Gepshtein S & Tyukin I (in press). Optimal measurement of visual motion across spatial and temporal scales.
In Favorskaya MN and Jain LC (Eds), Computer Vision in Advanced Control Systems using Conventional and Intelligent Paradigms, Intelligent Systems Reference Library, Springer-Verlag, Berlin.

Optimal measurement of visual motion across spatial and temporal scales

Sergei Gepshtein[1] and Ivan Tyukin[2,3]

[1] Systems Neurobiology Laboratories, Salk Institute for Biological Studies
10010 North Torrey Pines Road, La Jolla, CA 92037, USA
sergei@salk.edu

[2] Department of Mathematics, University of Leicester
University Road, Leicester LE1 7RH, UK
i.tyukin@le.ac.uk

[3] Department of Automation and Control Processes
Saint-Petersburg State Electrotechnical University
5 Professora Popova Str., 197376, Saint-Petersburg, Russia

April 26, 2014

Abstract. Sensory systems use limited resources to mediate the perception of a great variety of objects and events. Here a normative framework is presented for exploring how the problem of efficient allocation of resources can be solved in visual perception. Starting with a basic property of every measurement, captured by Gabor's uncertainty relation about the location and frequency content of signals, prescriptions are developed for optimal allocation of sensors for reliable perception of visual motion. This study reveals that a large-scale characteristic of human vision (the spatiotemporal contrast sensitivity function) is similar to the optimal prescription, and it suggests that some previously puzzling phenomena of visual sensitivity, adaptation, and perceptual organization have simple principled explanations.

Keywords: resource allocation, contrast sensitivity, perceptual organization, sensory adaptation, automated sensing

Table of Contents

Optimal measurement of visual motion across spatial and temporal scales 1
Sergei Gepshtein and Ivan Tyukin
1 Introduction . . . 2
2 Gabor's uncertainty relation in one dimension . . . 5
2.1 Single sensors . . . 5
2.2 Sensor populations . . . 7
2.3 Cooperative measurement . . . 9
3 Gabor's uncertainty in space-time . . . 10
3.1 Uncertainty in two dimensions . . . 10
3.2 Equivalence classes of uncertainty . . . 11
3.3 Spatiotemporal interaction: speed . . . 13
4 Optimal conditions for motion measurement . . . 15
4.1 Minima of uncertainty . . . 15
4.2 The shape of optimal set . . . 16
5 Sensor allocation . . . 18
5.1 Adaptive allocation . . . 18
5.2 Mechanism of adaptive allocation . . . 20
5.3 Conclusions . . . 21
6 Appendices . . . 25
6.1 Appendix 1. Additivity of uncertainty . . . 25
6.2 Appendix 2. Improving resolution by multiple sampling . . . 26

1 Introduction

Biological sensory systems collect information from a vast range of spatial and temporal scales. For example, human vision can discern modulations of luminance that span nearly seven octaves of spatial and temporal frequencies, while many properties of optical stimulation (such as the speed and direction of motion) are analyzed within every step of the scale.

The large amount of information is encoded and transformed for the sake of specific visual tasks using limited resources. In biological systems, it is a large but finite number of neural cells. The cells are specialized: sensitive to a small subset of optical signals, presenting sensory systems with the problem of allocation of limited resources. This chapter is concerned with how this problem is solved by biological vision. How are the specialized cells distributed across the great number of potential optical signals in the environments that are diverse and variable?

The extensive history of vision science suggests that any attempt of vision theory should begin with an analysis of the tasks performed by visual systems. Following Aristotle, one may begin with the definition of vision as "knowing hat

is where by looking" [1]. The following argument concerns the basic visual tasks captured by this definition.

The "what" and "where" of visual perception are associated with two characteristics of optical signals: their frequency content and locations, in space and time. The last statement implicates at least five dimensions of optical signals (which will become clear in a moment).

The basic visual tasks are bound by first principles of measurement. To see that, consider a measurement device (a "sensor" or "cell") that integrates its inputs over some spatiotemporal interval. An individual device of an arbitrary size will be more suited for measuring the location *or* the frequency content of the signal, reflected in the uncertainties of measurement. The uncertainties associated with the location and the frequency content are related by a simple law formalized by Gabor [2], who showed that the two uncertainties trade off across scales. As the scale changes, one uncertainty rises and the other falls.

Assuming that the visual systems in question are interested in both the locations and frequency content of optical signals ("stimuli"), the tradeoff of uncertainties will attain a desired ("optimal") balance of uncertainties at some intermediate scale. The notion of the optimal tradeoff of uncertainty has received considerable attention in studies of biological vision. This is because the "receptive fields" of single neural cells early in the visual pathways appear to approximate one or another form of the optimal tradeoff [3–10].

Here the tradeoff of uncertainties is formulated in a manner that is helpful for investigating its consequences outside of the optimum: across many scales, and for cell populations rather than for single cells. Then the question is posed of how the scales of multiple sensory cells should be selected for simultaneously minimizing the uncertainty of measurement for all the cells, on several stimulus dimensions.

The present focus is on how visual motion can be estimated at the lowest overall uncertainty of measurement across the entire range of useful sensor sizes (in artificial systems) or the entire range of receptive fields (in biological systems). In other words, the following is an attempt to develop an economic *normative theory* of motion-sensitive systems. Norms are derived for efficient design of such systems, and then the norms are compared with facts of biological vision.

This approach from first principles of measurement and parsimony helps to understand the forces that shape the characteristics of biological vision, but which had appeared intractable or controversial using previous methods. These characteristics include the spatiotemporal contrast sensitivity function, adaptive transformations of this function caused by stimulus change, and also some characteristics of the higher-level perceptual processes, such as perceptual organization.

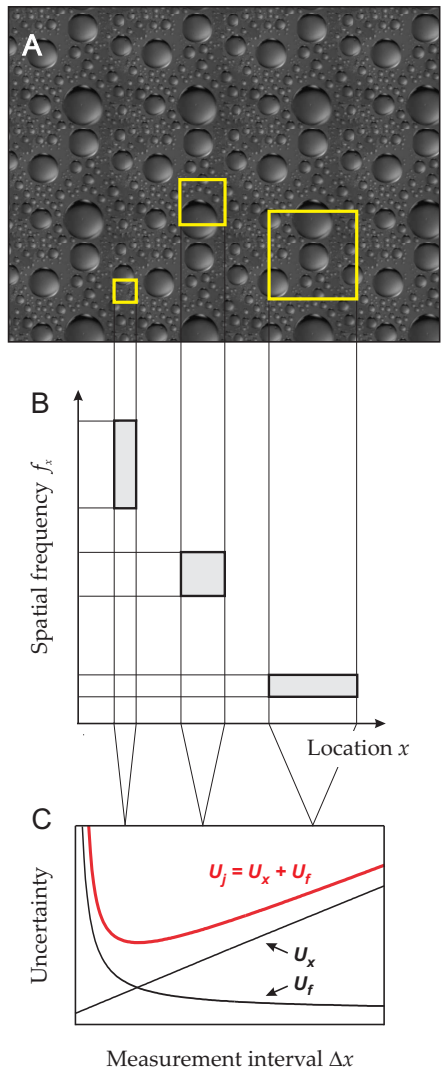

Fig. 1: Components of measurement uncertainty. (A) The image is sampled by three sensors of different sizes. (B) The three sensors are associated with Gabor's *logons*: three rectangles that have the same areas but different shapes, according to the limiting condition of the uncertainty relation in Eq. 3. (C) Functions U_x and U_f represent the uncertainties about the location and content of the measured signal (the horizontal and vertical extents of the logons in B, respectively), and function U_j represents the joint uncertainty about signal location and content.

2 Gabor's uncertainty relation in one dimension

The outcomes of measuring the location and the frequency content of any signal by a single sensory device are not independent of one another. The measurement of location assigns the signal to interval Δx on some dimension of interest x. The smaller the interval the lower the uncertainty about signal location. The uncertainty is often described in terms of the precision of measurements, quantified by the dispersion of the measurement interval or, even simpler, by the size of the interval, Δx. The smaller the interval, the lower the uncertainty about location, and the higher the precision of measurement.

The measurement of frequency content evaluates how the signal varies over x, i.e., the measurement is best described on the dimension of frequency of signal variation, f_x. That is, the measurement of frequency content is equivalent to localizing the signal on f_x: assigning the signal to some interval Δf_x. Again, the smaller the interval, the lower the uncertainty of measurement and the higher the precision.[4]

The product of uncertainties about the location and frequency content of the signal is bounded "from below" [2, 11–13]. The product cannot be smaller than some positive constant C_x:

$$U_x U_f \geq C_x, \tag{1}$$

where U_x and U_f are the uncertainties about the location and frequency content of the signal, respectively, measured on the intervals Δx and Δf_x.

