

with high sensitivity, as exemplified by neural activity in the primary and secondary visual cortex, V1 and V2 [42, 43].

However, primates are not the only sharp-eyed animals in which visual motion sensitivity and visual acuity are shaped by natural selection – insects rival primates [44]. Among insects, dragonflies have the highest visual acuity and are the swiftest and most acrobatic fliers [45]. The ancestry of modern dragonflies can be traced back over 300 Myr; primates appeared only 55 Myr ago [46]. Dragonflies spend most of their active lives in the air chasing down their aerial prey, actively defending their territorial boundaries, and even mating in mid-air. They possess remarkable movement vision [47]. No less than primates, dragonflies rely on their ability to detect visual motion for their survival since both are predominantly diurnally-active animals. Could this common evolutionary imperative lead to common, convergent mechanisms in deep neural processing or computation of visual motion in the brains of primates and dragonflies? In other words, what would it take to support the claim that primates and insects use the same strategies for motion estimation? We here present evidence that in spite of obviously vast morphological differences in their eyes (simple lensed vs. compound), brain size, brain organization, number of neurons, body plan, life histories, and over 300 Myr of evolutionary separation, single units recorded from brain visual areas in rhesus monkey and dragonfly are remarkably similar in their responses to multiple kinds of motion cues. Differences are subtle and may be attributed to terrestrial vs. aerial lives.

3.3 Results

To compare motion processing in the brain of macaque and dragonfly (Figure 3.1, we used recording and analysis techniques that were as similar as possible across the two visual systems, and probed their visual systems with representatives of the diverse motion signals that exist in natural stimuli [8, 17]. For recording, we used metal microelectrodes to record visual responses from primary and secondary visual cortices, which are considered to be the mainstream visual area, (area V1 or V2) of the monkey and from the medulla or lobula complex of the dragonfly (Figure 3.1A). Although many studies had been conducted on invertebrates and in particular on flies, only a few had been done on dragonflies. This is the first substantial recording of spiking neurons from the latter species. These extracellular recordings yield one to several neural units that are parsed, and identified single units on the basis of their amplitude and waveform using spike-sorting software (see SM 3.5). Example visual unit responses are shown for each species as a raster plot (Figure 3.1B). Although, in general, neurons that came from similar brain areas of the same animal may present some dependencies in their response, the claims of similarity made in this chapter hold, even when statistical tests (i.e. jackknife statistics) that ignore these dependencies are employed.

Figure 3.1C shows the two most important motion probes that we used, and examples of neural responses to them. The first probed the classic local motion signal (Figure 3.1C, top row) [1], consisting of correlations between two nearby points in space and time. The black-and-white movie shown in figure 3.1C (as a series of frames, and also as a slice across time) exemplifies this signal, because every check on one frame is always displaced in the same direction on

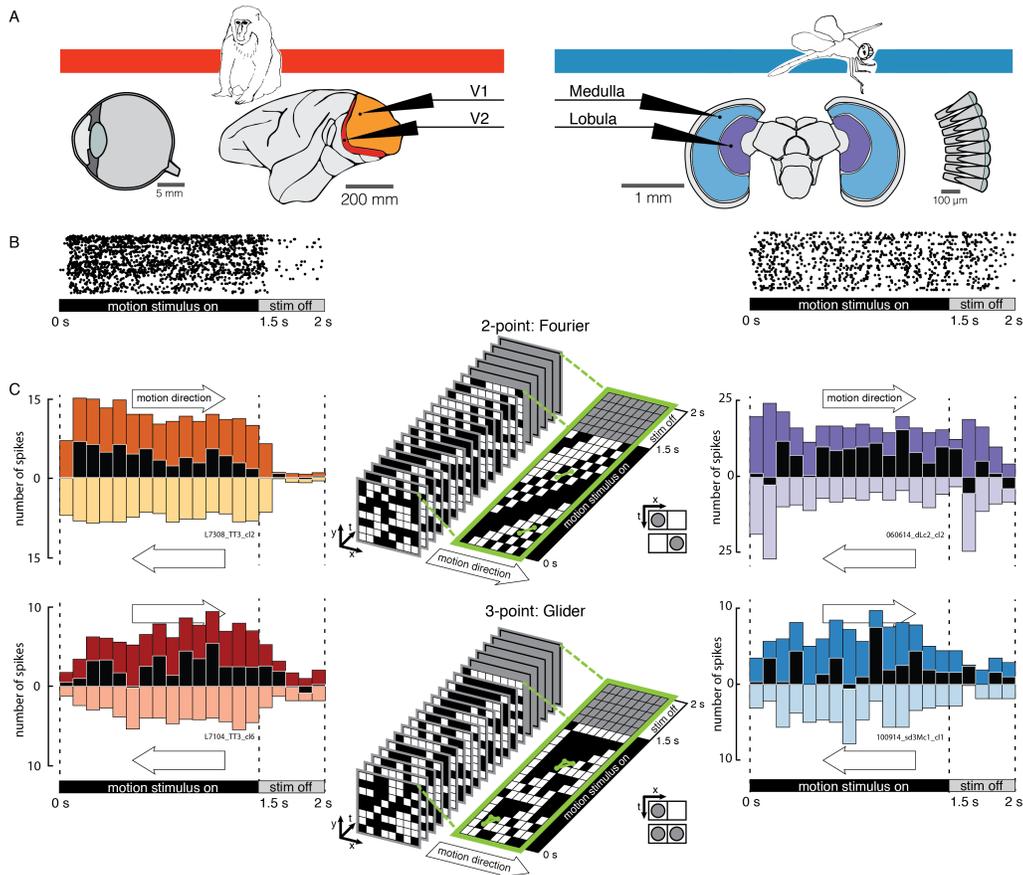


Figure 3.1: Analysis of single unit responses to elementary motion signals in the brain of macaque monkeys and dragonflies.

(A) The two species have extreme differences in body plan and size, brain, and eye morphology (“simple” camera-like eye in the macaque, versus compound eye in the dragonfly), but in both species, initial visual processing in the brain is hierarchical. (B) Extracellular methods enabled recordings from single units in two early sequential regions of the visual brain (V1 and V2 in the macaque; medulla and lobula in the dragonfly) in response to motion stimuli presented on a video screen. (C) Stimuli were black-and-white movies on a 16×16 grid that isolated different kinds of local motion signals. Top row: a movie that isolates the Fourier motion signal. After an initial random frame, checks are colored to match within a diagonal space-time template (small icon in middle). Bottom row: a movie that isolates glider motion (here, black expansion subtype). A check is colored black if both or neither of its neighbors in the triangular template area also black. Each movie segment was shown for 1500ms (15 frames), followed by 500ms of gray, and repeated 32 times in interleaved trials. See SM 3.6 for details. Histograms show example neural responses to corresponding motion types presented moving in opposing directions, with black bars showing the difference. Histogram colors correspond with the brain area of the recording site.

the next. This stimulus property can be summarized by a "template" consisting of two checks along a space-time diagonal (inset): for every placement of this template on the movie, either both checks are black, or both checks are white. This motion signal has also been called the "Fourier" motion signal, since it can be extracted from the spatiotemporal Fourier amplitudes of the movie.

Figure 3.1C (bottom row) shows another kind of motion signal – the "glider" motion signal [3]. It consists of correlations between three nearby points in space and time. The movie spatiotemporal slices shown in the bottom row exemplifies this signal: every triangular template in a particular space-time orientation contains either 1 or 3 black checks. As shown, this leads to a movie containing rows of black checks that expand in time (i.e., which appear to loom), and is therefore designated "black glider expansion." We also used related glider motion signal subtypes, in which the rows of black checks contract in time, or in which the expanding or contracting elements consist of white checks (see Figure 3.5). These subtypes indicate looming vs. receding of an object, and its contrast relative to the background. Fourier and glider motion signals are computationally distinct: the glider movie of Figure 3.1C (bottom) has no pairwise correlations corresponding to the Fourier template, and the Fourier movie of Figure 3.1C (top) has no three-point correlations corresponding to the glider template. This means that the glider motion is not detectable by the classical Reichardt model, nor are the similarities that imply from it.

A third kind of motion signal used in these studies is an example of the "non-Fourier" motion signal, identified by Chubb & Sperling [2]. Non-Fourier motion is often referred to as "feature motion" or "second-order" motion [17] and contains correlations among four points in space and time. Since our main

concern is with Fourier and glider motion we leave description of the non-Fourier motion for Supplemental information. See SI section 3.5.2 and Figure 3.5 for details on non-Fourier stimuli and their construction. These motion elements (Fourier, glider, and non-Fourier) are known to elicit behavioral responses in a wide range of species (Drosophila: [7]; zebrafish: [6]; human: [3, 48]); for two of the elements (Fourier and non-Fourier), neurophysiologic correlates in the mammalian visual cortex are documented (macaque: [49]; cat: [50]). Furthermore, from purely mathematical viewpoint, observing responses to one motion kind, does not contribute any information about the presence of another kind of motion signal. Here we show that all of these motion elements, including the multiple subtypes of glider motion, are detected by central visual neurons in the macaque and the dragonfly, and that the characteristics of this neural response are strikingly similar.

Figure 3.2 and Figure 3.6 in the SI, summarize the prevalence of neurons with directionally-selective responses to each motion type. Each kind of stimulus elicits directionally-selective responses in V1 and V2 of the monkey and the medulla and lobula of the dragonfly. In both species, the proportions of neurons that are sensitive to each motion type (detailed in Figure 3.5) are similar. Notably, the neural circuitry for analyzing local motion signals does not follow the clean mathematical segregation between Fourier, glider, and non-Fourier motion signals, which typically coexist in natural visual inputs [8]. Rather, in both species, there are neurons that respond to more than one motion type, and the frequency of neurons that respond to each combination of motion types is similar across species.

While Figure 3.2 shows that the two species are similar in terms of sen-

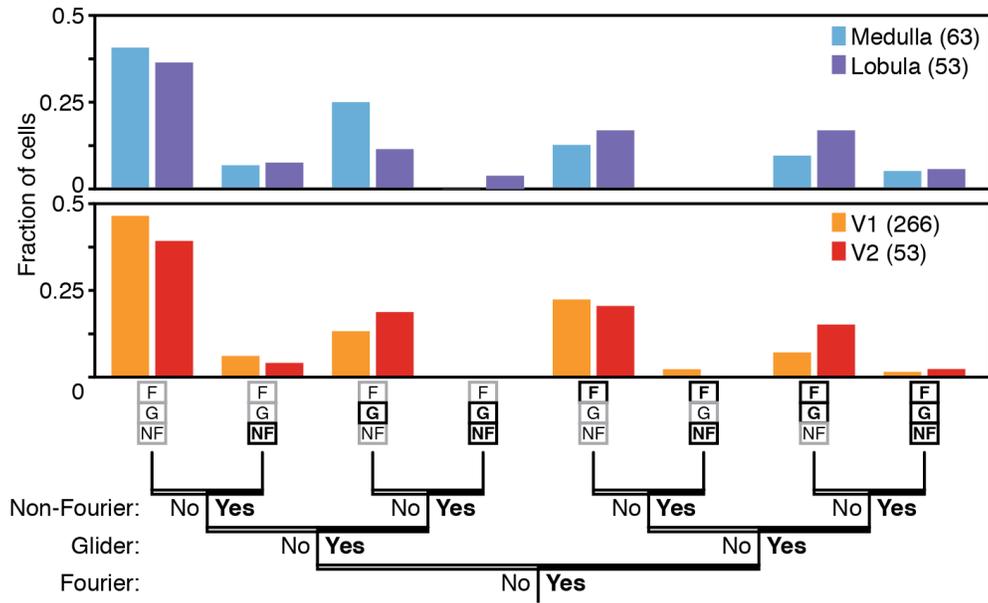


Figure 3.2: In dragonflies (top) and macaques (bottom), similar fractions of central neurons respond to Fourier, glider, and non-Fourier motion signals.

Neurons were considered to have a significant response to a kind of motion if their firing rate was significantly larger for stimuli with motion in one direction than in the other ($p < 0.05$, two-tailed paired t-test). In both species and for both brain regions, similar fractions of neurons responded to each of the three major motion signals, and similar fractions of neurons responded to more than one kind of motion. In both species, many cells did not respond selectively to any motion stimulus (far left bars), as is typical of recordings from the visual brain.

Similarities to motion signals at the population level, Figure 3.3 shows that the species are similar on a neuron-by-neuron basis. For each neuron, we compare its direction selectivity for Fourier (standard) motion signals with its direction selectivity for the other types: the glider subtypes mentioned above, "reverse phi" Fourier signals [51] and their non-Fourier analogs. As is standard [52], we quantify direction selectivity for Fourier motion as the Direction selectivity Index (DI)

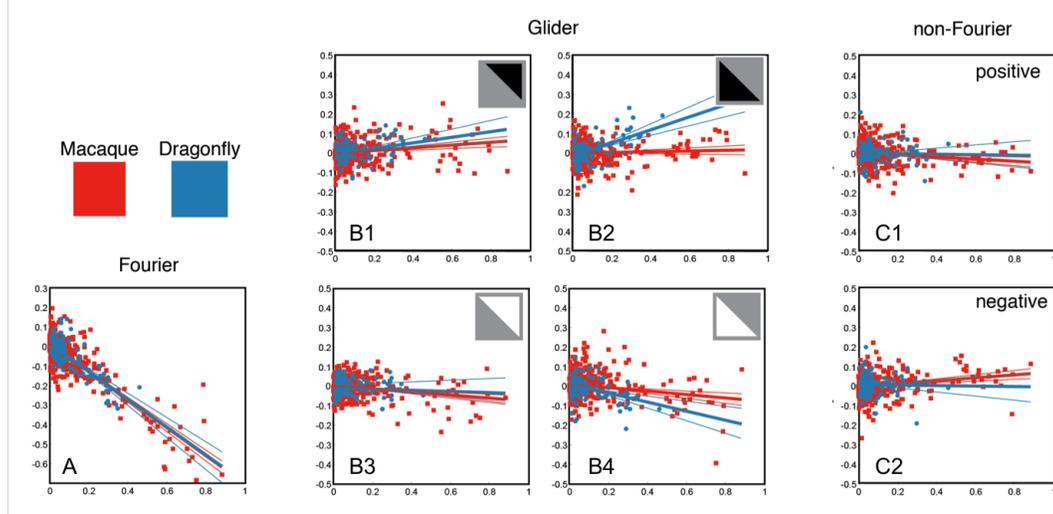


Figure 3.3: In dragonflies (blue) and macaques (red), combination of motion signals at a cellular level is similar.