Eq. 1 means that any measurement has a limit at $U_x U_f = C_x$. At the limit, decreasing one uncertainty is accompanied by increasing the other. For simplicity, let us quantify the measurement uncertainty by the size of the measurement interval. Gabor's uncertainty relation may therefore be written as

$$\Delta x \Delta f_x \geq C_x, \tag{2}$$

and its limiting condition as

$$\Delta x \Delta f_x = C_x. \tag{3}$$

2.1 Single sensors

Let us consider how the uncertainty relation constrains the measurements by a single measuring device: a "sensor." Fig 1 illustrates three spatial sensors of different sizes. In Fig 1A, the measurement intervals of the sensors are defined on two spatial dimensions. For simplicity, let us consider just one spatial dimension, x, so the interval of measurement ("sensor size") is Δx, as in Fig 1B–C.

The limiting effect of the uncertainty relation for such sensors has a convenient graphic representation called "information diagram" (Fig 1B). Let the two multiplicative terms of Eq. 3 be represented by the two sides of a rectangle in coordinate plane (x, f_x). Then C_x is the rectangle area. Such rectangles are called "information cells" or "logons." Three logons, of different shapes but of the same area C_x, are shown in Fig 1B, representing the three sensors:

[4] For brevity, here "frequency content" will sometimes be shortened to "content."

- The logon of the smallest sensor (smallest Δx, left) is thin and tall, indicating that the sensor has a high precision on x and a low precision on f_x.
- The logon of the largest sensor (right) is thick and short, indicating a low precision on x and a high precision on f_x.
- The above sensors are specialized for measuring either the location or frequency content of signals. The medium-size sensor (middle) offers a compromise: its uncertainties are not as low as the lowest uncertainties (but not as high as the highest uncertainties) of the specialized sensors. In this respect, the medium-size sensor trades one kind of uncertainty for another.

The medium-size sensors are most useful for jointly measuring the locations and frequency content of signals.

So far, the ranking of sensors has been formalize using an additive model of uncertainty (Fig 1C). The motivation for such an additive model is presented in Appendix 1. This approach is motivated by the assumption that visual systems have no access to complete prior information about the statistics of measured signals (such as the joint probability density functions for the spatial and temporal locations of stimuli and their frequency content). The assumption is, instead, that the systems can reliably estimate only the means and variances of the measured quantities.

Accordingly, the overall uncertainty in Fig 1C has the following components. The increasing function represents the uncertainty about signal location: $U_x = \Delta x$. The decreasing function represents the uncertainty about signal content: $U_f = \Delta f_x = C_x/\Delta x$ (from Eq. 3). The joint uncertainty of measuring signal location *and* content is represented by the non-monotonic function U_j:

$$U_j = \lambda_x U_x + \lambda_f U_f = \lambda_x \Delta x + \lambda_f \frac{1}{\Delta x}, \tag{4}$$

where λ_x and λ_f are positive coefficients reflecting how important the components of uncertainty are relative to one another.

The additive model of Eq. 4 implements a worst-case estimate of the overall uncertainty (as it is explained in section *The Minimax Principle* just below). The additive components are weighted, while the weights are playing several roles. They bring the components of uncertainty to the same units, allowing for different magnitude of C_x,[5] and representing the fact that the relative importance of the components depends on the task at hand.

The joint uncertainty function (U_j in Fig 1C) has its minimum at an intermediate value of Δx. This is a *point of equilibrium* of uncertainties, in that a sensor of this size implements a perfect balance of uncertainties about the location and frequency content of the signal [14]. If measurements are made in the interest of high precision, and if the location and the frequency content of the signal are equally important, then a sensor of this size is the best choice for jointly measuring the location and the frequency content of the signal.

[5] Different criteria of measurement and sensor shapes correspond to different magnitudes of C_x.

The Minimax Principle. What is the best way to allocated resources in order to reduce the chance of gross errors of measurement. One approach to solving this problem is using the minimax strategy devised in game theory for modeling choice behavior [15, 16]. Generally, the minimax strategy is used for estimating the maximal expected loss for every choice and then pursuing the choices, for which the expected maximal loss is minimal. In the present case, the choice is between the sensors that deliver information with variable uncertainty.

In the following, the minimax strategy is implemented by assuming the maximal (worst-case) uncertainty of measurement on the sensors that span the entire range of the useful spatial and temporal scales. This strategy is used in two ways. First, the consequences of Gabor's uncertainty relation are investigated by assuming that the uncertainty of measurement is as high as possible (i.e., using the limiting case of uncertainty relation; Eq. 3). Second, the outcomes of measurement on different sensors are anticipated by *adding* their component uncertainties, i.e., using the joint uncertainty function of Eq. 4. (The choice of the additive model is explained in Appendix 1.) It is assumed that sensor preferences are ranked according to the expected maximal uncertainty: the lower the uncertainty, the higher the preference.

2.2 Sensor populations

Real sensory systems have at their disposal large but limited numbers of sensors. Since every sensor is useful for measuring only some aspects of the stimulus, sensory systems must solve an economic problem: they must distribute their sensors in the interest of perception of many different stimuli. Let us consider this problem using some simple arrangements of sensors.

First, consider a population of identical sensors in which the measurement intervals do not overlap. Fig 2A contains three examples of such sensors, using the information diagram introduced in Fig 1. Each of the three diagrams in Fig 2A portrays four sensors, identical to one another except they are tuned to different intervals on x (which can be space or time). Each panel also contains a representation of a narrow-band signal: the yellow circle, the same across the three panels of Fig 2A. The different arrangements of sensors imply different *resolutions* of the system for measuring the location and frequency content of the stimulus.

- The population of small sensors (small Δx on the left of Fig 2A) is most suitable for measuring signal location: the test signal is assigned to the rightmost quarter on the range of interest in x. In contrast, measurement of frequency content is poor: signals presented anywhere within the vertical extent of the sensor (i.e., within the large interval on f_x) will all give the same response. This system has a good location resolution and poor frequency resolution.

- The population of large sensors (large Δx on the right of Fig 2A) is most suitable for measuring frequency content. The test signal is assigned to a small interval on f_x. Measurement of location is poor. This system has a good frequency resolution and poor location resolution.

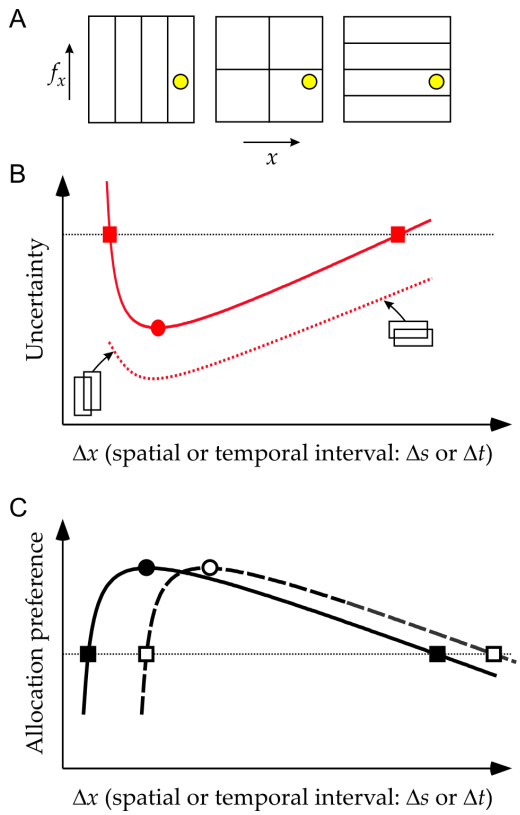

Fig. 2: Allocation of multiple sensors. (A) Information diagrams for a population of four sensors, using sensors of the same size within each population, and of different sizes across the populations. (B) Uncertainty functions. The red curve is the joint uncertainty function introduced in Fig 1, with the markers indicating special conditions of measurement: the lowest joint uncertainty (the circle) and the equivalent joint uncertainty (the squares), anticipating the *optimal sets* and the *equivalence classes* of measurement in the higher-dimensional systems illustrated in Figs 3–4. (C) Preference functions. The solid curve is a function of allocation preference (here reciprocal to the uncertainty function in B): an optimal distribution of sensors, expected to shift (dashed curve) in response to change in stimulus usefulness.

– The population of medium-size sensors can obtain useful information about both locations and frequency content of signals. It has a better frequency resolution than the population of small sensors, and a better location resolution than the population of large sensors.

Consequences of the different sensor sizes are summarized by the joint uncertainty function in Fig 2B. (For non-overlapping sensors, the function has the same shape as in Fig 1C). The figure makes it clear that the sensors or sensor populations with very different properties can be equivalent in terms of their joint uncertainty. For example, the two filled squares in Fig 2B mark the uncertainties of two different sensor populations: one contains only small sensors and the other contains only large sensors.

The populations of sensors in which the measurement intervals overlap are more versatile than the populations of non-overlapping sensors. For example, the sensors with large overlapping intervals can be used to emulate measurements by the sensors with smaller intervals (Appendix 2), reducing the uncertainty of stimulus localization. Similarly, groups of the overlapping sensors with small measurement intervals can emulate the measurements by sensors with larger intervals, reducing the uncertainty of identification. Overall, a population of the overlapping sensors can afford lower uncertainties across the entire range of measurement intervals, represented in Fig 2B by the dotted curve: a lower-envelope uncertainty function. Still, the new uncertainty function has the same shape as the previous function (represented by the solid line) because of the limited total number of the sensors.