For each unit, responses to the several motion subtypes were quantified by the direction selectivity index (DI), and plotted against the unit's DI for Fourier (standard) motion, on the abscissa. Motion subtypes consisted of: **(A)** negatively-correlated Fourier motion ("reverse-phi"), **(B1)** black glider contraction, **(B2)** black glider expansion, **(B3)** white glider contraction, **(B4)** white glider expansion, **(C1)** positively-correlated non-Fourier motion, and **(C2)**, negatively-correlated non-Fourier motion. Heavy lines indicate the regression line through the origin; thin lines are 95% confidence limits for slopes. Negative values of the DI indicate a direction preference that is opposite to the direction preference for Fourier motion. Lines are drawn to the limits of the data along the abscissa.

$$DI = \frac{M_{pref} - M_{non-pref}}{M_{pref} + M_{non-pref}} \quad (3.1)$$

where M_{pref} is the firing rate elicited by stimulus movement in the direction that is preferred for Fourier motion, and $M_{non-pref}$ is the firing rate for stimulus movement in the opposite direction. Thus a $DI = 0$ indicates a cell has no direction selectivity and a $DI > 0$ indicates the presence of direction selectivity (with maximum value of 1). We define the DI similarly for the other kinds of motion, but maintain the convention that M_{pref} is the preferred direction for Fourier motion. Thus, a $DI > 0$ indicates a motion preference in the same direction as for Fourier

motion and $DI < 0$ indicates a motion preference in the opposite direction.

In both species, neurons that have directionally-selective responses for one subtype often have directionally-selective responses for another (Figure 3.3). For some subtypes, (for example, glider expansion and contraction, black), the DI is mostly positive, and is also positively-correlated with the DI for standard (Fourier) motion. That is, neurons that prefer Fourier motion in one direction tend to prefer the other motion subtype in the same direction. But for other subtypes, (for example, glider expansion, white), the DI is mostly negative, and is negatively-correlated with the DI for standard motion. A neuron that is directionally selective to one of these motion signal subtypes tends to prefer Fourier motion in the opposite direction. This pattern of correlations is consistent across the two species, and is also found when each brain region is analyzed separately (figures 3.7, 3.8, 3.9, 3.10).

Thus, the fractions of neurons that are directionally selective to the motion signal subtypes (Figure 3.2) and the patterns of responses to these subtypes (the correlations in Figure 3.3) are similar across species. Overall levels of direction selectivity tend to be higher in the macaque, but this likely reflects a methodological difference (see SI 3.5.1).

These similarities suggest that extraction of motion is an ethologically crucial operation that, from a computational viewpoint, is achieved through similar algorithms in the macaque and dragonfly. The only exception is a quantitative one, related to glider expansion. Units in the dragonfly are more sensitive to glider expansion than comparable units in the macaque, and their directional responses to glider expansion are more strongly correlated with their selectivity for Fourier motion (Figure 3.3). We hypothesize that for fast-flying dragonflies,

rapidly looming objects could signal an impending collision, hence accounting for increased sensitivities to such signals compared to macaques.

Finally, we asked whether there was a parallel progression of motion processing from V1 to V2 and from medulla to lobula. To characterize the overall nature of local motion processing in each area, we focused on opponency, a computational characteristic that is relevant to extraction of all kinds of motion elements. Specifically, in the standard Hassenstein-Reichardt model [1], the motion signal carried by a directionally-selective neuron is the net result of a two-step process. First, pairwise spatiotemporal correlations are computed. Then, correlation strengths in opposite directions are compared via a subtraction, and the resulting signal is manifest in the neuron's firing rate. This basic architecture is shared by extensions of the Hassenstein-Reichardt model that extract glider [3] and non-Fourier motion [48] in which the opponent step is retained, but the initial computation of pairwise correlation is replaced by a more elaborate computation. The computational benefit of opponency is that it eliminates false-positive signals due to uniform flicker and static edges. However, this entails a functional disadvantage: negative correlations in one direction are indistinguishable from positive correlations in the other. A classic example of this confound is the "reverse phi" phenomenon: standard Fourier motion in one direction is confused with negative correlations in the opposite direction [51]. Another concerns the glider motion types: black glider expansion in one direction would be confused with white glider expansion in the opposite direction. Effective use of motion signals to guide action may require distinguishing between these alternatives – a distinction that, given its computational complexity, may be expected to occur downstream from computations based on simple opponency. That is, we hypothesize that motion computations by neurons in

V2 and lobula will deviate from simple opponency to a greater extent than the computations in V1 and medulla.

To test this hypothesis, we indexed the deviation from opponency by a “motion complexity” (MC) score.

$$MC = \sum_{m \in \{F, G\text{-exp}, G\text{-cont}, NF\}} |DI(m, +) + DI(m, -)| \quad (3.2)$$

,where $DI(m, +)$ is the DI of motion m with respect to the preferred direction and $DI(m, -)$ is the DI of motion m with respect to the non-preferred one.

The Motion Complexity (MC) score quantifies whether a neuron responds in the same way to positive correlations in one direction, and to negative correlations in its opposite: $MC = 0$ in this is the case, and $MC > 0$ if not. We hypothesized that in both species, it would increase as motion processing unfolds.

Figure 3.4 shows the MC score distributions in the two species. Consistent with the hypothesis that later stage visual processing begins to distinguish between negative correlations in one direction and positive correlations in the other, the distribution of MC scores shifts towards higher values in the second processing area (V2 or lobula), compared to the earlier processing (V1 or medulla). For each motion type there is a trend to higher values of MC scores as motion processing progresses (Figure 3.4A), and this trend becomes significant ($p < 0.001$ for V1/V2, Wilcoxon ranksum test) when all motion types are considered together (Figure 3.4B), indicating that the progressive deviation from opponency applies to processing of all motion elements.

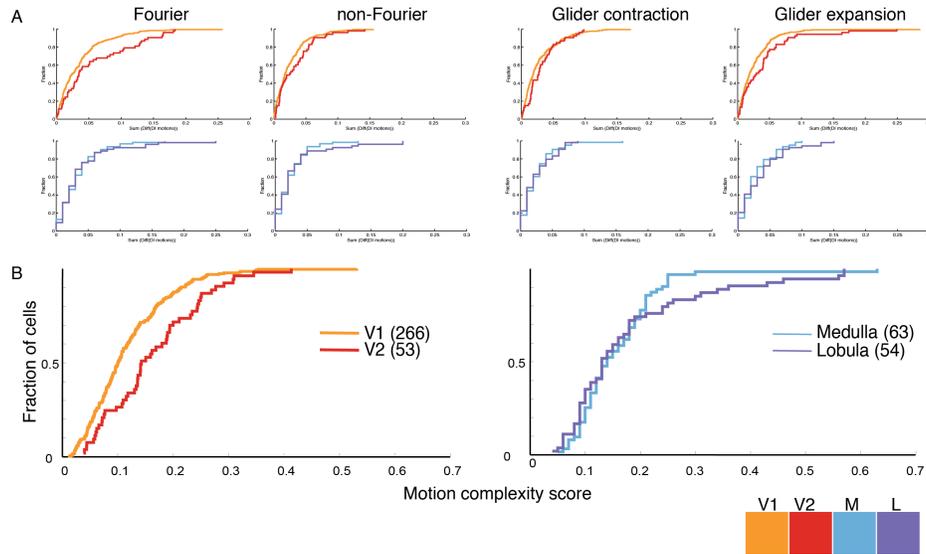


Figure 3.4: In macaques and dragonflies, motion processing unfolds in a similar fashion.

The motion complexity score (MC, see text above) indicates whether an individual neuron's response can resolve the ambiguities inherent in an opponency calculation (see text): neurons with higher MC scores are able to distinguish a positive motion signal in one direction from a negative motion signal in the opposite direction. We determined the distribution of MC subscores (A) and the total MC score (B). In macaques (B, left) and dragonflies (B, right), the distribution of MC scores shifts towards higher values from V1 (orange) to V2 (brown), and from medulla (light blue) to lobula (dark blue). Thus, as motion processing unfolds, neural computations progress beyond opponency, allowing neurons to distinguish positive spatiotemporal correlation in one direction from negative spatiotemporal correlation in the opposite direction.

3.4 Discussion

The ability to detect moving objects is as important as correctly identifying them. Neurophysiologic evidence in primates indicates that in mammalian cortex, high-order, nonlinear mechanisms are used to extract and process motion cues. Motion sensitivity in insects rivals that of mammals. Dragonflies are swift, acrobatic fliers whose acutely motion-sensitive vision enables them to capture aerial prey. We present neurophysiologic findings that mammalian visual cortex and dragonfly brain process motion using similar algorithms: they respond

to multiple types of motion cues in very similar ways. The commonality of motion computations at the neuronal level indicates that two extremely dissimilar animals possessing non-homologous visual brain centers have converged on a common computational scheme for detecting visual motion.

In sum, we compared the analysis of visual motion at the neuronal level in the central nervous systems of two extremely different visual specialists, whose evolutionary and anatomical divergences are profound: the macaque monkey and the dragonfly. We used artificial synthetic stimuli, designed to probe the nature of the neuronal computations by isolating individual motion elements that normally co-occur together in the natural environment (Figure 3.1). Across species, we find a similar spectrum of sensitivities to these elements, in terms of the proportions of neurons that are sensitive to each element (Figure 3.2), the way that these elements are combined by individual neurons (Figure 3.3), and how motion processing progresses from one central area to the next (Figure 3.4). We conclude that the ethological demands drive biologic motion processing to a highly convergent solution at the neuronal level despite major differences in their phylogenies and the architectures of their eyes and brains.

3.5 Methods

3.5.1 Physiology and Recording Methods

Macaque

Physiological preparation. Standard acute preparation techniques were used for electrophysiological recordings from V1 and V2 of cynomolgus monkeys (*Macaca fascicularis*) weighing 2.2 to 10 kg (12 males, 1 female). All procedures were approved by the Animal Care and Use Committee of the Weill Cornell Medical College and consistent with Institutional and National Institutes of Health guidelines for the care and experimental use of animals. Procedures were previously described in detail [53–56] and are summarized here. Animals were premedicated with atropine (0.05 mg/kg, i.m.; Henry Schein). Following ketamine (Ketaset, 10 mg/kg, i.m.; Fort Dodge Animal Health) or Telazol (4 mg/kg, i.m.; Fort Dodge Animal Health) and under isoflurane (12%; Hospira) surgical anesthesia, an endotracheal tube was placed, catheters were inserted in both femoral veins and one femoral artery, and a craniotomy was made near coordinates P10, L15. During recording, anesthesia was maintained with propofol (PropoFlo, 220 mg/kg/h, i.v.; Abbott) and sufentanil (Sufenta, 0.11 micrograms/kg/h, i.v.; Janssen) and neuromuscular blockade was established (following all surgical procedures) with vecuronium bromide (0.25 mg/kg, i.v. bolus, 0.25 mg/kg /h, i.v.; Bedford Laboratories) or rocuronium bromide (1.5 mg/kg, i.v. bolus, 1.5 mg/kg/h, i.v.; Mylan Institutional). During the experiment, heart rate and rhythm, arterial blood pressure, body temperature, end-expiratory pCO₂, urine output, and EEG were monitored. Routine maintenance

included intravenous fluids, periodic O₂ supplementation, antibiotics, dexamethasone, application of local anesthetics to surgical sites, and ocular instillation of atropine (1%; Bausch & Lomb), and flurbiprofen (Ocufer, 0.03%; Allergan), and periodic cleaning of the gas-permeable contact lenses (Metro Optics) behind 2-mm artificial pupils. Lenses with spherical correction, subsequently adjusted to maximize the responses of isolated single units to high-spatial-frequency visual stimuli, were used to focus the stimulus on the retina. With these measures, the preparation remained physiologically stable for 4-5 days.

Recording and visual stimulation. Through a small durotomy over V1 and/or V2, an array of 3 or 6 tetrodes (quartz-coated platinum-tungsten fibers; Thomas Recording, Giessen, Germany) was inserted, avoiding surface blood vessels via a custom headstage (Thomas Recording) that allowed for adjustments of the array geometry and for each tetrode to be independently lowered. Signals from each tetrode channel were amplified, filtered (0.3- 6 kHz), and digitized (25 or 30.303 kHz). Once spiking activity from one or more units was encountered, the region of the receptive field(s) was hand-mapped and then centered on the display of a 21-inch gamma-corrected CRT monitor, (1280 x 1024 raster, 100 Hz refresh), either a ViewSonic G225f 21-inch monitor (mean luminance 47 cd/m²) or a Sun GDM5410 21-inch monitor (mean luminance 46 cd/m²) at a distance of 114 cm. Control signals for the monitors were provided by a PC-hosted system optimized for OpenGL (NVidia GeForce3 chipset) programmed in Delphi.

Following hand-mapping, computer-controlled presentation of drifting sinewave gratings were used to characterize neural responses, including orientation tuning, spatial frequency tuning, temporal frequency tuning, and the

contrast-response function. One unit whose extracellularly-recorded action potential was identifiable by on-line spike sorting was chosen as the "target neuron." This neuron's orientation and direction preference was used to determine the orientation of the motion stimuli (see SI 3.5.2), and its spatial frequency optimum was used to determine the check size (approximately 2 checks per lobe of the optimal grating). For the motion stimuli, contrast was always 1.0 and stimulus velocity was always 10 checks per second. With the typical check size of 0.2 deg (rarely less than 0.1 deg or greater than 0.5 deg), stimulus velocity was 2 deg/sec.

Offline, recordings were spike-sorted as described in [53], based on automated clustering via KlusterKwik [57] operating on 17 features (peaks and troughs on each of the four channels, the first eight principal components of the wave shapes, and spike time), followed by hand merging and reclustering in Klusters. Criteria for single neuron isolation included waveform shape, its gradual change over time, and the number of refractory period violations.

Histology. Procedures were identical to that of [53]. In brief, lesions were made after all recordings were completed, and, following a waiting period of 1 h, the animal was deeply anesthetized and perfused (4% paraformaldehyde; EMS). The border between V1 and V2 was identified via the distinct appearance of layer 4 in V1 and its disappearance in V2. We marked each unit as certain V1, certain V2 or uncertain V1/V2. The latter units were only included in the analyses that were not subdivided according to area.

Dragonfly

Physiological preparation. Multiple dragonfly genera were used in these experiments including *Anax junius*, *Aeshna verticalis* and others, which came from one of two sources: wild-caught (Ithaca, NY; May-October 2013) and laboratory reared (Carolina Biological Supply Co.). Gathering data from several genera in dragonfly is a common procedure. Recordings were made from a total of 26 animals.

After capture, wild-caught dragonflies were held for short periods that did not exceed 15 hours in the laboratory before use. Laboratory reared animals arrived as penultimate nymphs and were raised in house (12:12 light/dark cycle; 80% humidity; 27 deg C) in individual containers and fed a diet of mosquito larva until eclosion. All experiments with laboratory reared dragonflies took place no more than 48 hours after eclosion.

Just prior to the start of each experiment, dragonflies were cold anesthetized for 2-4 minutes in a freezer (-4 deg C). Dragonflies were then restrained and affixed to a plastic post using Kerr dental sticky wax (58 deg C melting point, Syborn Kerr, Emeryville, CA, USA) heated by a cool soldering iron (Antex model C, Antex (Electronics) Limited, Trivestock, Devon, UK) with the voltage limited to 55V using a variable transformer (Powerstat type 3PN116B, The Superior Electric Co., Bristol, CT, USA). The animal was positioned ventral-side-down on the post, and the head was tilted downward such that the dorsal high-acuity fovea was pointed towards the screen. A small flap of cuticle on the anterior portion of the head, between the eyes and the neck (thorax), was removed to expose the right optic lobe of the brain over the medulla and lobula. A drop of fresh extracellular saline solution containing (in mM): 185 NaCl, 4

KCl, 6 CaCl₂, 2 MgCl₂, 10 HEPES, and 35 D-glucose (solution adjusted to pH 7.2 with NaOH and 430 mOsm with glucose; [58]) was placed on the brain at least every 30 minutes.

Recording and visual stimulation. Recordings were made using tungsten microelectrodes (4M Ω ; MicroProbe Inc., Gaithersburg, MD, USA) mounted to stereotactic micromanipulators (Narishige International USA, Inc. , East Meadow, NY, USA) and advanced using a hydraulic microdrive (Model 607W, David Kopf Instruments, Tujunga, CA, USA) at 1 μ m steps once inserted into the brain. Electrode placement into the medulla or lobula was determined visually by the experimenters using anatomical landmarks [59]. Electrical activity was acquired via an extracellular headstage (Model 1800 A-M Systems, Sequim, WA, USA) and amplified 10,000x and filtered (100Hz-5000Hz bandpass, 60Hz notch) using a differential AC microelectrode amplifier (Model 1800 A-M Systems, Sequim, WA, USA), followed by an A/D converter (NI PCI-MIO-16E-1, National Instruments, National Instruments, Austin, TX, USA) fitted with a breakout box (NI BNC-2090, National Instruments, Austin, TX, USA). All recordings were made at 15kHz sample rate using the Spike Hound data acquisition software (formerly called g-Prime; [60] on a computer running Windows 7 (64-bit; Microsoft Corporation, Redmond, WA, USA). All recordings were done on an air table (Micro-G, Technical Manufacturing Corporation, Woburn, MA, USA) with a custom-built wire-mesh Faraday cage.

Visual stimuli were presented using a conventional 37 by 22cm LCD computer monitor (ViewPanel VE150m, ViewSonic, Walnut, CA, USA) at a refresh rate of 60Hz and resolution of 1920x960 pixels and mean luminance of 53 cd/m². Animals were positioned 22.8 cm from the screen, which resulted in

stimulus check sizes that were approximately 2.5 degrees. Note that stimulus parameters were not optimized to the tuning of the recorded neurons, due to the limited stability of the extracellular recordings (typically about 30 min). Stimuli were presented using a custom-made video player, which I designed (E.I.N.). The program presents the stimuli and synchronizes the recordings.

Once acquired, single units were isolated using a customized version of WaveClus [61]. WaveClus processing entailed a bandpass filter (300-6000 Hz), thresholding to detect candidate spikes, decomposition of each candidate spike into eight Haar wavelet features, and clustering of spike events based on the wavelet coefficients for each. For each recording, sorting was carried out using different amplitude thresholds and cluster partitioning until we were confident that single units were isolated using similar criteria as in the macaque.

3.5.2 Motion Stimuli (both species)

The motion stimuli consisted of a temporal sequence of "motion blocks." Each motion block consisted of a segment containing a particular motion signal subtype in one direction (1500 ms duration, containing 15 frames of 100 ms each, followed by 500 ms of gray (50%)), followed by a similar segment containing the same motion signal subtype in the opposite direction. The opposite-direction segments paired in each motion block were presented in pseudorandom order. Note that the stimuli by nature are sampled in time so we used the same sampling rate for both species (60Hz).

We analyzed responses to 8 kinds of motion blocks that contain Fourier, glider, and non-Fourier signals (see Figure 3.1), along with 5 other kinds of

motion blocks that served as controls (These control motion blocks are based on three- and four-point gliders that did not yield strong percepts of motion [3] , and did not yield direction-selective neural responses). For each kind of motion block, movies were generated with pseudorandom seeds and contained, on average, 50% black checks and 50% white checks, according to the procedure of Hu & Victor [3]. These motion blocks were each repeated 32 times, totaling 26 minutes. In the macaque, this sequence was repeated up to four times. Each segment in each motion block was constructed with a different random seed. Example movies of all motion stimuli can be found in supplemental video S1.

Note that all stimuli (Fourier, glider, non-Fourier, and the controls) were defined by "templates" (spatiotemporal correlations) confined to a 2x2x2 spatiotemporal volume of checks (Figure 3.5), so that they had comparable spatial and temporal extent. As a consequence, the non-Fourier motion stimuli used here consist of motion of an edge that is parallel to the motion direction. Other studies of non-Fourier motion in mammalian V1 and V2 [49, 50] used non-Fourier stimuli defined by contours that were orthogonal to the motion direction, and this may account for the higher fraction of neurons that were found in those studies to be sensitive to non-Fourier motion.

Kind of Motion	Template	Subtype	Number of black checks	Number of white checks
2-point: Fourier		Standard motion	0 or 2	0 or 2
		Reverse phi	1	1
example: standard				
3-point: Glider		Black triangles	1 or 3	0 or 2
		White triangles	0 or 2	1 or 3
	contraction	Black triangles	1 or 3	0 or 2
		White triangles	0 or 2	1 or 3
example: expansion; black triangles				
4-point: Non-Fourier		Positive correlations	0 or 2 or 4	0 or 2 or 4
		Negative correlations	1 or 3	1 or 3
example: positive				

Figure 3.5: Summary of motion stimuli.

Each kind of motion (first column) corresponds to a “template” (second column) containing two, three, or four points in space and time. For each kind of motion, there are two or four subtypes (third column), which are defined by parity rules (last two columns), and, for glider motion, the orientation of the template. One subtype is illustrated in detail for each kind of motion; movies of all subtypes can be found in supplemental video S1. Fourier motion (top panel): the template consists of two checks on a space-time diagonal. In the illustrated “standard motion” subtype, there are always 0 or 2 black checks within the template, so the colors of the two checks along the diagonal must match. In the “reverse-phi” subtype, the number of black checks within the template is always 1, so the colors of the checks alternate along a diagonal. Glider motion (middle panel): the template contains three checks in a space-time triangle. For “expansion” subtypes, the triangle is oriented so that it expands as time progresses; for “contraction” subtypes, the template orientation is reversed. In the illustrated “black expansion” subtype, there are always 1 or 3 black checks within the template. This generates a stimulus containing expanding black regions. Non-Fourier motion (bottom panel): the template contains four checks in a space-time parallelogram. In the illustrated “positive correlation” subtype, the number of black checks within the template is always 0, 2, or 4. This generates a stimulus in which an edge between the two checks in the template at time t is always followed by an edge between the other two checks of the template at time $t+1$. In the “negative correlation” subtype, the number of black checks in the template must be 1 or 3. This generates a stimulus in which edges within the template are present either at time t or $t+1$, but not both.

3.5.3 Data analysis

To determine whether a neuron had a directionally-selective response for each motion subtype (e.g., white glider expansion), we compared the total number of spikes that occurred between 50ms and 1600ms following the presentation of the motion movie in the two directions within each motion block (paired t-test, criterion $p = 0.05$).

A neuron was considered to have a directionally-selective response to a kind of motion (Figure 3.2) if it had a directionally-selective response to any of its subtypes: for Fourier motion, the subtypes consist of standard and reverse-phi motion; for glider, the subtypes consist of expansion and contraction, each black or white; for non-Fourier motion, the subtypes consist of positive and negative correlation.

For each motion subtype, direction selectivity was quantified by a direction selectivity index (DI), given by equation 1:

$$DI = (M_{pref} - M_{non-pref}) / (M_{pref} + M_{non-pref}).$$

Directions were labeled "preferred" and "non-preferred" based on their responses to Fourier (standard) motion. Thus, for a neuron that responded to another subtype with the opposite direction preference compared to its response to Fourier motion, the DI was negative.

3.6 Supplemental Information

3.6.1 Additional analyses

3.6.2 Alternative definition of responsiveness to each kind of motion.

In figure 3.2 of the main text, we defined a neuron to be sensitive to a given kind of motion if it was had a significant ($p < 0.05$) directionally selective response to any of its subtypes (Figure 3.5). Figure 3.6 uses an alternative definition that includes a Bonferroni correction to account for the multiple subtypes and shows that the same result holds: the pattern of motion sensitivity is very similar between macaque and dragonfly. In this analysis, the significance cutoff for Fourier motion (two subtypes) is $p < 0.025$, for glider motion (four subtypes) the cutoff is $p < 0.0125$, and for non-Fourier motion (two subtypes) the cutoff is $p < 0.025$. This correction is conservative, since it assumes that responses to motion signals within a type are uncorrelated, when in fact there is evident for correlation (see Figure 3.3).

3.6.3 Analysis of combination of motion signals, subdivided by brain area.

Figure 3.3 in the main text showed that macaque and dragonfly share the same pattern of combination of motion signals at the neuronal level for example, neu-

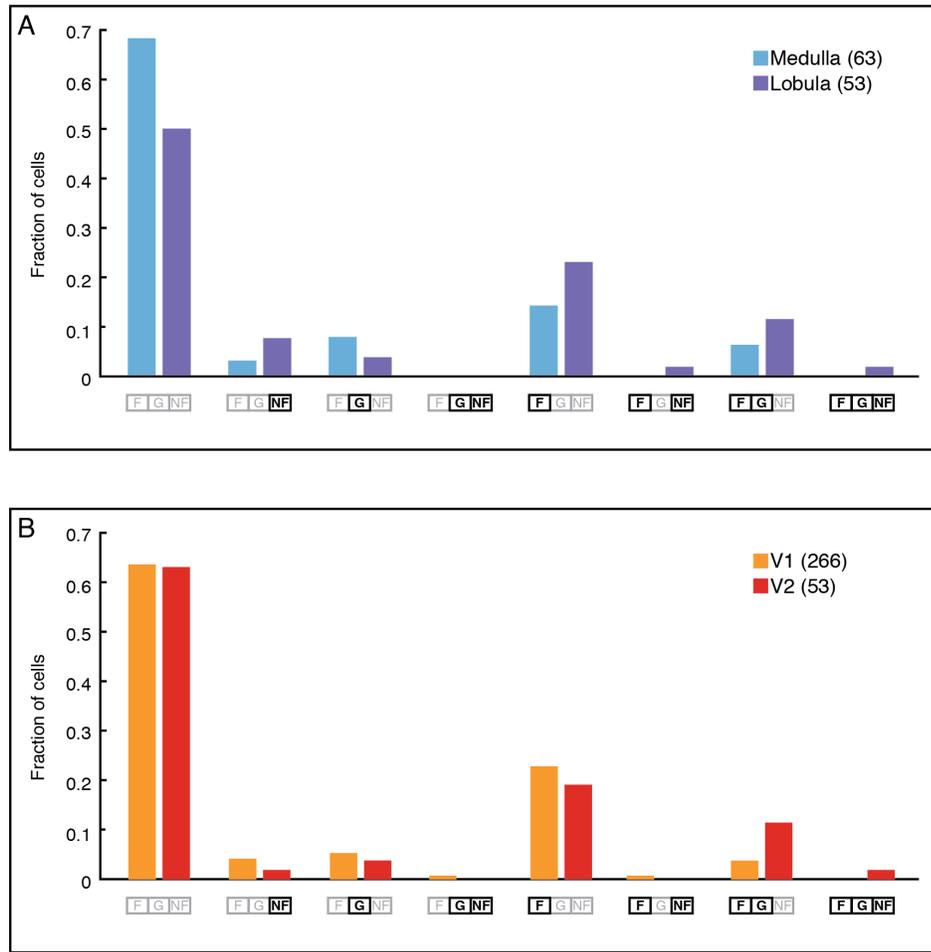


Figure 3.6: Fraction of neurons with a significant response to one or more kinds of motion for dragonflies (A) and macaques (B).