2.3 Cooperative measurement

To illustrate the benefits of measurement using multiple sensors, suppose that the stimulation was uniform and one could vary the number of sensors in the population at will, starting with a system that has only a few sensors, toward a system that has an unlimited number of sensors.

– A system equipped with very limited resources, and seeking to measure both the location and the frequency content of signals, will have to be unmitigatedly frugal. It will use only the sensors of medium size, because only such sensors offer useful (if limited) information about both properties of signals.

– A system enjoying unlimited resources, will be able to afford many specialized sensors, or groups of such sensors (represented by the information diagrams in Fig 2A).

– A moderately wealthy system: a realistic middle ground between the extremes outlined above, will be able to escape the straits of Gabor's uncertainty relation using different specialized sensors and thus measuring the location and content of signals with a high precision.

As one considers systems with different numbers of sensors, from small to large, one expects to find an increasing ability of the system to afford the large and

small measurement intervals. As the number of sensors increases, the allocation of sensors will expand in two directions, up and down on the dimension of sensor scale: from using only the medium-size sensors in the poor system, to using also the small and large sensors in the wealthier systems. This allocation policy is illustrated in Fig 2C. The *preference function* in Fig 2C indicates that, as the more useful sensors are expected to grow in number, the distribution of sensors will form a smooth function across the scales. As mentioned, the sensitivity of the system is expected to follow a function monotonically related to the preference function.

Increasing the number of sensors selective to the same stimulus condition is expected to improve sensory performance, manifested in lower sensory thresholds. One reason for such improvement in biological sensory systems is the fact that integrating information across multiple sensors will help to reduce the detrimental effect of the noisy fluctuations of neural activity, in particular when the noises are uncorrelated.

The preference function in Fig 2C is exceedingly simple: it merely mirrors the joint uncertainty function of Fig 2B. This example helps to illustrate some special conditions of the uncertainty of measurement and to anticipate their consequences for sensory performance. First, the minimum of uncertainty corresponds to the maximum of allocation preference, where the highest sensitivity is expected. Second, equal uncertainties correspond to equal allocation preferences, where equal sensitivities are expected. Allocation policies are considered again in Sections 4–5, where the relationship is studied between a normative prescription for resource allocation and a characteristic of performance in biological vision.

3 Gabor's uncertainty in space-time

3.1 Uncertainty in two dimensions

Now consider a more complex case where signals vary on two dimensions: space and time. Here, the measurement uncertainty has four components, illustrated in Fig 3A. The bottom of Fig 3A is a graph of the spatial and temporal sensor sizes $(T, S) = (\Delta t, \Delta s)$. Every point in this graph corresponds to a "condition of measurement" associated with the four properties of sensors.[6] By Gabor's uncertainty relation, spatial and temporal intervals $(\Delta t, \Delta s)$ are associated with, respectively, the spatial and temporal frequency intervals $(\Delta f_t, \Delta f_s)$.

The four-fold dependency is explained on the side panels of the figure using Gabor's logons, each associated with a sensor labeled by a numbered disc. For example, in sensor 7 the spatial and temporal intervals are small, indicating a good precision of spatial and temporal localization (i.e., concerning "where" and "when" the stimuli occurs). But the spatial and temporal frequency intervals are large, indicating a low precision in measuring spatial and temporal frequency

[6] Here the sensors are characterized by intervals following the standard notion that biological motion sensors are maximally activated when the stimulus travels some distance Δs over some temporal interval Δt [17].

content (a low capacity to serve the "what" task of stimulus identification). This pattern is reversed in sensor 3, where the precision of localization is low but the precision of identification is high.

As in the previous example (Fig 1B–C), here the one-dimensional uncertainties are summarized using joint uncertainty functions: the red curves on the side panels of Fig 3B. Each function has the form of Eq. 4, applied separately to spatial:

$$U_S = \lambda_1 S + \lambda_2 / S$$

and temporal:

$$U_T = \lambda_3 T + \lambda_4 / T,$$

dimensions, where $S = \Delta s$ and $T = \Delta t$. Next, spatial and temporal uncertainties are combine for every spatiotemporal condition:

$$U_{ST} = U_T + U_S,$$

to obtain a bivariate *spatiotemporal uncertainty function*:

$$U_{ST} = \lambda_1 S + \frac{\lambda_2}{S} + \lambda_3 T + \frac{\lambda_4}{T} \tag{5}$$

represented in Fig 3B by a surface.

The spatiotemporal uncertainty function in Fig 3B has a unique minimum, of which the projection on graph (T, S) is marked by the red dot: the point of perfect balance of the four components of measurement uncertainty. Among the conditions of imperfect balance of uncertainties, consider the conditions of an equally imperfect balance. These are the *equivalence classes* of measurement uncertainty, represented by the level curves of the surface. The concentric contours on the bottom of Fig 3B are the projections of some of the level curves.

3.2 Equivalence classes of uncertainty

Contours of equal measurement uncertainty are reproduced in Fig 4 from the bottom of Fig 3B. The pairs of connected circles indicate that the slopes of equivalence contours vary across the conditions of measurement. This fact has several interesting implications for the perception of visual motion.

First, if the equivalent conditions of motion perception were consistent with the equivalent conditions of uncertainty, then some lawful changes in the perception of motion would be expected for stimuli that activate sensors in different parts of the sensor space. This prediction was confirmed in studies of apparent motion, which is the experience of motion from discontinuous displays, where the sequential views of the moving objects (the "corresponding image parts") are separated by spatial (σ) and temporal (τ) distances. Perceptual strength of apparent motion in such displays was conserved: sometimes by changing σ and τ in the same direction (both increasing or both decreasing), which is the regime of space-time coupling [22], and sometimes by trading off one distance for another: the regime of space-time tradeoff [23]. Gepshtein and Kubovy [20] found that

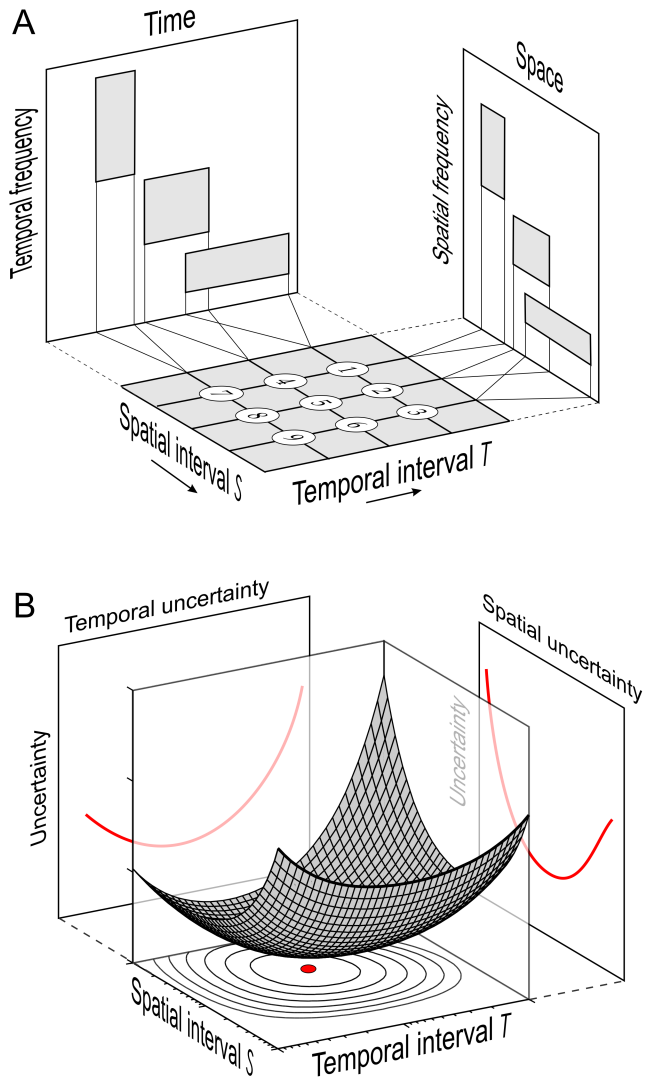

Fig. 3: Components of measurement uncertainty in space-time. (A) Spatial and temporal information diagrams of spatiotemporal measurements. The numbered discs each represents a sensor of particular spatial and temporal extent, $S = \Delta s$ and $T = \Delta t$. The rectangles on side panels are the spatial and temporal logons associated with the sensors. (B) The surface represents the joint uncertainty about signal location and frequency content of signals across sensors of different spatial and temporal size. The contours in the bottom plane (S, T) are sets of equivalent uncertainty (reproduced for further consideration in Fig 4). Panel A is adopted from [18] and panel B from [19].

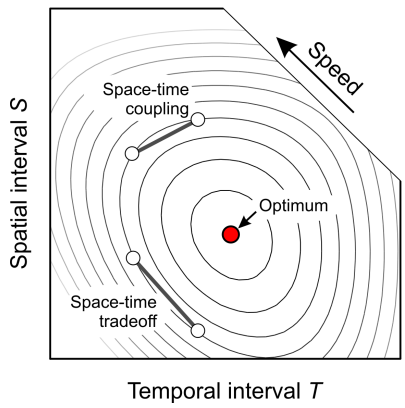

Fig. 4: Equivalence classes of uncertainty. The contours represent equal measurement uncertainty (reproduced from the bottom panel of Fig 3B) and the red circle represents the minimum of uncertainty. The pairs of connected circles labeled "space-time coupling" and "space-time tradeoff" indicate why some studies of apparent motion discovered different regimes of motion perception in different stimuli [20, 21].

the two regimes of apparent motion were special cases of a lawful pattern: one regime yielded to another as a function of speed, consistent with the predictions illustrated in Fig 4.

Second, the regime of space-time coupling undermines one of the cornerstones of the literature on visual perceptual organization: the *proximity principle* of perceptual grouping [24, 25]. The principle is an experimental observations from the early days of the Gestalt movement, capturing the common observation that the strength of grouping between image parts depends on their distance: the shorter the distance the stronger the grouping. In space-time, the principle would hold if the strength of grouping had not changed, when increasing one distance (σ or τ) was accompanied by decreasing the other distance (τ or σ): the regime of tradeoff [26]. The fact that the strength of grouping is maintained by increasing both σ and τ, or by or decreasing both σ and τ, is inconsistent with the proximity principle [21].

3.3 Spatiotemporal interaction: speed

Now let us consider the interaction of the spatial and temporal dimensions of measurement. A key aspect of this interaction is the speed of stimulus variation: the rate of temporal change of stimulus intensity across spatial location. The dimension of speed has been playing a central role in the theoretical and empirical studies of visual perception [27, 17, 28]. Not only is the perception of speed crucial for the survival of mobile animals, but it also constitutes a rich source of auxiliary information for parsing the optical stimulation [29, 30].

What is more, speed appears to play the role of a control parameter in the organization of visual sensitivity. The shape of a large-scale characteristic of visual sensitivity (measured using continuous stimuli) is invariant with respect to speed [31, 32]. And a characteristic of the strength of perceived motion in discontinuous stimuli (giving rise to "apparent motion") collapse onto a single function when plotted against speed [20].

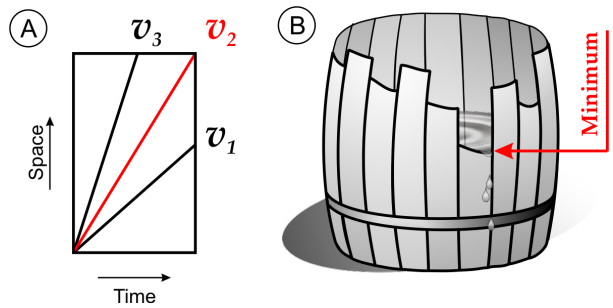

Fig. 5: Economic measurement of speed. (A) The rectangle represents a sensor defined by spatial and temporal intervals (S and T). From considerations of parsimony, the sensor is more suitable for measurement of speed $v_2 = S/T$ than v_1 or v_3 since no part of S or T is wasted in measurement of v_2. (B) Liebig's barrel. The shortest stave determines barrel's capacity. Parts of longer staves are wasted since they do not affect the capacity.

From the present normative perspective, the considerations of speed measurement (combined with the foregoing considerations of measuring the location and frequency content) of visual stimuli have two pervasive consequences, which are reviewed in some detail next. First, in a system optimized for the measurement of speed, the expected distribution of the quality of measurement has an invariant shape, distinct from the shape of such a distribution conceived before one has taken into account the measurement of speed (Fig 4). Second, the *dynamics* of visual measurement, and not only its static organization, will depend on the manner of interaction of the spatial and temporal aspects of measurement.

In Figs 3-4, a distribution of the expected uncertainty of measurement was derived from a local constraint on measurement. The local constraint was defined separately for the spatial and temporal intervals of the sensor. The considerations of speed measurement add another constraint, which has to do with the relationship between the spatial and temporal intervals.

The ability to measure speed by a sensor defined by spatial and temporal intervals depends on the extent of these intervals. As it is shown in Fig 5A, different ratios of the spatial extent to the temporal extent make the sensor differently suitable for measuring different magnitudes of speed.

This argument is one consequence of the *Law of The Minimum* [33], illustrated in Fig 5B using Liebig's barrel. A broken barrel with the staves of different lengths can hold as much content as the shortest stave allows. Using the staves of different lengths is wasteful because a barrel with all staves as short as the shortest stave would do just as well. In other words, the barrel's capacity is limited by the shortest stave.

Similarly, a sensor's capacity for measuring the speed is limited by the extent of its spatial and temporal intervals. The capacity is not used fully if the spatial and temporal projections of vector v are larger or smaller than the spatial and

temporal extents allow (v_1 and v_3 in Fig 5B). Just as the extra length of the long staves is wasted in the Liebig's barrel, the spatial extent of the sensor is wasted in measurement of v_1 and the temporal extent is wasted in measurement of v_3. Let us therefore start with the assumption that the sensor defined by the intervals S and T is best suited for measuring speed $v = S/T$.

4 Optimal conditions for motion measurement

4.1 Minima of uncertainty

The optimal conditions of measurement are expected where the measurement uncertainty is the lowest. Using a shorthand notation for the spatial and temporal partial derivatives of U_{ST} in Eq. 4, $\partial U_S = \partial U_{ST}/\partial S$ and $\partial U_T = \partial U_{ST}/\partial T$, the minimum of measurement uncertainty is the solution of

$$\partial U_T \, dT + \partial U_S \, dS = 0. \tag{6}$$

The optimal condition for the entire space of sensors, disregarding individual speeds, is marked as the red point in Fig 4. To find the minima for specific speeds v_i, let us rewrite Eq. 6 such that speed appears in the equation as an explicit term. By dividing each side of Eq. 6 by dT, and using the fact that $v = dS/dT$, it follows that

$$\partial U_S v_i + \partial U_T = 0. \tag{7}$$

The solution of Eq. 7 is a set of optimal conditions of measurement across speeds. To illustrate the solution graphically, consider the vector form of Eq. 7, i.e., the scalar product

$$\left\langle \mathbf{g}_{(T,S)}, \ \mathbf{v}_{(T,S)} \right\rangle = 0, \tag{8}$$

where the first term is the gradient of measurement uncertainty function,

$$\mathbf{g}_{(T,S)} = (\partial u_T, \partial u_S), \tag{9}$$

and the second term is the speed,

$$\mathbf{v}_{(T,S)} = (T, vT), \tag{10}$$

for sensors with parameters (T, S). For now, assume that the speed to which a sensor is tuned is the ratio of spatial to temporal intervals ($v = S/T$) that define the logon of the sensor. (Normative considerations of speed tuning are reviewed in section *Spatiotemporal interaction: speed.*)

The two terms of Eq. 8 are shown in Fig 6: separately in panels A-B and together in panel C. The blue curve in panel C represents the set of conditions where vectors $\mathbf{v}$ and $\mathbf{g}$ are orthogonal to one another, satisfying Eq. 8. This curve is the *optimal set* for measuring speed while minimizing the uncertainty about signal location and content.

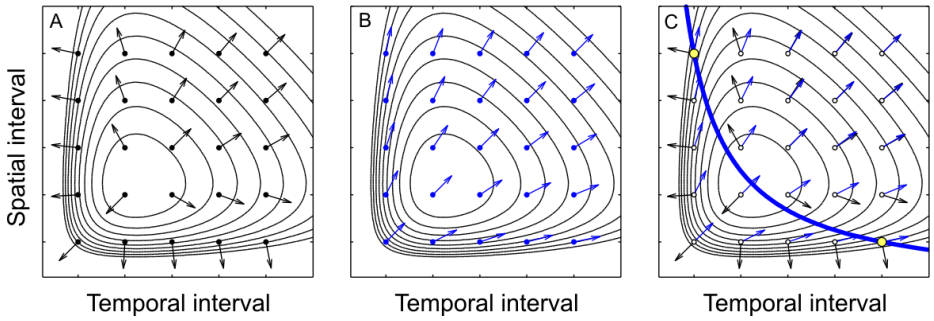

Fig. 6: Graphical solution of Eq. 8 without integration of speed. (A) Local gradients of measurement uncertainty **g**. (B) Speeds **v** to which the different sensors are tuned. (C) Optimal conditions (blue curve) arise where **g** and **v** are orthogonal to one another (Eq. 8). The yellow circles are two examples of locations where the requirement of orthogonality is satisfied. (Arrow lengths are normalized to avoid clutter.)