Significance levels include Bonferroni corrections. Other details as in Figure 3.4 and the main text.

rons that had strong responses to Fourier motion in one direction had tended to have responses with the opposite direction preference to white glider expansion. Here we show that this also holds when individual brain regions are compared. Figures 3.7 and 3.8 subdivide, respectively, the macaque data into V1 and V2, and the dragonfly data into medulla and lobula (restricting the analysis only to those recordings in which the recording site was unambiguous). Figures 3.9 and 3.10 compare macaque V1 to dragonfly medulla, and macaque V2

to dragonfly lobula.

As mentioned in the main text, the level of sensitivity to Fourier motion is different across species, with lower levels seen in dragonfly neurons than in the macaque. This difference may be due to the fact that for the macaque, stimuli were optimized for velocity (by adjusting spatial frequency), as well as for preferred orientation and contrast, but no optimization was done for the dragonfly (see SM 3.5.1, above). Movie frames were updated at 10 Hz in both species, and this may also contribute to a lower level of responsiveness in dragonflies, as they are known to capture changes in moving images at more than 200Hz framerate [62].

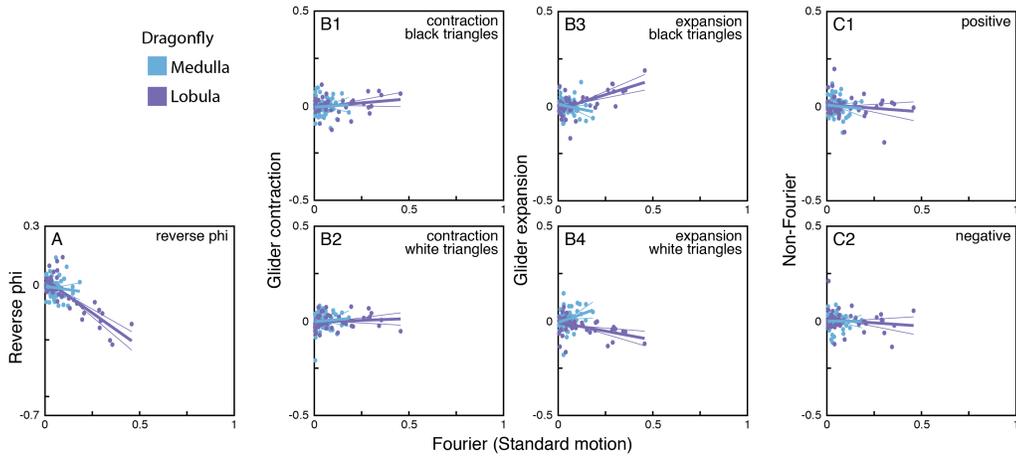


Figure 3.7: Comparison of direction selectivity index (DI) for Fourier (standard) motion, and each of the other subtypes, for macaque neurons localized to either V1 or V2.

(A) Negatively-correlated Fourier motion ("reverse-phi"), (B1) black glider contraction, (B2) black glider expansion, (B3) white glider contraction, (B4) white glider expansion, (C1) positively-correlated non-Fourier motion, and (C2), negatively-correlated non-Fourier motion. Heavy lines indicate the regression line through the origin, thin lines are 95% confidence limits on the slopes. Negative values of the DI indicate a direction preference that is opposite to the direction preference for Fourier motion. Lines are drawn to the limits of the data along the abscissa.

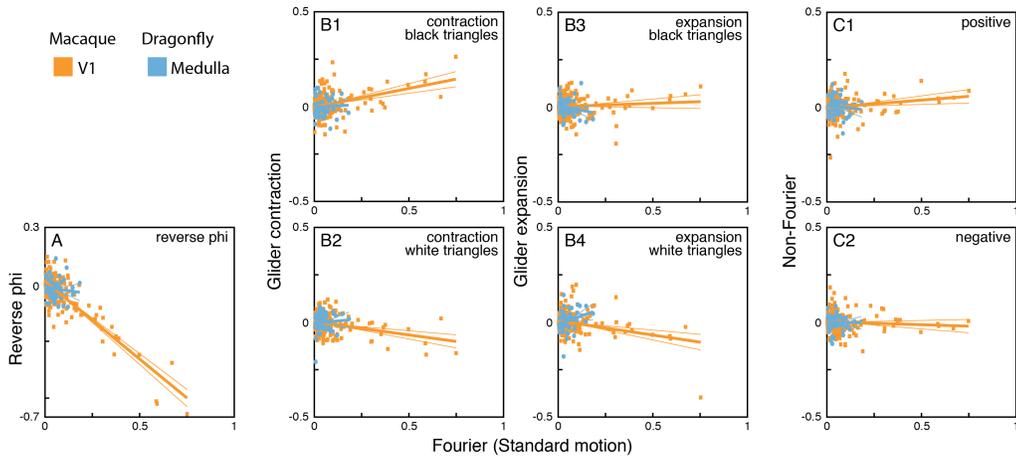


Figure 3.8: Comparison of direction selectivity index (DI) for Fourier (standard) motion, and each of the other subtypes, for dragonfly neurons localized to either medulla or lobulla.

For other details see Figure 3.7

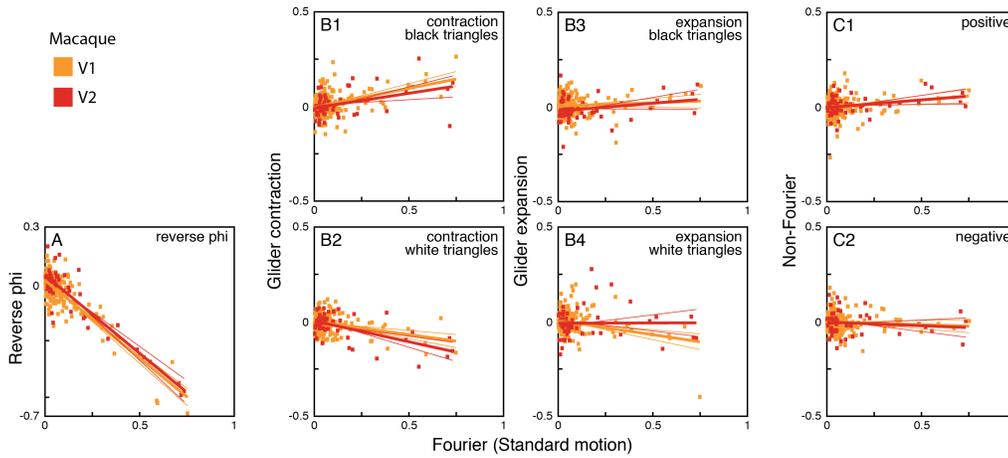


Figure 3.9: Comparison of direction selectivity index (DI) for Fourier (standard) motion, and each of the other subtypes, for dragonfly medulla and macaque V1 neurons.

For other details see Figure 3.7

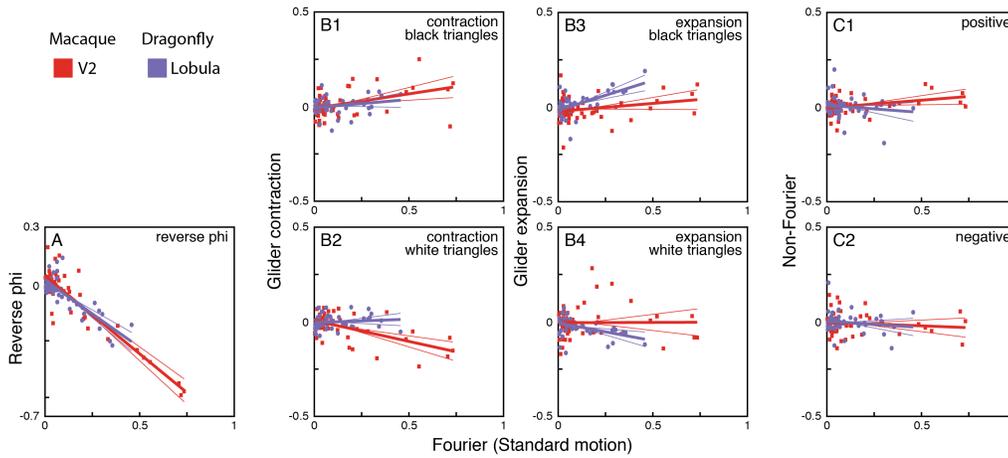


Figure 3.10: Comparison of direction selectivity index (DI) for Fourier (standard) motion, and each of the other subtypes, for dragonfly lobula and macaque V2 neurons.

For more details see Figure 3.7

3.6.4 Video. Demonstration of the motion stimuli used in these experiments.

Short example clips show the following, with each shown moving to the right, then to the left:

1. Fourier standard motion
2. Glider black expansion
3. Glider black contraction
4. Glider white expansion
5. Glider white contraction
6. Non-Fourier positive correlation
7. Non-Fourier negative correlation
8. Random

CHAPTER 4

PERCEPTUAL INTERACTION OF LOCAL MOTION SIGNALS

4.1 Introduction

Motion is crucial for everyday tasks, such as navigation [29] and figure ground segregation [15]. Motion analysis is generally considered to begin with the extraction of local motion signals, and following this step, motion signals are analyzed across space. Many kinds of local motion signals are well recognized, including Fourier (F), non-Fourier (NF), and glider (G) (see below). These can be distinguished by the nature of the local computation required to extract them. It is straightforward to construct stimuli that isolate each of these kinds of signals, enabling experimental analysis of visual responses to each. Yet natural scenes contain all of these signals showed [8], and they typically co-occur in the same location. Moreover, theoretical studies indicate that motion extraction can be made more efficient through the combined use of multiple local motion cues [13]. Thus, it is of interest to determine how local motion signals are jointly processed; this is the focus of this chapter.

The archetypal local motion signal consists of spatiotemporal correlation between pairs of points [1], i.e., two points along a diagonal in space-time. This is often also called Fourier motion, as two-point correlations can be determined from the power spectrum. However, motion can also be signaled by multipoint correlations in any slanted region in space-time. These higher-order motion signals include classic "non-Fourier" motion, which involves four points in a space-time parallelogram [2], and "glider" motion [3, 8], which involves three points in a space-time triangle.

While several studies have examined the relationship between different kinds of local motion signals, they have focused on F and NF signals, and primarily on whether these two kinds of motion were processed by one vs. two systems [17]. Here, we are specifically interested in a different question: how distinct kinds of local motion signals (here, F and G) interact.

A wide range of techniques are available to study processing of different kinds of motions [40, 49, 63–73], including neurophysiological experiments [49, 66, 67], studies of patients with neurologic disorders [64, 68–71], imaging studies [40], and psychophysical experiments [63, 65, 68, 72, 73], with the latter approach either based on comparing the tuning of responses to different motion types [65, 68], or using adaptation paradigms [63, 73]. In this study, we use a psychophysical strategy that explicitly combines F motion signals with several variants of G motion signals in a controlled fashion. This enables us to look directly at how F and G signals are integrated, and whether the responses to F signals depend on context.

As we show, F and G signals are integrated prior to perceptual threshold, but this only occurs in one context, and in a way that suggests a light/dark asymmetry. In a different context, integration was not found, and moreover, sensitivity to F motion was significantly reduced.

4.2 Methods

Subjects performed a two-alternative forced-choice task, indicating their judgment of the direction of motion of a 1-second movie segment via a button-press. Movie segments consisted of 10 frames, 100 ms each, each composed of a 16 x

16 array of 26-min black and white checks (display size, 7×7 deg), with a fixation point superimposed on the center of the display prior to the onset of the movie. Viewing distance was approximately 30 cm. Trials were self-paced, and no feedback was given. Data were collected after a small number of practice trials to ensure that the subject understood the task.

4.2.1 Visual stimuli

Movies were specified by the strength of the Fourier (F) motion signal, C_F , and the strength of the glider (G) signal, C_G . Either signal could range from 0 (absent) to 1 (maximal). When both signals were present, the motions they defined were in the same direction (randomly left or right).

As mentioned above, both kinds of local motions are defined by correlations in slanted spatiotemporal regions, which we designate as their "templates" [8]. For F motion, the template consists of two checks in a space-time diagonal; for glider motion, the template consists of three checks in a space-time triangle (see table 1). Since the checks contain either black or white checks, their correlations are defined by parity rules – whether the number of black (or white) checks is an even or an odd number. Thus, for F motion, full strength ($C_F=1$) corresponds to all templates containing matching checks (either 2 white checks or 2 black checks; see table 4.1 and figure 4.1A); at zero strength ($C_F=0$), half of the templates contain matching checks and half contain one check of each color. For G motion, full strength ($C_G=1$) stimuli can be of two polarities – "black triangles", in which all triangles have either 3 black checks or 1 black check, and "white triangles", in which all triangles have either 3 white checks or 1 white

Kind	Subtype	Template	Subtype	# black	# white
2-point (Fourier)			Standard	0 or 2	0 or 2
			Reverse phi	1	1
3-point (Glider)	Contraction		White exp.	0 or 2	1 or 3
			Black exp.	1 or 3	0 or 2
	Expansion		White cont.	0 or 2	1 or 3
			Black cont.	1 or 3	0 or 2

Table 4.1: Examined motion types.

The two kinds of motion signals examined here. Fourier (F) motion is characterized by two-point correlations along a spatiotemporal diagonal, and glider (G) motion is characterized by three-point motion in a spatiotemporal triangle. G motion has contraction and expansion subtypes, depending on the orientation of the triangle in space-time. Motion signals are further subdivided by the sign of the correlation. For binary stimuli (used here), this corresponds to whether the number of black checks within the correlation template is an even or an odd number.

check (see table 4.1 and figure 4.1A). Additionally, the glider template can have two different orientations in space-time – expanding or contracting. For further information about glider motion we refer the interested reader to Hu & Victor [3] and Nitzany & Victor [8].

To make a stimulus that combines a F motion signal of strength C_F with a G motion signal of strength C_G , we adapted the maximum-entropy texture generation algorithms of Victor & Conte [74] to spatiotemporal stimuli (see below). The maximum-entropy property means that the movies are as random as possible, given the two specified component motion signal strengths. That is, they contain no other cues to motion other than what is implied by the component motion signals.

To adapt texture generation algorithms to synthesize movies, we interpret the vertical spatial dimension of a texture as time. Thus, the XY -plane of a tex-

ture becomes an XT -slice of a movie (where X is the horizontal axis). With this replacement, the second-order image statistics $\beta_{/}$ and β_{\setminus} capture the pairwise correlations along diagonals in space-time (see Victor and Conte [74] for background on the visual textures). Thus, to generate a F motion signal to the left or right, we set one of these β 's equal to C_F . The third-order image statistics θ_{\uparrow} , θ_{\downarrow} , θ_{\lrcorner} and θ_{\llcorner} capture three-point correlations within a triangular glider template. Thus, to generate a G motion signal, we set one of the four θ 's equal to $\pm C_G$. $\theta = -C_G$ generates a preponderance of templates with an odd number of black checks (black gliders), while $\theta = +C_G$ generates a preponderance of templates with an odd number of white checks (white gliders). The selection among the four θ 's is determined by whether the glider subtype is expansion or contraction, and whether the motion signal is to the left or the right. Specifically, θ_{\uparrow} and θ_{\downarrow} correspond to glider contraction, while θ_{\lrcorner} and θ_{\llcorner} correspond to glider expansion; θ_{\uparrow} and θ_{\downarrow} are leftward motions, while θ_{\lrcorner} and θ_{\llcorner} are rightward motions.

The algorithms of Victor and Conte [74] include procedures for making maximum-entropy textures for any pairwise combination of values of the β 's and θ 's, provided that the total strength does not exceed 1. These algorithms translate directly into making maximum-entropy movies containing mixtures of F and G motion signals. We used this approach, along with the correspondence above, to generate 16 independent XT slices for each movie, and then stacked them together to make a single XYT movie.

In this manner, F and G cue strengths could be varied independently, and thus constituted coordinates a planar "combination space" of motion signals. We carried out experiments in two such planes: F combined with G expansion, and F combined with G contraction. Since pure F stimuli were common

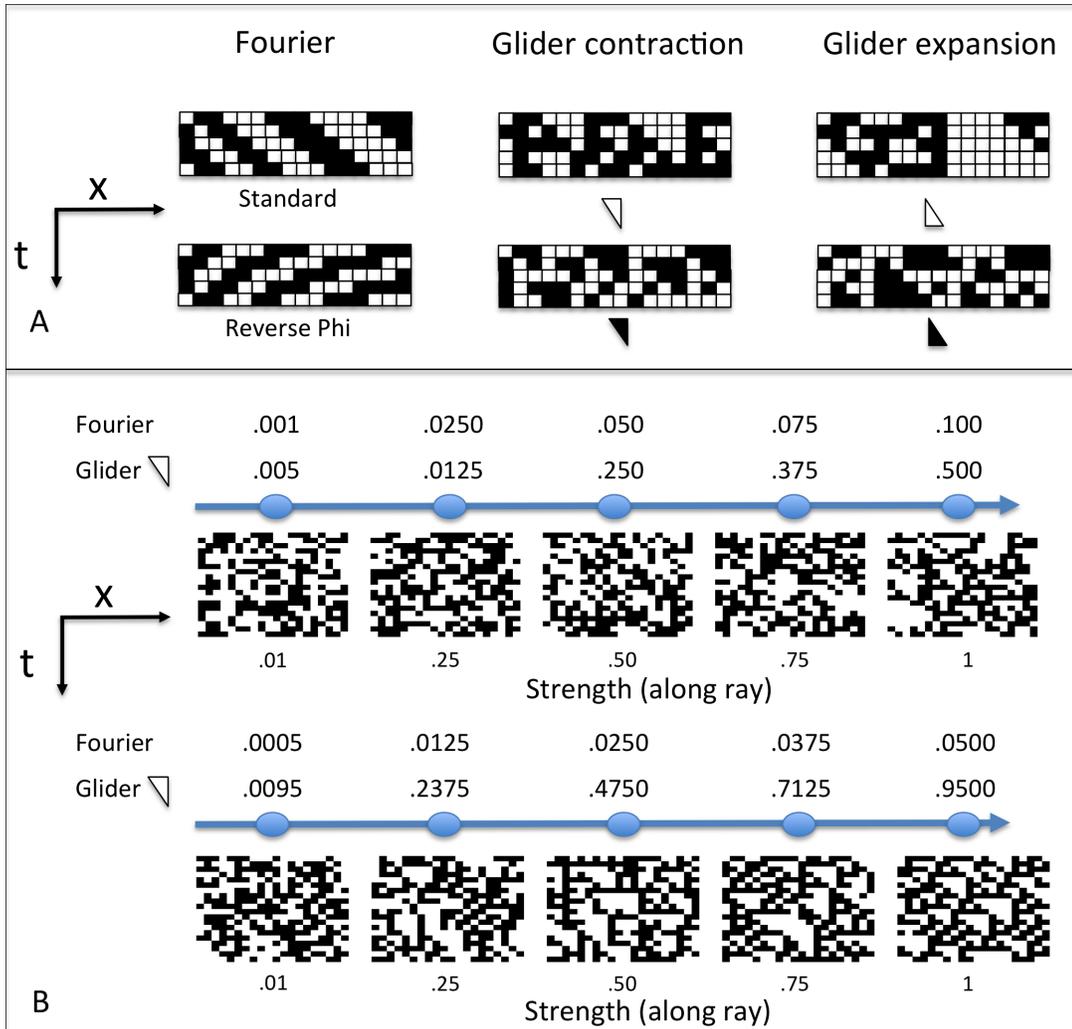


Figure 4.1: Stimulus examples

Panel A: space-time slices of stimuli containing Fourier (left column), glider (G) contraction (middle column) and G expansion (right column) signals at maximum correlation strength. Top row: even parity, corresponding to standard Fourier (F) motion, white G contraction, and white G expansion (see Table 4.1). Bottom row: odd parity, corresponding to reverse-phi F motion, black G contraction, and black G expansion. **Panel B:** space-time slices of stimuli containing mixtures of F and white G contraction, in the proportions used in these experiments. Top row: five example stimuli along a ray ending with $C_F=0.1$ and $C_G=0.5$. Bottom row: five example stimuli along a ray ending with $C_F=0.05$ and $C_G=0.95$.

to both experiments, this design enabled us to study two kinds of interactions: direct interactions between F and G cues that were simultaneously presented, and modulatory interactions, i.e., changes in sensitivity to F cues depending on whether they were in the context of G contraction, vs. G expansion.

Within each combination space, experiments were organized as follows. Test points were located along 7 rays emanating from the origin of the combination space. Along each ray, 5 points were examined. Specifically, each ray was defined by a maximum motion strength of each signal type $(C_{F,max}, C_{G,max})$, and the stimuli along each ray were defined by $(C_F, C_G) = R_i * (C_{F,max}, C_{G,max})$, where $R_i = 0.01, 0.25, 0.5, 0.75, 1$. The above stimuli were generated via software written in Matlab (2010a), which also recorded the subjects' responses. Figure 4.1B shows examples of combined stimuli along two rays.

4.2.2 Data analysis

Data from the two kinds of interactions studied (F combined with G expansion, and F combined with G contraction) were analyzed separately. In each of these planes, the first step was to fit the measured fraction correct with a Weibull function along each ray (as in Victor et. al. [75], using the maximum likelihood approach [75]):

$$FC(x) = \frac{1}{2} + \frac{1}{2}(1 - 2^{-(\frac{x}{a_r})^{b_r}}) \quad (4.1)$$

In this equation, x is the distance along the ray from the origin, given by $x = \sqrt{C_F^2 + C_G^2}$, where C_F and C_G are the individual motion strengths; b_r is

the Weibull shape parameter; and a_r is the motion strength at which the fraction correct is 0.75. The value of the Weibull shape parameter (b_r) was found to be in the range of 0.85 to 2.96, across all ray, interaction planes, and subjects. We then fit each dataset (all the rays within a single interaction plane for a single subject) with a uniform shape parameter value b , allowing the threshold parameter a_r to vary across rays. This yielded consensus values of b in the range 1.5 to 2.3 (across subjects and glider expansion vs. contraction). The fitted value of a_r for each ray was then taken to be the threshold along that ray. 95% confidence intervals were determined from the empiric distribution of 1000 bootstrapped samples.

4.2.3 Subjects and display

Subjects (4 male, 1 female) had either normal or corrected-to-normal vision, and ranged in age from 20 to 38. Subjects EIN was an author; subject PSS was not an author but aware of the purpose of the experiments. For EIN, the display consisted of a 17-inch LED monitor; for PSS, the display was a 15-inch Retina display on a Macbook Pro. For AB and TS, the display was a 17-inch Retina display on a Macbook Pro. Human subject procedures were approved by the Institutional Review Committee of Weill Cornell Medical College and by Institutional Review Committee of University of Chicago. Each subject was studied in four 1-hour sessions. Within each session, approximately 3,500 trials were presented.

4.3 Results

To determine how Fourier (F) and glider (G) motion cues are combined perceptually, we measured the ability to determine direction of motion in a stimulus containing both kinds of cues at a range of levels. Figure 4.2 shows how performance depended on F and G strength for G contraction (top row), and G expansion (bottom row). The most obvious feature is that when motion cues were presented in isolation, there was a much greater sensitivity to F cues than to G cues: when only the F cue present (along the abscissa), a strength $C_F=0.1$ typically sufficed for a fraction correct of 0.75, but when only the G cue was present (along the ordinate), a strength of $C_G=0.5$ was required to achieve the same level of performance.

For combinations of F and G contraction (top row), the consistently curved contour lines showed that F and G contraction signals are integrated – a given level of performance can be achieved by a combination of signals, even if neither signal by itself would have sufficed. If, alternatively, there were no integration of the motion signals, then the contours would have been rectilinear, because a criterion performance level would only be reached in the combination stimulus when one of its components meets threshold. The elliptical shape of these contours suggests that the signals are combined in an approximately quadratic fashion, with F signals given a stronger weight.

There is a suggestion of an asymmetry between white and black G contraction. For combined stimuli that differ in the polarity of the G contraction signal but not their strengths (open symbols superimposed on contour maps) performance, as indicated by the contour lines, is generally higher for white G

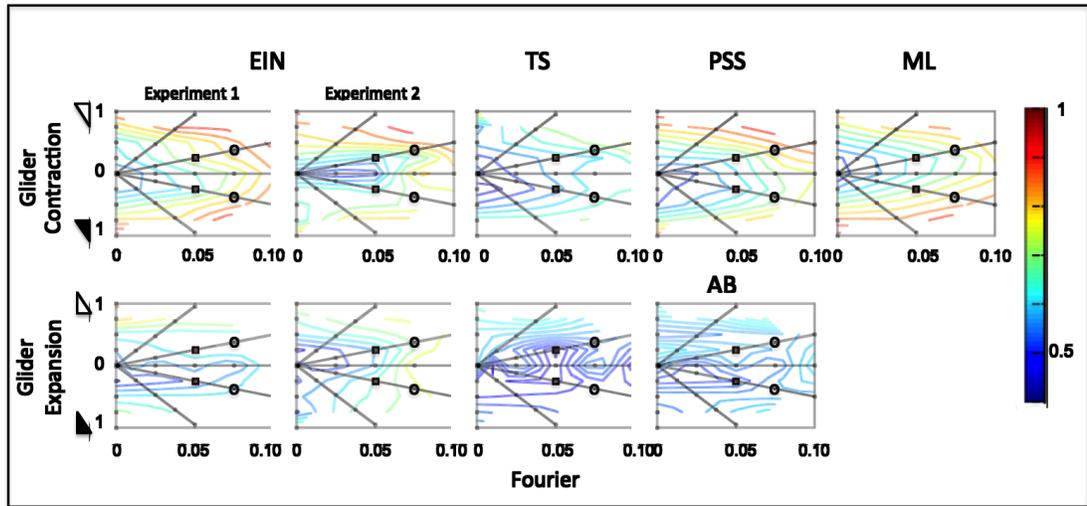


Figure 4.2: Psychophysical performance for direction judgments for stimuli containing mixtures of Fourier (F) and glider (G) signals.