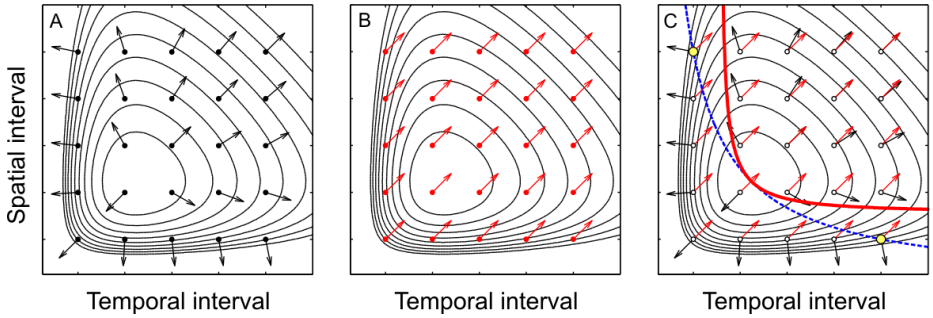

Fig. 7: Graphical solution of Eq. 8 with integration of speed. (A) Local gradient of measurement uncertainty **g** as in Fig 6A. (B) Speeds **v** integrated across multiple speeds. (C) Now the optimal conditions (red curve) arise at locations different from those in Fig 6 (the blue curve is a copy from Fig 6C).

4.2 The shape of optimal set

The solution of Eq. 8 was derived for speed defined at every point in the space of intervals (T,S): the blue arrows in Fig 6B. This picture is an abstraction that disregards the fact that measurements are performed while the sensors integrate stimulation over sensor extent. The solution of Eq. 8 that takes this fact into account is described in Fig 7. The integration reduces differences between the directions of adjacent speed vectors (panel B), and so the condition of orthogonality of **g** and **v** is satisfied at locations other than in Fig 6.

The red curve Fig 7C is the *integral optimal set* for measuring speed. This figure presents an extreme case, where speeds are integrated across the entire range of stimulation, as if every sensor had access to the expected speed of

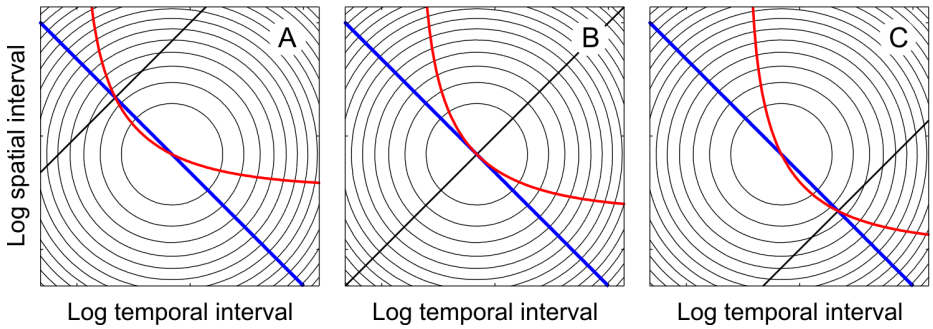

Fig. 8: Effect of expected stimulus speed. The red and blue curves are the optimal sets derived in Figs 6-7, now shown in logarithmic coordinates to emphasize that the "integral" optimal set (red) has the invariant shape of a rectangular hyperbola, whereas the "local" optimal set (blue) does not. From A to C, the expected stimulus speed (Eq. 11) decreases, represented by the black lines. The position of the integral optimal set changes accordingly.

stimulation across the entire range of stimulus speed:

$$v_e = \int_0^\infty p(v)\, v\, dv, \tag{11}$$

where $p(v)$ is the distribution of speed in the stimulation. At this extreme, every $\mathbf{v}$ is co-directional with the expected speed.

In comparison to the local optimal set (the blue curve in Fig 7C), many points of the integral optimal set (the red curve) are shifted away from the origin of the parameter space. The shift is small in the area of expected speed v_e (the black line in Fig 8), yet the shift increases away from the expected speed, such that the integral optimal set has the shape of a hyperbola.

The position of the optimal set in the parameter space depends on the prevailing speed of stimulation [19], as Fig 8 illustrates. This dependence is expected to be more pronounced in cases where the integration by receptive fields is large.

To summarize, the above argument has been concerned with how speed integration affects the optimal conditions for speed measurement. At one extreme, with no integration, the set of optimal conditions could have any shape. At the other extreme, with the scope of integration maximally large, the optimal set is a hyperbola. In between, the larger the scope of integration, the more the optimal set resembles a hyperbola. The position of this hyperbola in the parameter space depends on the prevalent speed of stimulation.

This argument has two significant implications. First, the distribution of resources in the visual system is predicted to have an invariant shape, which is consistent with results of measurements in biological vision (Fig 9) using a variety of psychophysical tasks and stimuli [34, 35, 27, 36]. Second, it implies that changes in statistics of stimulation will have a predictable effect on allocation

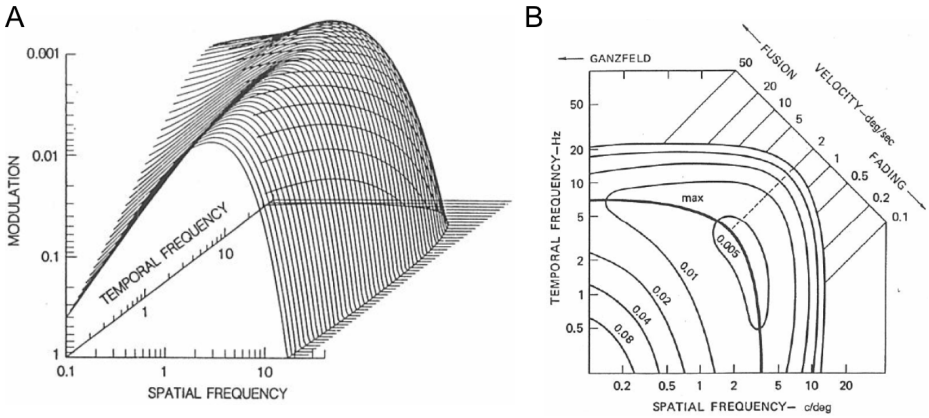

Fig. 9: Human spatiotemporal contrast sensitivity function, shown as a surface in A and a contour plot in B. Conditions of maximal sensitivity across speeds form the thick curve labeled "max." The maximal sensitivity set has the shape predicted by the normative theory: the red curve in Fig 7. The mapping from measurement intervals to stimulus frequencies is explained in [27, 19]. Both panels are adopted from [31].

of resources, helping the systems adapt to the variable stimulation, a theme developed in the next section.

5 Sensor allocation

5.1 Adaptive allocation

Allocation of sensors is likely to depend on several factors that determine sensor usefulness, such as sensory tasks and properties of stimulation. For example, when the organism needs to identify rather than localize the stimulus, large sensors are more useful than small ones. Allocation of sensors by their usefulness is therefore expected to shift, for example as shown in Fig 2C.

Such shifts of allocation are expected also because the environment is highly variable. To insure that sensors are not allocated to stimuli that are absent or useless, biological systems must monitor their environment and the needs of measurement. As the environment or needs change, the same stimuli become more or less useful. The system must be able to reallocate its resources: change properties of sensors such as to better measure useful stimuli.

Because of the large but limited pool of sensors at their disposal, real sensory systems occupy the middle ground between extremes of sensor "wealth." Such systems can afford some specialization but they cannot be wasteful. They are therefore subject to Gabor's uncertainty relation, but they can alleviate consequences of the uncertainty relation, selectively and to some extent, by allocating

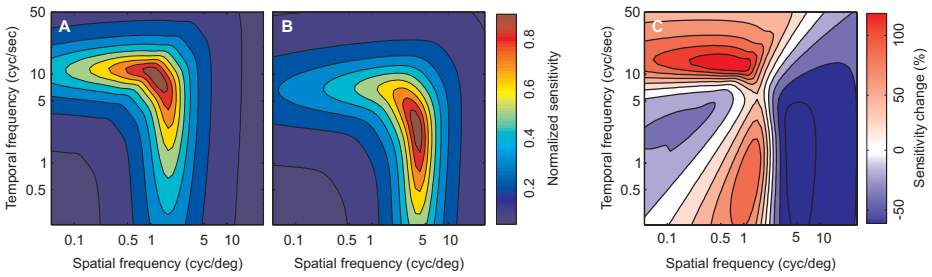

Fig. 10: Predictions for adaptive reallocation of sensors. (A–B) Sensitivity maps predicted for two stimulus contexts: dominated by high speed in A and low speed in B. The color stands for normalized sensitivity. (C) Sensitivity changes computed as $100 \times a/b$ where a and b are map entries in A and B, respectively. Here, the color stands for sensitivity change: gain in red and loss in blue.

sensors to some important classes of stimuli. Allocation preferences of such systems is expected to look like that in Fig 2C, yet generalized to multiple stimulus dimensions.