Contour maps show fraction correct as a function of C_F (abscissa) and C_G (ordinate). Upper quadrant shows responses for stimuli containing white Gs; lower quadrant shows responses for stimuli containing black Gs. The abscissa corresponds to pure F stimuli. Top row: mixtures of F and G contraction. Bottom row: mixtures of F and G expansion. The lines indicate the rays that were studied, and the points on the rays the specific signal combinations. Black circles and squares mark examples of stimulus pairs containing mixtures matched for the strengths of F and G motion, but differ in the parity of G contraction (white vs. black).

contraction, than black G contraction.

For combinations of F signals with G expansion, the pattern of interaction is different. Only one subject (EIN in Experiment 1) shows approximately elliptical contours indicating integration of F and G expansion. For the others (including one subject, TS, who had elliptical contours for G contraction), contour lines are approximately horizontal, showing that performance depended primarily on the G expansion component. Note that the stimuli on the abscissa (the pure F stimuli) were identical to those used in the G contraction experiment, and elicited strong motion percepts in that context – but here, in the context of stimuli containing G expansion, sensitivity to F appeared to be reduced. For

subjects TS and AB, low levels of black G expansion led to performance that was less than chance – they perceived the stimulus as moving in the direction opposite to its true motion – a finding reported in Hu & Victor [3], who only tested this signal type at a strength of 1.

To quantify these observations, we fit the psychophysical data along each ray to a Weibull function (see Methods). For G contraction, this provided a reasonable fit to the data, as shown in Figure 4.3A. Thus, psychophysical performance for G contraction in the entire C_F, C_G domain can be summarized by the Weibull threshold parameter (a_r) along each ray, with a consensus value of the Weibull shape parameter b (here, 1.65) used for all rays.

Figure 4.4 shows this summary of thresholds for all four subjects, and confirms the observations made from figure 4.2. There is integration between the two types of motion signals, as manifest by the curved trajectory of the performance threshold. Four of the five datasets show an asymmetry in the way that F signals combine with white vs. black G contraction: for stimuli combining F motion with equal strengths of G contraction (for example, the pairs of similarly marked points), thresholds are lower for white G contraction than black G contraction.

For the experiment examining combinations of F with G expansion, Weibull fits (e.g. see figure 4.3B) could not capture qualitative aspects of the performance, since (as mentioned above) two subjects perceived motion opposite to the true direction for low levels of black G expansion. However, the Weibull fits along the F ray allowed for a direct comparison to the parallel condition in the G contraction experiment (Figure 4.5). In 2 subjects (EIN experiment 1 and TS) sensitivity to F in the context of the G expansion experiment was signifi-

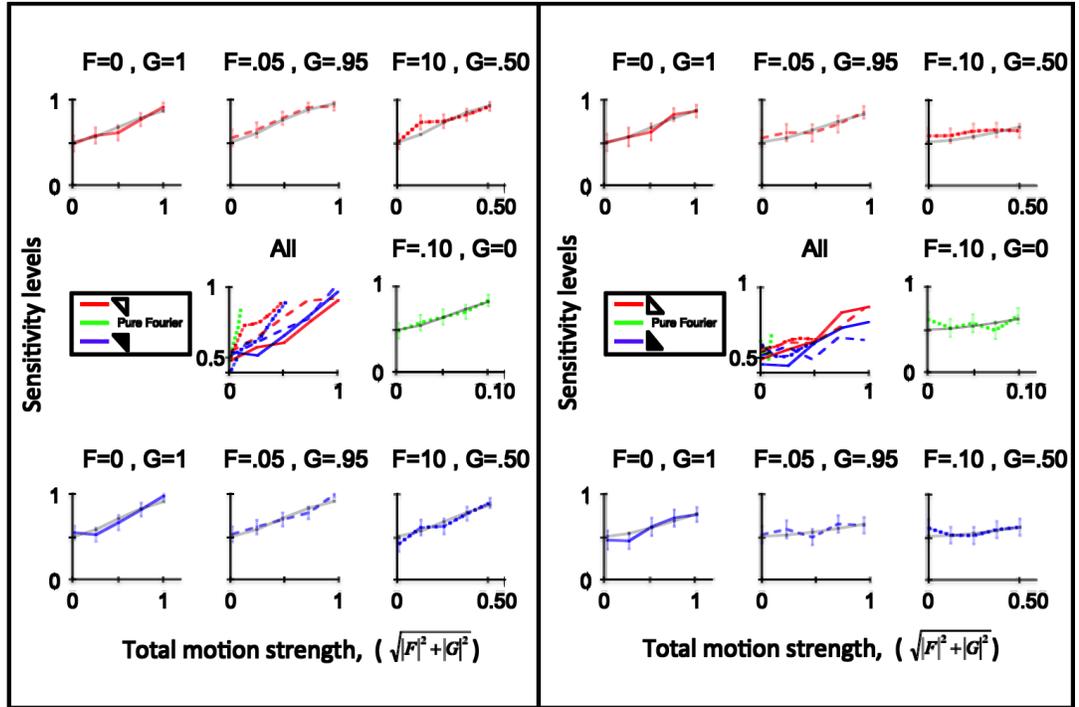


Figure 4.3: Fits of Weibull functions to performance along each ray.

Panel A: Fourier (F) and glider (G) contraction; **Panel B:** F and G expansion. In each panel, the seven individual plots compare the measured fraction correct along a single ray with a Weibull function, with the abscissa scaled to the maximal total motion strength on that ray $\sqrt{|C_F|^2 + |C_G|^2}$; error bars indicate 95% confidence intervals determined by bootstrap. Data are shown with a line color indicating glider parity (red: white G; blue: black G; green: pure F (no G)) and line style indicating the maximal motion strength at the end of the ray (solid: $C_F=0, C_G=1$; dashed: $C_F=0.05, C_G=0.95$; dot-dash: $C_F=0.1, C_G=0.5$; dotted: $C_F=0.1, C_G=0$). Black solid lines show the Weibull fit. The central plot shows all of the fits superimposed, scaled to a motion signal strength of 1. In each panel, Weibull functions had the same shape parameter b : 1.65 in Panel A, 1.70 in panel B. Subject EIN, experiment 1.

cantly decreased compared to sensitivity to F in the G contraction experiment.

Results from the 3 other subjects (PSS, ML and AB) support the notion that sensitivity to F is lower in the context of G expansion than in the context of G contraction, although each of these subjects were only tested under one condition, so no within-subject comparison could be made. We note that EIN experiment 2 did not show that trend, but this experiment included fewer number of repeats.

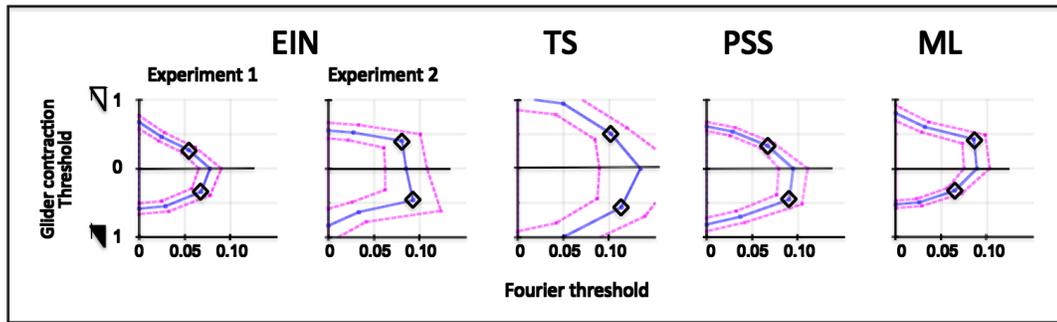


Figure 4.4: Isodiscrimination curves for combinations of Fourier (F) and glider (G) contraction.

Curves connect the distances a_r along each ray at which a fraction correct of 0.75 is reached (see methods). Ordinate shows G contraction strength, upper quadrant for white triangles and lower quadrant for black triangles. Diamonds mark points along rays that have the same proportions of F and G motion, but differ in the parity of G contraction (white vs. black).

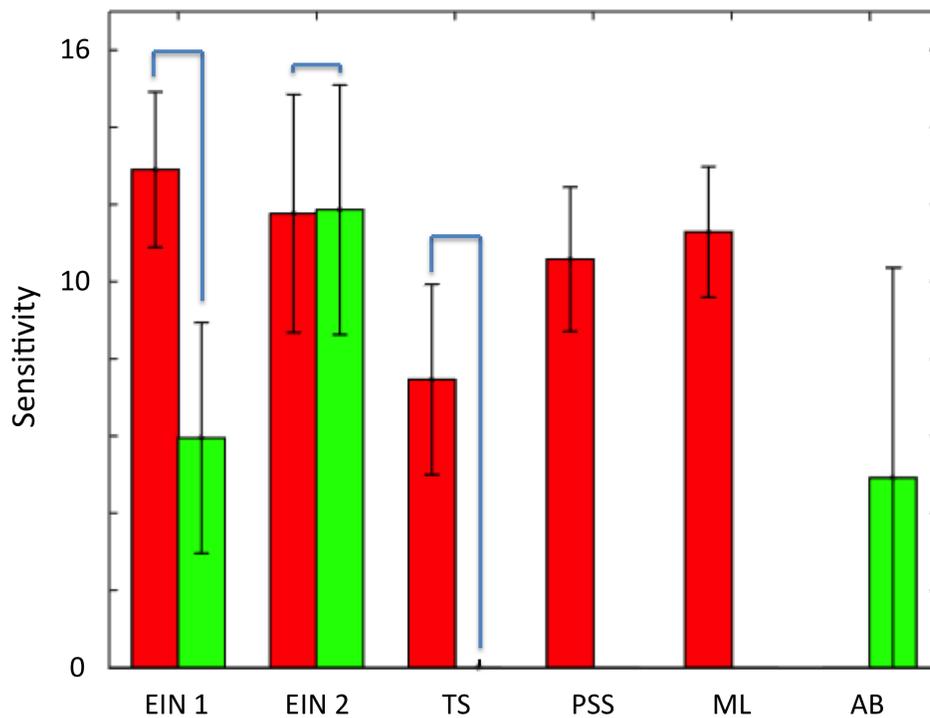


Figure 4.5: Comparison of sensitivity to Fourier (F) motion in the context of glider (G) contraction (red) and expansion (green).

Error bars indicate 95% confidence intervals. Braces connect data from subjects run in both kinds of experiments. Note that for subject TS, there was no measurable sensitivity to F motion in the context of G expansion.

4.4 Discussion

Local motion signals can be distinguished by mathematical characterization [1, 2, 17] of the underlying spatiotemporal correlations. This approach identifies several kinds of motion elements (Fourier, non-Fourier, and glider) that are mathematically independent, and are also independent in an operational sense: each kind of motion signal can be isolated experimentally, by creating artificial stimuli that drive it, and none of the others. These different kinds of motion signals have been shown to elicit behavioral changes responses in humans [3, 48] and other species (Drosophila: [7] ; zebrafish: [6]; dragonfly: [5]; Macaque: [5, 49]).

However, this subdivision of motion signal types is arguably un-natural, in that they rarely occur in isolation in natural sensory inputs. Rather, as we recently showed [8], several kinds of local motion signals occur together in naturalistic inputs, and typically in overlapping locations. While different naturalistic inputs have similar mixtures of motion signal types, and the co-occurrence of these signal types is correlated on a scene-by-scene basis, they are not redundant: the presence of one motion signal only predicts the strength of another within about a factor of two.

Fitzgerald and colleagues [13] developed the theoretical framework for the utility of these motion signals for biological motion processing. However, it is unclear whether the visual system takes advantage of the co-occurrence of those signals by integrating them. Moreover, since the complements of these signals depend on the cause of the motion (e.g. translation leads to primarily Fourier (F) motion, while looming of objects generates prominent glider (G)

expansion signals and receding of objects generates prominent G contraction signals), biological motion processing may be context-dependent.

To address these questions, we asked subjects to report the perceived direction of synthetic motion signals that contained controlled mixtures of F and G signals. Experiments were carried out in blocks - one consisting of mixtures of F with G contraction, one consisting of mixtures of F with G expansion. This enabled us to examine integration of motion signals that were simultaneously present, as well as context-dependence of processing of F motion.

We found that integration occurred prior to the decision stage when F and G signals were simultaneously present (N=4); this is the behavior of computational models that make optimal use of two kinds of signals that are non-redundant [13]. But surprisingly, this integration only occurred for G contraction. For G expansion, we observed a context effect (N=3, including two studied with G contraction): in blocks in which G expansion was present, sensitivity to F signals was reduced.

What possible function does this modulatory interaction serve? We speculate that it arises because biological systems adapt to their environments by prioritizing information according to its importance. A successful system is expected to focus on signals that serve a task, but also to place a premium on those that have greater implications for survival. Here, if the visual system has prior knowledge that one kind of signal, G expansion, is associated with looming objects, it makes sense to focus on this kind of signal, even if the cost includes temporarily ignoring other motion signals, e.g. F. This shift in focus may be a consequence of limited resources, but even if capacity is unlimited, the less-important kind of signal may be discounted at the decision level.

Our findings are in line with those of Hu et. al. [3], but the correspondence is not complete - and we suspect that this difference reflects the contextual modulation of motion processing. As in Hu et al. [3], found that sensitivity to G motion was substantially less than sensitivity to F motion, and that responses to G contraction were veridical - i.e., that inversion of the sign of contrast (black vs. white gliders) did not invert the perceived direction. The latter is noteworthy, since a strict opponent model would have predicted that such an inversion is present (see Table 1 of Hu et, al. [3]).

For G expansion, however, the findings differed somewhat: Hu & Victor [3] found that inversion of contrast led to a reversal of the perceived motion direction: white ("even" parity) G expansion elicited perceived motion in the veridical direction, while black G expansion elicited perceived motion opposite to the veridical direction. Here, we only found this inversion at low levels of G strength, and only in the two naïve subjects (TS and AB, Figure 4.2). We hypothesize that this difference may result from a difference in the experimental paradigm. Specifically, Hu et. al. [3] presented stimuli with many kinds of gliders at full correlation strength, but randomly interleaved on a trial-by-trial basis, so there was no G expansion or contraction "context". Here, only G contraction (or only G expansion) was presented, for blocks of up to 3500 trials, across sittings with a total duration of several hours, enabling a context to develop.

4.4.1 Summary

Using novel synthetic stimuli that contain controlled amounts of two kinds of local motion signals, we find two kinds of interactions at the perceptual level:

subthreshold integration, and context-dependent changes in sensitivity.

CHAPTER 5

DISCUSSION AND FUTURE WORK

5.1 Discussion and future directions

Extraction of local motion signals is critical task for our survival. It is an example of a process that serves to articulate abstract rules from many pieces of information, which are later used to infer expectations about diverse outcomes. This is a complex task that require a tremendous amount of resources for implementation. Nevertheless, many species seem to accomplish it effortlessly and successfully, and use motion information for critical tasks such as avoiding collision and separating figure from ground. Here I summarize the studies that I carried out concerning motion signals and the neural computations that extract them.

The thesis examined (i) the statistics of local motion signals in natural scenes, (ii) how neurons in the visual systems of different species respond to them, and (iii) how humans perceive combinations of different kinds of motion signals when presented simultaneously. The motivation for our studies was to fill in some missing pieces in the puzzle of how do we extract, perceive and use local motion cues.

Historically, researchers such as Riechardt, Chubb, Sperling, Adelson, Victor and many more [1–3, 14, 17, 48, 76, 77] have studied the various characteristics of local motion signals. Two approaches have been taken to define different kinds of motion signals: via their mathematical properties [1, 14, 19], or via exemplars [2, 3, 17, 48, 77]. We use the former approach here, as it leads to an

unambiguous and rich classification of motion signal types. The specific kinds of motion signals that we deal with here include Fourier [1], non-Fourier [2] and glider motion [3]. In general, it is possible to create synthetic stimuli that contain one kind of motion signal, and not any of the others. For instance, Chubb & Sperling [2] developed a "non-Fourier" stimuli, which does not include any first-order (i.e. Fourier) signals and Hu & Victor [3] designed another kind (i.e. glider), which does not contain either Fourier or non-Fourier signals. Using these synthetic stimuli, it was shown [2, 3] that humans are sensitive to each of these kinds of motion signals.

However, it was not clear whether these signals appear in nature, and if so, whether they were independent or correlated. We tackled this issue by conducting a rigorous analysis of naturalistic (Hollywood) movies, quantifying the prevalence of these several kinds of local motion signals and examining potential correlations between them (chapter 2). To ensure the robustness of our findings, we made use of two ways to quantify local motion signals and their spatial consistency (Pattern Match Opponent (PMO) and Rule Match Opponent (RMO)), a range of resolutions (1x1; to 16x16) and several ways of binarizing the movies. In brief, we found that all three kinds (Fourier, non-Fourier, and glider) of local motion cues appeared in the movies. While there were strong correlations between the co-occurrences of the different motion types, they were also partially independent: knowing the strength of one motion signal type in scene allowed one to estimate the strength of another, but only within about a factor of two.

Based on the mathematical properties of the different kinds of local motion signals, we speculate that specific scenarios preferentially generate one or

another signal type — for example, that glider motion is associated with looming, and non-Fourier motion is associated with motion transparency [19]. Thus, different kinds of motion signals might be selectively used in particular tasks, such as collision avoidance and object recognition. Further investigation to confirm this speculation is required. A logical next step would be to identify the real-world scenarios that generate specific kinds of local motion signals.

We further speculate that any system (not just a biological visual system) whose goals include the analysis of motion would benefit from using heuristics concerning the correlations of local motion signal types in natural scenes. Thus, we think that our findings may be of interest to the computer vision community and to the telecom industry, as they suggest ways to extract motion information, and a novel strategy for video compression algorithms.

Chapter 3 deals with the ability of neurons in visual systems to extract different kinds of local motion signals and is thus directly motivated by Chapter 2. It is known that these motion signals (Fourier, glider, and non-Fourier) elicit behavioral responses in a wide range of species (Drosophila: [7]; zebrafish: [6] and human: [3, 48]); for two of the elements (Fourier and non-Fourier), neurophysiologic correlates in the mammalian visual cortex are documented (macaque: [49]; cat: [50]). However, it is not clear whether individual neurons in the visual system respond to glider motion, nor whether neurons that respond to one kind of motion signal also respond to others. We tackled these issue, by presenting visual system neurons of two divergent species (i.e. macaque and dragonfly) with similar stimuli that isolated each kind of local motion signals. We found very similar responses from the two species. Specifically, a similar fraction of neurons responded to each motion signal type; neurons that responded to mul-

multiple motion signal types did so in a similar way; and the extent of motion-opponency diminished as processing unfolded from V1 to V2 in the macaque, or from medulla to lamina in the dragonfly. These findings indicate that at an algorithmic level, and at the level of its cellular implementation, both systems had converged to similar solution. This was especially striking in view of the gross differences between the scale and organization of the two species' visual systems. We conclude that the ethological demands and survival pressures drive biologic motion processing in these species to a highly convergent solution at the neuronal level, despite major differences in their phylogenies and the architectures of their eyes and brains.

On this background of similarities, there was one notable difference between the responses of the two species: for a subtype of glider motion (black expansion), neurons in the dragonfly tended to be more sensitive than in the macaque. This specific motion type is associated, among other things, with looming of dark objects on a light background. Thus it makes sense that a flying creature would be more sensitive to this motion kind, as it directly influences its ability to survive. More generally, this finding emphasizes the need to analyze a system within its context. Neurons, as part of the brain, must serve the requirements of the organismic system and therefore must be tuned to solve the tasks critical to its survival.

Our findings that in naturalistic stimuli, motion signal types do not occur in isolation [8] motivated us to explore the perception of stimuli that contained two kinds of local motion signals. Most earlier research either focused on a single kind of local motion signal (i.e. Fourier [1, 17], non-Fourier [2, 76], glider [3]), or a mixture of Fourier and non-Fourier signals. Here we follow the latter

approach and, as in the study of Ledgeway & Smith [78] for Fourier and non-Fourier signals, measure human sensitivity to stimuli containing mixtures of local motion cues.

Our experimental design enabled us to study two kinds of interactions: direct interactions between motion cues that were simultaneously presented, and modulatory interactions, i.e., changes in sensitivity to one kind of cue depending on whether it was in the context of another. We found that both kinds of interactions were present — direct interactions between glider contraction and Fourier, and a modulatory interaction between glider expansion and Fourier.

The manifestation of the first kind of interaction was that subjects were able to report motion direction of a stimulus containing weak Fourier and glider contraction cues, even though neither cue, when presented in isolation, led to a clear motion percept. This implies an interaction at a level of neural computation prior to perception.

The manifestation of the second kind of interaction was that sensitivity to Fourier motion was diminished when presented in the context of glider expansion. I interpret this as a hint about the importance of glider expansion signals. The visual system is expected to focus on signals that are associated with important outcomes. For example, if the visual system had found that glider expansion signals, which are associated with looming, tend to end with bad outcome (e.g. hitting walls), then it is expected to prioritize them, possibly ignoring less-ominous cues in when glider expansion cues are presently. We note, though, that these findings were obtained with stimuli that are highly artificial. Further investigation is required in to determine whether the phenomenon also occurs with more natural stimuli.

While this study demonstrates that a form of selective adaptation to motion signals exists at the perceptual level, it is not clear where in the brain this adaptation occurs. Since neurons in early visual cortices (e.g. V1 and V2) have selective sensitivity to local motion signals, adaptation may occur at this point. But the adaptation may also occur at later processing stages, and additional study is required to resolve this issue.

To summarize, here I describe a series of experiments that explore several aspects of local motion analysis. First, we showed that many different kinds of local motion signals appear in natural environments, in stereotyped proportions and with moderate correlations between them. Next, we showed that neurons in two very different visual systems (macaque and dragonfly) are sensitive to these motion cues, and the pattern of sensitivity indicated convergence both at the algorithmic and neural-implementation levels. Finally, we explored perceptual responses to combinations of local motion cues. Two kinds of interactions were found: a sub-threshold summation between Fourier and glider contraction cues, and attenuation of sensitivity to Fourier cues in the context of glider expansion. This latter result was interpreted in terms of the importance of these cues for survival, recognizing that further experimentation is necessary to solidify this viewpoint.

5.1.1 Specific predictions and suggested frameworks

Above, I mentioned several predictions and future directions for research that emerged from this research. Here, I elaborate on these ideas and suggest experimental frameworks in which they might be pursued.

The statistics of local motion signals in naturalistic movies

Our work on natural movies showed that there were statistical relationships between the different kinds of local motion signals; these relationships held for the four movies that were presented, and in several more (~50; data not shown). However, the correlations between motion types are only moderate, and we speculate that the reason that motion signals do not tightly covary is that specific kinds of local motion signals are preferentially associated with specific phenomena.

This speculation is motivated by canonical examples of stimuli that generate each kind of motion signal. Fourier motion signals are generated by frontoparallel translation — bars or other objects moving across the field of view. Non-Fourier signals — as exemplified by the stimuli introduced by Chubb & Sperling [2] — are likely to appear in the context of stereotyped occlusion and transparency [19]. Glider motion signals [3] are generated by correlations among three points in space-time, and this arises whenever objects change their distance to the observer. These observations suggest that motion associated with various kinds of natural phenomena may have specific “signatures” in terms of local motion signals. The obvious candidates are translation, occlusion, transparency, looming and receding, but other phenomena — such as self-motion — may also have such signatures. One way to test this idea is to collect a large video database of each kind of phenomenon (e.g. translation), quantify the amount of each kind of local motion signal, and look for patterns. The prediction is that specific kinds of motion signals will be associated (probably dominate) with specific natural phenomena.

Evolutionary convergence of motion computation in monkey and dragonfly

In this study we showed that neurons from two very divergent species respond in similar ways to different kinds of local motion stimuli. This suggests that at the neuronal level, similar kinds of data are being extracted. On the background of these commonalities, we observed one striking difference: that for one motion signal subtype — black glider expansion — dragonfly neurons were more sensitive than macaque. We speculated that this difference is due to the habitat and life-style of the dragonfly, and the selective pressure exerted by evolution.

To test this idea, one could study other species that live in different habitats (e.g. rabbits, fish) and have different life-styles and/or selective pressures (e.g. chameleon, whale). The prediction is that species with similar lifestyles in similar environments will share characteristics of motion processing at a fine level of detail, even if their evolutionary lineages are widely divergent. Conversely, species with similar evolutionary lineages are predicted to differ in the attributes of their motion processing, if their habitats and lifestyles are different.

Perceptual interaction of local motion signals

Our psychophysical study of how motion signals combine showed a surprising finding — that sensitivity to the most basic kind of local motion signal (Fourier motion) can be suppressed in an environment in which another kind of local motion signal (glider expansion) is prominent. I believe that this study calls for follow-ups to understand the nature of the adaptation phenomenon. Examples of questions that can be answered by straightforward variations of our paradigm are: how long does the adaptation take to develop once glider ex-

pansion signals are added to the environment, how long does it take for the adaptation to disappear once glider expansion signals are removed from the environment, whether it operates across spatial scales, and whether it transfers between eyes.

Beyond visual motion

In every study described in this thesis, motion occurs in the visual domain. However, I believe that the concept of local motion signals is much more general, and is applicable to other modalities. For example, the visual stimuli used here can be translated into auditory stimuli by considering an X-T "slice" of the stimulus to be a sound spectrogram, where the spatial dimension of the visual stimulus becomes the temporal frequency of the auditory stimulus. Extensions to other modalities — such as olfaction — may also be possible, provided one can define an appropriate metrical domain for the stimulus. Characterization of these local motion signals in natural stimuli and neural processing of them may reveal relationships between the different sensory modalities, and broad principles of neural computation.

BIBLIOGRAPHY

1. Reichardt, W. Autocorrelation, a principle for the evaluation of sensory information by the central nervous system. *Sensory communication*, 303–317 (1961).
2. Chubb, C. & Sperling, G. Drift-balanced random stimuli- A general basis for studying non-Fourier motion perception. *Optical Society of America, Journal, A: Optics and Image Science*, **5**, 1986–2007 (1988).
3. Hu, Q. & Victor, J. D. A set of high-order spatiotemporal stimuli that elicit motion and reverse-phi percepts. *Journal of vision* **10**, 9 (2010).
4. Nitzany, E., Hu, Q., Schmid, A., Yu, Y & Victor, J. D. *Characterization of primate visual cortex responses to local motion signals: Fourier, non-Fourier, and glider* San Diego, Nov. 2013.
5. Nitzany, E. I. *et al. Evolutionary convergence in computation of local motion signals in monkey and dragonfly* Salt Lake City, Feb. 2014.
6. Orger, M. B., Smear, M. C., Anstis, S. M. & Baier, H. Perception of Fourier and non-Fourier motion by larval zebrafish. *Nature neuroscience* **3**, 1128–1133 (2000).
7. Clark, D. A. *et al. Flies and humans share a motion estimation strategy that exploits natural scene statistics. Nature neuroscience* **17**, 296–303 (2014).
8. Nitzany, E. I. & Victor, J. D. The statistics of local motion signals in naturalistic movies. *Journal of vision* **14**, 10 (2014).
9. Nitzany, E., Hammel, I. & Meilijson, I. Quantal basis of vesicle growth and information content, a unified approach. *Journal of Theoretical Biology* **266**, 202–209. ISSN: 0022-5193 (Sept. 2010).