To summarize, the above analysis suggests that sensory systems are shaped by constraints of measurement and the economic constraint of limited resources. This is because the sensors of different sizes are ordered according to their usefulness in view of Gabor's uncertainty relation. These considerations are exceedingly simple in the one-dimensional analysis undertaken so far. In a more complex case considered in the next section, this approach leads to nontrivial conclusions. In particular, this approach helps to explain several puzzling phenomena in perception of motion and in motion adaptation.

A prescription has been derived for how receptive fields of different spatial and temporal extents ought to be distributed across the full range of visual stimuli. By this prescription, changes in usefulness of stimuli are expected to cause changes in receptive field allocation. Now consider some specific predictions of how the reallocation of resources is expected to bring about systematic changes in spatiotemporal visual sensitivity. Because the overall amount of resources in the system is limited, an improvement of visual performance (such as a higher sensitivity) at some conditions will be accompanied by a deterioration of performance (a lower sensitivity) at other conditions, leading to counterintuitive patterns of sensitivity change.

Assuming that equivalent amounts of resources should be allocated to equally useful stimuli, when certain speeds become more prevalent or more important for perception than other speeds, the visual system is expected to allocate more resources to the more important speeds.

For example, Fig 10A–B contains maps of spatiotemporal sensitivity computed for two environments, with high and low prevailing speeds. Fig 10C is a summary of differences between the sensitivity maps: The predicted changes form well-defined foci of increased performance and large areas of decreased per-

formance. Gepshtein et al. [37] used intensive psychometric methods [38] to measure the entire spatiotemporal contrast sensitivity function in different statistical "contexts" of stimulation. They found that sensitivity changes were consistent with the predictions illustrated in Fig 10.

These results suggest a simple resolution to some long-standing puzzles in the literature on motion adaptation. In early theories, adaptation was viewed as a manifestation of neural fatigue. Later theories were more pragmatic, assuming that sensory adaptation is the organism's attempt to adjust to the changing environment [39–42]. But evidence supporting this view has been scarce and inconsistent. For example, some studies showed that perceptual performance improved at the adapting conditions, but other studies reported the opposite [43, 44]. Even more surprising were systematic changes of performance for stimuli very different from the adapting ones [44]. According to the present analysis, such local gains and losses of sensitivity are expected in a visual system that seeks to allocate its limited resources in face of uncertain and variable stimulation (Fig 10). Indeed, the pattern of gains and losses of sensitivity manifests an optimal adaptive visual behavior.

This example illustrates that in a system with scarce resources optimization of performance will lead to reduction of sensitivity to some stimuli. This phenomenon is not unique to sensory adaptation [45]. For example, demanding tasks may cause impairment of visual performance for some stimuli, as a consequence of task-driven reallocation of visual resources [46, 47].

5.2 Mechanism of adaptive allocation

From the above it follows that the shape of spatiotemporal sensitivity function, and also transformations of this function, can be understood by studying the uncertainties implicit to visual measurement. This idea received further support from simulations of a visual system equipped with thousands of independent (uncoupled) sensors, each having a spatiotemporal receptive field [48, 49].

In these studies, spatiotemporal signals were sampled from known statistical distributions. Receptive fields parameters were first distributed at random. They were then updated according to a generic rule of synaptic plasticity [50–53]. Changes of receptive field amounted to small random steps in the parameters space, modeled as stochastic fluctuations of the spatial and temporal extents of receptive fields. Step length was proportional to the (local) uncertainty of measurement by individual receptive fields. The steps were small where the uncertainty was low, and receptive fields changed little. Where the uncertainty was high, the steps were larger, so the receptive fields tended to escape the high-uncertainty regions. The stochastic behavior led to a "drift" of receptive fields in the direction of low uncertainty of measurement [49], predicted by standard stochastic methods [54], as if the system sought stimuli that could be measured reliably (*cf.* [55]).

Remarkably, the independent stochastic changes of receptive fields (their uncoupled "stochastic tuning") steered the system toward the distribution of receptive field parameters predicted by the normative theory, and forming the

distribution observed in human vision (Fig 9). When the distribution of stimuli changed, mimicking a change of sensory environment, the system was able to spontaneously discover an arrangement of sensors optimal for the new environment, in agreement with the predictions illustrated in Fig 10 [56]. This is an example of how efficient allocation of resources can emerge in sensory systems by way of self-organization, enabling a highly adaptive sensory behavior in face of the variable (and sometimes unpredictable) environment.

5.3 Conclusions

A study of allocation of limited resources for motion sensing across multiple spatial and temporal scales revealed that the optimal allocation entails a shape of the distribution of sensitivity similar to that found in human visual perception. The similarity suggested that several previously puzzling phenomena of visual sensitivity, adaptation, and perceptual organization have simple principled explanations. Experimental studies of human vision have confirmed the predictions for sensory adaptation. Since the optimal allocation is readily implemented in self-organizing neural networks by means of unsupervised leaning and stochastic optimization, the present approach offers a framework for neuromorphic design of multiscale sensory systems capable of automated efficient tuning to the varying optical environment.

Acknowledgments

This work was supported by the European Regional Development Fund, National Institutes of Health Grant EY018613, and Office of Naval Research Multidisciplinary University Initiative Grant N00014-10-1-0072.

References

1. Marr, D.: Vision: A computational investigation into the human representation and processing of visual information. W. H. Freeman, San Francisco (1982)
2. Gabor, D.: Theory of communication. Institution of Electrical Engineers **93 (Part III)** (1946) 429–457
3. Marcelja, S.: Mathematical description of the response by simple cortical cells. Journal of the Optical Society of America **70** (1980) 1297–1300
4. MacKay, D.M.: Strife over visual cortical function. Nature **289** (1981) 117–118
5. Daugman, J.G.: Uncertainty relation for the resolution in space spatial frequency, and orientation optimized by two-dimensional visual cortex filters. Journal of the Optical Society of America A **2**(7) (1985) 1160–1169
6. Glezer, V.D., Gauzel'man, V.E., Iakovlev, V.V.: Principle of uncertainty in vision. Neirofiziologiia [Neurophysiology] **18**(3) (1986) 307–312 PMID: 3736708.
7. Field, D.J.: Relations between the statistics of natural images and the response properties of cortical cells. J. Opt. Soc. Am. A **4** (1987) 2379–2394
8. Jones, A., Palmer, L.: An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. Journal of Neurophysiology **58** (1987) 1233–1258
9. Simoncelli, E.P., Olshausen, B.: Natural image statistics and neural representation. Annual Review of Neuroscience **24** (2001) 1193–1216
10. Saremi, S., Sejnowski, T.J., Sharpee, T.O.: Double-Gabor filters are independent components of small translation-invariant image patches. Neural Computation **25**(4) (2013) 922–939
11. Gabor, D.: Lectures on communication theory. Technical report **238** (1952) Fall Term, 1951.
12. Resnikoff, H.L.: The illusion of reality. Springer-Verlag New York, Inc., New York, NY, USA (1989)
13. MacLennan, B.: Gabor representations of spatiotemporal visual images. Technical report (1994) University of Tennessee, Knoxville, TN, USA.
14. Gepshtein, S., Tyukin, I.: Why do moving things look as they do? Vision. The Journal of the Vision Society of Japan, Supp. **18** (2006) 64
15. von Neumann, J.: Zur Theorie der Gesellschaftsspiele. [On the theory of games of strategy]. Mathematische Annalen **100** (1928) 295–320 English translation in [57].
16. Luce, R.D., Raiffa, H.: Games and Decisions. John Wiley, New York (1957)
17. Watson, A.B., Ahumada, A.J.: Model of human visual-motion sensing. Journal of the Optical Society of America A **2**(2) (1985) 322–341
18. Gepshtein, S.: Two psychologies of perception and the prospect of their synthesis. Philosophical Psychology **23** (2010) 217–281
19. Gepshtein, S., Tyukin, I., Kubovy, M.: The economics of motion perception and invariants of visual sensitivity. Journal of Vision **7**(8:8) (2007) 1–18
20. Gepshtein, S., Kubovy, M.: The lawful perception of apparent motion. Journal of Vision **7**(8) (2007) 1–15 doi: 10.1167/7.8.9.
21. Gepshtein, S., Tyukin, I., Kubovy, M.: A failure of the proximity principle in the perception of motion. Humana Mente **17** (2011) 21–34
22. Korte, A.: Kinematoskopische Untersuchungen. Zeitschrift für Psychologie **72** (1915) 194–296
23. Burt, P., Sperling, G.: Time, distance, and feature tradeoffs in visual apparent motion. Psychological Review **88** (1981) 171–195