10. Geisler, W. S., Najemnik, J. & Ing, A. D. Optimal stimulus encoders for natural tasks. en. *Journal of Vision* **9**, 17. ISSN: , 1534-7362 (Dec. 2009).
11. Michotte, A. *The perception of causality* (Basic Books, Oxford, England, 1963).
12. Marr, D. *Vision: A computational investigation into the human representation and processing of visual information* (WH Freeman, 1982).
13. Fitzgerald, J. E., Katsov, A. Y., Clandinin, T. R. & Schnitzer, M. J. Symmetries in stimulus statistics shape the form of visual motion estimators. *Proceedings of the National Academy of Sciences* **108**, 12909–12914. ISSN: 0027-8424, 1091-6490 (July 2011).
14. Adelson, E. H. & Bergen, J. R. Spatiotemporal energy models for the perception of motion. *J. Opt. Soc. Am. A* **2**, 284–299 (1985).
15. Grossberg, S. Theory and Evaluative Reviews 3-D vision and figure-ground separation by visual cortex. *Perception & Psychophysics* **55**, 48–120 (1994).
16. Ullman, S. *The interpretation of visual motion*. (Oxford, England, 1979).
17. Lu, Z. L. & Sperling, G. Three-systems theory of human visual motion perception: review and update. *JOSA A* **18**, 2331–2370 (2001).
18. Bracewell, R. *The Fourier Transform & Its Applications* 3 edition. English. ISBN: 0073039381 (McGraw-Hill Science/Engineering/Math, June 1999).
19. Fleet, D. J. & Langley, K. Computational analysis of non-Fourier motion. *Vision Research* **34**, 3057–3079 (1994).
20. Van Santen, J. P. H. & Sperling, G. Elaborated reichardt detectors. *Journal of the Optical Society of America A* **2**, 300–321 (1985).

21. Reichardt, W. Evaluation of optical motion information by movement detectors. *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology* **161**, 533–547 (1987).
22. Dong, D. W. & Atick, J. J. Statistics of natural time-varying images. *Network: Computation in Neural Systems* **6**, 345–358 (1995).
23. Kayser, C., Einhäuser, W. & König, P. Temporal correlations of orientations in natural scenes. *Neurocomputing* **52**, 117–123 (2003).
24. Field, D. J. Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America. A, Optics and image science* **4**, 2379–2394. ISSN: 0740-3232 (Dec. 1987).
25. Reinagel, P & Zador, A. M. Natural scene statistics at the centre of gaze. *Network (Bristol, England)* **10**, 341–350. ISSN: 0954-898X (Nov. 1999).
26. Van Hateren, J. H. & van der Schaaf, A. Independent component filters of natural images compared with simple cells in primary visual cortex. en. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **265**, 359–366. ISSN: 0962-8452, 1471-2954 (Mar. 1998).
27. Cutting, J. E., DeLong, J. E. & Brunick, K. L. Visual activity in Hollywood film: 1935 to 2005 and beyond. *Psychology of Aesthetics, Creativity, and the Arts* **5**, 115 (2011).
28. Cutting, J. E., Brunick, K. L., DeLong, J. E., Iricinschi, C. & Candan, A. Quicker, faster, darker: Changes in Hollywood film over 75 years. *i-Perception* **2**, 569–576. ISSN: 2041-6695 (2011).
29. Ullman, S. The interpretation of structure from motion. *Proceedings of the royal society* **203**, 405–426 (Jan. 1979).

30. Vaina, L. M., Lemay, M., Bienfang, D. C., Choi, A. Y. & Nakayama, K. Intact biological motion and structure from motion perception in a patient with impaired motion mechanisms: A case study. *Visual neuroscience* **5**, 353–369 (1990).
31. Ahlstrom, V., Randolph, B. & Ahlstrom, U. Perception of biological motion. *Perception* **26**, 1539–1548 (1997).
32. Beintema, J. A. & Lappe, M. Perception of biological motion without local image motion. *Proceedings of the National Academy of Sciences* **99**, 5661 (2002).
33. Burr, D. & Thompson, P. Motion Psychophysics: 1985-2010. *Vision research* (2011).
34. Fox, R & McDaniel, C. The perception of biological motion by human infants. *Science* **218**, 486–487 (Oct. 1982).
35. Grill-Spector, K. & Malach, R. The human visual cortex. *Annual Review of Neuroscience* **27**, 649–677. ISSN: 0147-006X, 1545-4126 (July 2004).
36. Johansson, G. Visual perception of biological motion and a model for its analysis. *Attention, Perception, & Psychophysics* **14**, 201–211 (1973).
37. Koenderink, J. J. & Van Doorn, A. J. Affine structure from motion. *JOSA A* **8**, 377–385 (1991).
38. Nakayama, K. biological image motion processing. *Vision Research* **25**, 625–660 (1985).
39. Grossman, E. *et al.* Brain areas involved in perception of biological motion. *Journal of Cognitive Neuroscience* **12**, 711–720 (2000).

40. Smith, A. T., Greenlee, M. W., Singh, K. D., Kraemer, F. M. & Hennig, J. The Processing of First- and Second-Order Motion in Human Visual Cortex Assessed by Functional Magnetic Resonance Imaging (fMRI). en. *The Journal of Neuroscience* **18**, 3816–3830. ISSN: 0270-6474, 1529-2401 (May 1998).
41. Van Essen, D. C., Anderson, C. H. & Felleman, D. J. Information processing in the primate visual system: an integrated systems perspective. *Science* **255**, 419–423 (1992).
42. McKeefry, D. J., Watson, J. D. G., Frackowiak, R. S. J., Fong, K. & Zeki, S. The Activity in Human Areas V1/V2, V3, and V5 during the Perception of Coherent and Incoherent Motion. *NeuroImage* **5**, 1–12. ISSN: 1053-8119 (Jan. 1997).
43. Sincich, L. C. & Horton, J. C. The circuitry of V1 and V2: integration of color, form, and motion. *Annu. Rev. Neurosci.* **28**, 303–326 (2005).
44. Land, M. F. & Nilsson, D. E. *Animal eyes* (visited on 05/18/2014) (Oxford University Press, 2012).
45. Olberg, R. M., Worthington, A. H. & Venator, K. R. Prey pursuit and interception in dragonflies. *Journal of Comparative Physiology A* **186**, 155–162 (2000).
46. Silcox, M. T., Dalmyn, C. K. & Bloch, J. I. Virtual endocast of *Ignacius graybullianus* (Paromomyidae, Primates) and brain evolution in early primates. en. *Proceedings of the National Academy of Sciences* **106**, 10987–10992. ISSN: 0027-8424, 1091-6490 (July 2009).
47. Olberg, R. M. Identified target-selective visual interneurons descending from the dragonfly brain. *Journal of Comparative Physiology A* **159**, 827–840 (1986).

48. Lu, Z.-L. & Sperling, G. The functional architecture of human visual motion perception. *Vision Research* **35**, 2697–2722. ISSN: 0042-6989 (Oct. 1995).
49. O’Keefe, L. P. & Movshon, J. A. Processing of first-and second-order motion signals by neurons in area MT of the macaque monkey. *Visual neuroscience* **15**, 305–317 (1998).
50. Marescal, I. & Baker, C. L. Cortical processing of second-order motion. *Visual neuroscience* **16**, 527–540 (1999).
51. Anstis, S. M. Phi movement as a subtraction process. *Vision research* **10**, 1411–IN5 (1970).
52. Red, R. C., Soodak, R. E. & Shapley, R. M. Linear mechanisms of directional selectivity in simple cells of cat striate cortex. *Proceedings of the National Academy of Sciences* **84**, 8740–8744 (1987).
53. Schmid, A. M., Purpura, K. P. & Victor, J. D. Responses to Orientation Discontinuities in V1 and V2: Physiological Dissociations and Functional Implications. *The Journal of Neuroscience* **34**, 3559–3578 (2014).
54. Mechler, F., Reich, D. S. & Victor, J. D. Detection and discrimination of relative spatial phase by V1 neurons. *The Journal of neuroscience* **22**, 6129–6157 (2002).
55. Schmid, A. M., Purpura, K. P., Ohiorhenuan, I. E., Mechler, F. & Victor, J. D. Subpopulations of Neurons in Visual Area V2 Perform Differentiation and Integration Operations in Space and Time. *Frontiers in Systems Neuroscience* **3**. ISSN: 1662-5137. (Visited on 07/02/2014) (Nov. 2009).
56. Victor, J. D., Mechler, F., Repucci, M. A., Purpura, K. P. & Sharpee, T. Responses of V1 neurons to two-dimensional hermite functions. *Journal of neurophysiology* **95**, 379–400 (2006).

57. Hazan, L., Zugaro, M. & Buzski, G. Klusters, NeuroScope, NDManager: A free software suite for neurophysiological data processing and visualization. *Journal of Neuroscience Methods* **155**, 207–216. ISSN: 0165-0270 (Sept. 2006).
58. Husch, A., Paehler, M., Fusca, D., Paeger, L. & Kloppenburg, P. Distinct Electrophysiological Properties in Subtypes of Nonspiking Olfactory Local Interneurons Correlate With Their Cell Type Specific Ca²⁺ Current Profiles. *Journal of Neurophysiology* **102**, 2834–2845. ISSN: 0022-3077, 1522-1598 (Nov. 2009).
59. Strausfeld, N. J. Atlas of an insect brain. English (1976).
60. Lott, G. K. *Hybridizing cellular and behavioral neurobiology with modern engineering tools: microelectronics, microfabricated devices, and software solutions for physiology* PhD thesis (Cornell University, 2007).
61. Quiroga, R. Q., Nadasdy, Z. & Ben-Shaul, Y. Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering. *Neural computation* **16**, 1661–1687 (2004).
62. Ruck, P. A comparison of the electrical responses of compound eyes and dorsal ocelli in four insect species. *Journal of Insect Physiology* **2**, 261–274 (1958).
63. Ledgeway, T. Adaptation to second-order motion results in a motion aftereffect for directionally-ambiguous test stimuli. *Vision Research* **34**, 2879–2889. ISSN: 0042-6989 (Nov. 1994).
64. Vaina, L. M. & Soloviev, S. in *Progress in Brain Research* (ed Heywood, M. and Blakemore) 197–212 (Elsevier, 2004).

65. Yo, C. & Wilson, H. R. Perceived direction of moving two-dimensional patterns depends on duration, contrast and eccentricity. *Vision Research* **32**, 135–147. ISSN: 0042-6989 (Jan. 1992).
66. Albright, T. D. Form-cue invariant motion processing in primate visual cortex. eng. *Science (New York, N.Y.)* **255**, 1141–1143. ISSN: 0036-8075 (Feb. 1992).
67. Chaudhuri, A. & Albright, T. D. Neuronal responses to edges defined by luminance vs. temporal texture in macaque area V1. *Visual Neuroscience* **14**, 949–962. ISSN: 1469-8714 (Sept. 1997).
68. Piponnier, J.-C. *et al.* First- and second-order stimuli reaction time measures are highly sensitive to mild traumatic brain injuries. *Journal of Neurotrauma*. ISSN: 0897-7151. doi:10.1089/neu.2014.3832. (Visited on 05/17/2015) (May 2015).
69. Plant, G. T., Laxer, K. D., Barbaro, N. M., Schiffman, J. S. & Nakayama, K. Impaired visual motion perception in the contralateral hemifield following unilateral posterior cerebral lesions in humans. en. *Brain* **116**, 1303–1335. ISSN: 0006-8950, 1460-2156 (Dec. 1993).
70. Plant, G. T. & Nakayama, K. The characteristics of residual motion perception in the hemifield contralateral to lateral occipital lesions in humans. en. *Brain* **116**, 1337–1353. ISSN: 0006-8950, 1460-2156 (Dec. 1993).
71. Greenlee, M. W. & Smith, A. T. Detection and Discrimination of First- and Second-Order Motion in Patients with Unilateral Brain Damage. en. *The Journal of Neuroscience* **17**, 804–818. ISSN: 0270-6474, 1529-2401 (Jan. 1997).

72. Victor, J. D. & Conte, M. M. Evoked potential and psychophysical analysis of Fourier and non-Fourier motion mechanisms. *Visual Neuroscience* **9**, 105–123. ISSN: 1469-8714 (Aug. 1992).
73. Edwards, M. & Badcock, D. R. Global motion perception: No interaction between the first- and second-order motion pathways. *Vision Research* **35**, 2589–2602. ISSN: 0042-6989 (Sept. 1995).
74. Victor, J. D. & Conte, M. M. Local image statistics: maximum-entropy constructions and perceptual salience. *Journal of the Optical Society of America A* **29**, 1313–1345 (July 2012).
75. Victor, J. D., Chubb, C. & Conte, M. M. Interaction of luminance and higher-order statistics in texture discrimination. *eng. Vision Research* **45**, 311–328. ISSN: 0042-6989 (Feb. 2005).
76. Chubb, C., McGowan, J., Sperling, G. & Werkhoven, P. *Non-Fourier motion analysis* in *CIBA Foundation Symposium* (JOHN WILEY & SONS LTD, 1994), 193–193.
77. Victor, J. D. & Conte, M. M. Motion mechanisms have only limited access to form information. *Vision Research* **30**, 289–301. ISSN: 0042-6989 (1990).
78. Ledgeway, T. & Smith, A. T. Evidence for separate motion-detecting mechanisms for first- and second-order motion in human vision. *Vision Research* **34**, 2727–2740. ISSN: 0042-6989 (Oct. 1994).