24. Wertheimer, M.: Untersuchungen zur Lehre von der Gestalt, II. Psychologische Forschung **4** (1923) 301–350
25. Kubovy, M., Holcombe, A.O., Wagemans, J.: On the lawfulness of grouping by proximity. Cognitive Psychology **35** (1998) 71–98
26. Koffka, K.: Principles of Gestalt psychology. A Harbinger Book, Harcourt, Brace & World, Inc., New York, NY, USA (1935/1963)
27. Nakayama, K.: Biological image motion processing: A review. Vision Research **25**(5) (1985) 625–660
28. Weiss, Y., Simoncelli, E.P., Adelson, E.H.: Motion illusions as optimal percepts. Nature Neuroscience **5**(6) (2002) 598–604
29. Longuet-Higgins, H.C., Prazdny, K.: The interpretation of a moving retinal image. Proceedings of the Royal Society of London. Series B, Biological Sciences **208**(1173) (1981) 385–397
30. Landy, M., Maloney, L., Johnsten, E., Young, M.: Measurement and modeling of depth cue combinations: in defense of weak fusion. Vision Research **35** (1995) 389–412
31. Kelly, D.H.: Motion and vision II. Stabilized spatio-temporal threshold surface. Journal of the Optical Society of America **69**(10) (1979) 1340–1349
32. Kelly, D.H.: Eye movements and contrast sensitivity. In Kelly, D.H., ed.: Visual Science and Engineering. (Models and Applications). Marcel Dekker, Inc., New York, USA (1994) 93–114
33. Gorban, A., Pokidysheva, L., Smirnova, E., Tyukina, T.: Law of the minimum paradoxes. Bull Math Biol **73**(9) (2011) 2013–2044
34. van Doorn, A.J., Koenderink, J.J.: Temporal properties of the visual detectability of moving spatial white noise. Experimental Brain Research **45** (1982) 179–188
35. van Doorn, A.J., Koenderink, J.J.: Spatial properties of the visual detectability of moving spatial white noise. Experimental Brain Research **45** (1982) 189–195
36. Laddis, P., Lesmes, L.A., Gepshtein, S., Albright, T.D.: Efficient measurement of spatiotemporal contrast sensitivity in human and monkey. In: 41st Annual Meeting of the Society for Neuroscience. (Nov 2011) [577.20].
37. Gepshtein, S., Lesmes, L.A., Albright, T.D.: Sensory adaptation as optimal resource allocation. Proceedings of the National Academy of Sciences, USA **110**(11) (2013) 4368–4373
38. Lesmes, L.A., Gepshtein, S., Lu, Z.L., Albright, T.: Rapid estimation of the spatiotemporal contrast sensitivity surface. Journal of Vision **9**(8) (2009) 696 http://journalofvision.org/9/8/696/.
39. Sakitt, B., Barlow, H.B.: A model for the economical encoding of the visual image in cerebral cortex. Biological Cybernetics **43** (1982) 97–108
40. Laughlin, S.B.: The role of sensory adaptation in the retina. Journal of Experimental Biology **146**(1) (1989) 39–62
41. Wainwright, M.J.: Visual adaptation as optimal information transmission. Vision Research **39** (1999) 3960–3974
42. Laughlin, S.B., Sejnowski, T.J.: Communication in Neuronal Networks. Science **301**(5641) (2003) 1870–1874
43. Clifford, C.W.G., Wenderoth, P.: Adaptation to temporal modulation can enhance differential speed sensitivity. Vision Research **39** (1999) 4324–4332
44. Krekelberg, B., Van Wezel, R.J., Albright, T.D.: Adaptation in macaque MT reduces perceived speed and improves speed discrimination. Journal of Neurophysiology **95** (2006) 255–270
45. Gepshtein, S.: Closing the gap between ideal and real behavior: scientific vs. engineering approaches to normativity. Philosophical Psychology **22** (2009) 61–75

46. Yeshurun, Y., Carrasco, M.: Attention improves or impairs visual performance by enhancing spatial resolution. Nature **396** (1998) 72–75
47. Yeshurun, Y., Carrasco, M.: The locus of attentional effects in texture segmentation. Nature Neuroscience **3**(6) (2000) 622–627
48. Jurica, P., Gepshtein, S., Tyukin, I., Prokhorov, D., van Leeuwen, C.: Unsupervised adaptive optimization of motion-sensitive systems guided by measurement uncertainty. In: International Conference on Intelligent Sensors, Sensor Networks and Information, ISSNIP 2007. 3rd, Melbourne, Qld (2007) 179–184 doi: 10.1109/ISSNIP.2007.4496840.
49. Jurica, P., Gepshtein, S., Tyukin, I., van Leeuwen, C.: Sensory optimization by stochastic tuning. Psychological Review **120**(4) (2013) 798–816 doi: 10.1037/a0034192.
50. Hebb, D.O.: The Organization of Behavior. John Wiley, New York (1949)
51. Bienenstock, E.L., Cooper, L.N., Munro, P.W.: Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex. Journal of Neuroscience **2** (1982) 32–48
52. Paulsen, O., Sejnowski, T.J.: Natural patterns of activity and long-term synaptic plasticity. Current Opinion in Neurobiology **10**(2) (2000) 172–180
53. Bi, G., Poo, M.: Synaptic modification by correlated activity: Hebb's postulate revisited. Annual Review of Neuroscience **24** (2001) 139–166
54. Gardiner, C.W.: Handbook of Stochastic Methods: For Physics, Chemistry and the Natural Sciences. Springer, New York (1996)
55. Vergassola, M., Villermaux, E., Shraiman, B.I.: 'Infotaxis' as a strategy for searching without gradients. Nature **445** (2007) 406–409
56. Gepshtein, S., Jurica, P., Tyukin, I., van Leeuwen, C., Albright, T.D.: Optimal sensory adaptation without prior representation of the environment. In: 40th Annual Meeting of the Society for Neuroscience. (Nov 2010) [731.7].
57. Taub, A.H., ed.: John von Neumann: Collected Works. Volume VI: Theory of Games, Astrophysics, Hydrodynamics and Meteorology. Pergamon Press, New York, NY, USA (1963)
58. Shannon, C.E.: A mathematical theory of communication. Bell System Technical Journal **27** (1948) 379–423, 623–656
59. Jaynes, E.T.: Information theory and statistical mechanics. Physical Review **106** (1957) 620–630
60. Gorban, A.: Maxallent: Maximizers of all entropies and uncertainty of uncertainty. Computers and Mathematics with Applications **65**(10) (2013) 1438–1456
61. Cover, T.M., Thomas, J.A.: Elements of Information Theory. John Wiley, New York (2006)

6 Appendices

6.1 Appendix 1. Additivity of uncertainty

For the sake of simplicity, the following derivations concern the stimuli that can be modeled by integrable functions $I : \mathbb{R} \to \mathbb{R}$ of one variable x. Generalizations to functions of more than one variable are straightforward. Consider two quantities:

- Stimulus location on x, where x can be space or time, the "location" indicating respectively "where" or "when" the stimulus occurred.
- Stimulus content on f_x, where f_x can be spatial or temporal frequency of stimulus modulation.

Suppose a sensory system is equipped with many measuring devices ("sensors"), each used to estimate both stimulus location and frequency content from "image" (or "input") $I(x)$. Assume that the outcome of measurement is a random variable with probability density function $p(x, f)$.

Let

$$\begin{aligned} p_x(x) &= \int p(x, f) df, \\ p_f(f) &= \int p(x, f) dx \end{aligned} \tag{A.1}$$

be the (marginal) means of $p(x, f)$ on dimensions x and f_x (abbreviated as f).

It is sometimes assumed that sensory systems "know" $p(x, f)$, which is not true in general. Generally, one can only know (or guess) some properties of $p(x, f)$, such as its mean and variance. Reducing the chance of gross error due to the incomplete information about $p(x, f)$ is accomplished by a conservative strategy: finding the *minima* on the function of *maximal* uncertainty, i.e., using a minimax approach [15, 16].

The minimax approach is implemented in two steps. The first step is to find such $p_x(x)$ and $p_f(f)$ for which measurement uncertainty is maximal. (Uncertainty is characterized conservatively, in terms of variance alone [2].) The second step is to find the condition(s) at which the function of maximal uncertainty has the smallest value: the minimax point.

Maximal uncertainty is evaluated using the well-established definition of entropy [58] (cf. [59, 60]):

$$H(X, F) = -\int p(x, f) \log p(x, f) dx\, df.$$

According to the *independence bound on entropy* (Theorem 2.6.6 in [61]):

$$H(X, F) \leq H(X) + H(F) = H^*(X, F), \tag{A.2}$$

where

$$H(X) = -\int p_x(x) \log p_x(x) dx,$$
$$H(F) = -\int p_f(f) \log p_f(f) df.$$

Therefore, the uncertainty of measurement cannot exceed

$$\begin{aligned} H^*(X, F) = &-\int p_x(x) \log p_x(x) dx \\ &-\int p_f(f) \log p_f(f) df. \end{aligned} \tag{A.3}$$

Eq. A.3 is the "envelope" of maximal measurement uncertainty: a "worst-case" estimate.

By the Boltzmann theorem on maximum-entropy probability distributions [61], the maximal entropy of probability densities with fixed means and variances is attained when the functions are Gaussian. Then, maximal entropy is a sum of their variances [61] and

$$p_x(x) = \frac{1}{\sigma_x \sqrt{2\pi}} e^{-x^2/2\sigma_x^2},$$
$$p_f(f) = \frac{1}{\sigma_f \sqrt{2\pi}} e^{-f^2/2\sigma_f^2},$$

where σ_x and σ_f are the standard deviations. Then maximal entropy is

$$H = \sigma_x^2 + \sigma_f^2. \tag{A.4}$$

That is, when $p(x, f)$ is unknown, and all one knows about marginal distributions $p_x(x)$ and $p_f(f)$ is their means and variances, the maximal uncertainty of measurement is the sum of variances of the estimates of x and f. The following *minimax* step is to find the conditions of measurement at which the sum of variances is the smallest.

6.2 Appendix 2. Improving resolution by multiple sampling

How does an increased allocation of resources to a specific condition of measurement improve resolution (spatial or temporal) at that condition? Consider set Ψ of sampling functions

$$\psi(s\sigma + \delta),\ \sigma \in \mathbb{R},\ \sigma > 0,\ \delta \in \mathbb{R},$$

where σ is a scaling parameter and δ is a translation parameter. For a broad class of functions $\psi(\cdot)$, any element of Ψ can be obtained by addition of weighted and shifted $\psi(s)$. The following argument proves that *any* function from a sufficiently

broad class that includes $\psi(s\sigma + \delta)$ can be represented by a weighted sum of translated replicas of $\psi(s)$.

Let $\psi^*(s)$ be a continuous function that can be expressed as a sum of a converging series of harmonic functions:

$$\psi^*(s) = \sum_i a_i \cos(\omega_i s) + b_i \sin(\omega_i s).$$

For example, Gaussian sampling functions of arbitrary widths can be expressed as a sum of $\cos(\cdot)$ and $\sin(\cdot)$. Let us show that, if $|\psi(s)|$ is Riemann-integrable, i.e., if

$$-\infty < \int_{-\infty}^{\infty} |\psi(s)| ds < \infty,$$

and its Fourier transform, $\widehat{\psi}$, does not vanish for all $\omega \in \mathbb{R}$: $\widehat{\psi}(\omega) \neq 0$ (i.e., its spectrum has no "holes"), then the following expansion of ψ^* is possible

$$\psi^*(s) = \sum_i c_i \psi(s + d_i) + \varepsilon(s), \tag{A.5}$$

where $\varepsilon(s)$ is a residual that can be arbitrarily small. This goal is attained by proving identities

$$\begin{aligned} \cos(\omega_0 s) &= \sum_i c_{i,1} \psi(s + d_{i,1}) + \varepsilon_1(s), \\ \sin(\omega_0 s) &= \sum_i c_{i,2} \psi(s + d_{i,2}) + \varepsilon_2(s), \end{aligned} \tag{A.6}$$

where $c_{i,1}$, $c_{i,2}$ and $d_{i,1}$, $d_{i,2}$ are real numbers, while $\varepsilon_1(s)$ and $\varepsilon_2(s)$ are arbitrarily small residuals.

First, write the Fourier transform of $\psi(s)$ as

$$\widehat{\psi}(\omega) = \int_{-\infty}^{\infty} \psi(s) e^{-i\omega s} ds,$$

and multiply both sides of the above expression by $e^{i\omega_0 \upsilon}$:

$$e^{i\omega_0 \upsilon} \widehat{\psi}(\omega) = e^{i\omega_0 \upsilon} \int_{-\infty}^{\infty} \psi(s) e^{-i\omega s} ds = \int_{-\infty}^{\infty} \psi(s) e^{-i(\omega s - \omega_0 \upsilon)} ds. \tag{A.7}$$

Change the integration variable:

$$x = \omega s - \omega_0 \upsilon \Rightarrow dx = \omega ds, \;\; s = \frac{x + \omega_0 \upsilon}{\omega},$$

such that Eq. A.7 transforms into

$$e^{i\omega_0 \upsilon} \widehat{\psi}(\omega) = \frac{1}{\omega} \int_{-\infty}^{\infty} \psi\left(\frac{x + \omega_0 \upsilon}{\omega}\right) e^{-ix} dx.$$

Notice that $\widehat{\psi}(\omega) = a(\omega) + ib(\omega)$. Hence

$$e^{i\omega_0 \upsilon}\widehat{\psi}(\omega) = e^{i\omega_0 \upsilon}(a(\omega) + ib(\omega)) = (\cos(\omega_0 \upsilon) + i\sin(\omega_0 \upsilon))(a(\omega) + ib(\omega))$$

and

$$e^{i\omega_0 \upsilon}\widehat{\psi}(\omega) = (\cos(\omega_0 \upsilon)a(\omega) - \sin(\omega_0 \upsilon)b(\omega)) + i(\cos(\omega_0 \upsilon)b(\omega) + \sin(\omega_0 \upsilon)a(\omega)).$$

Since $\widehat{\psi}(\omega) \neq 0$ is assumed for all $\omega \in \mathbb{R}$, then $a(\omega) + ib(\omega) \neq 0$. In other words, either $a(\omega) \neq 0$ or $b(\omega) \neq 0$ should hold. For example, suppose that $a(\omega) \neq 0$. Then

$$Re\left(e^{i\omega_0 \upsilon}\widehat{\psi}(\omega)\right) + \frac{b(\omega)}{a(\omega)} Im\left(e^{i\omega_0 \upsilon}\widehat{\psi}(\omega)\right) = \cos(\omega_0 \upsilon)\left(\frac{a^2(\omega) + b^2(\omega)}{a(\omega)}\right).$$

Therefore

$$\begin{aligned}\cos(\omega_0 \upsilon) = &\left(\frac{a(\omega)}{a^2(\omega) + b^2(\omega)}\right) Re\left(\frac{1}{\omega}\int_{-\infty}^{\infty} \psi\left(\frac{x + \omega_0 \upsilon}{\omega}\right) e^{-ix} dx\right) \\ &+ \left(\frac{b(\omega)}{a^2(\omega) + b^2(\omega)}\right) Im\left(\frac{1}{\omega}\int_{-\infty}^{\infty} \psi\left(\frac{x + \omega_0 \upsilon}{\omega}\right) e^{-ix} dx\right).\end{aligned} \tag{A.8}$$

Because function $\psi(s)$ is Riemann-integrable, the integrals in Eq. A.8 can be approximated as

$$Re\left(\frac{1}{\omega}\int_{-\infty}^{\infty} \psi\left(\frac{x + \omega_0 \upsilon}{\omega}\right) e^{-ix} dx\right) = \frac{\Delta}{\omega}\sum_{k=1}^{N} \psi\left(\frac{x_k + \omega_0 \upsilon}{\omega}\right)\cos(x_k) + \frac{\bar{\varepsilon}_1(\upsilon, N)}{2\omega}, \tag{A.9}$$

$$Im\left(\frac{1}{\omega}\int_{-\infty}^{\infty} \psi\left(\frac{x + \omega_0 \upsilon}{\omega}\right) e^{-ix} dx\right) = \frac{\Delta}{\omega}\sum_{p=1}^{N} \psi\left(\frac{x_p + \omega_0 \upsilon}{\omega}\right)\sin(x_p) + \frac{\bar{\varepsilon}_1(\upsilon, N)}{2\omega}, \tag{A.10}$$

where x_k and x_p are some elements of $\mathbb{R}$. To complete the proof, denote

$$\begin{aligned}&c_{2i,1} = \frac{\Delta}{\omega}\frac{a(\omega)}{a^2(\omega) + b^2(\omega)}\cos(x_i), \quad c_{2i-1,1} = \frac{\Delta}{\omega}\frac{a(\omega)}{a^2(\omega) + b^2(\omega)}\sin(x_i), \\ &d_{2i,1} = d_{2i-1,1} = \frac{x_i}{\omega}.\end{aligned}$$

From Eqs. A.8–A.10 it follows that

$$\cos(\omega_0 \upsilon) = \sum_{j=1}^{2N} c_{j,1}\psi\left(\frac{\omega_0 \upsilon}{\omega} + d_{j,1}\right) + \varepsilon_1(\upsilon, N).$$

Given that $\widehat{\psi}(\omega) \neq 0$ for all ω, and letting $\omega = \omega_0$, it follows that

$$\cos(\omega_0 \upsilon) = \sum_{j=1}^{2N} c_{j,1}\psi\left(\upsilon + d_{j,1}\right) + \varepsilon_1(\upsilon, N), \tag{A.11}$$

where

$$\varepsilon_1(\upsilon, N) = \bar{\varepsilon}_1(\upsilon, N)\frac{a(\omega_0)}{2\omega_0(a^2(\omega_0) + b^2(\omega_0))}. \tag{A.12}$$

An analogue of Eq. A.11 for $\sin(\omega_0\upsilon)$ follows from $\sin(\omega_0\upsilon) = \cos(\omega_0\upsilon + \pi/2)$. This completes the proof of Eq. A.6 and hence of Eq. A.5.