10/16/01 v. 3.3

STIMULUS REPRESENTATIONS THAT ARE INDEPENDENT OF SENSORS AND OBSERVATIONAL CONDITIONS

David N. Levin
Department of Radiology, University of Chicago

ABSTRACT

In this paper, we show how time-dependent sensory data from an evolving stimulus can be rescaled in a non-linear, time-dependent fashion in order to create a time series of stimulus representations that are invariant under any unknown invertible transformation of the sensory data. Conventional methods of multidimensional scaling and dimensional reduction do not produce data representations that are invariant in this way. Such invariant stimulus representations are unaffected by a wide variety of processes that invertibly remap sensor states, including: 1) altered performance of a device's detector, 2) changes in the observational environment external to the sensory device and the stimulus, and 3) certain modifications of the presentation of the stimuli themselves. For example, any two devices, possibly equipped with significantly different sensors, will create the same rescaled representation of an evolving stimulus, as long as they are sensitive to the same internal degrees of freedom of the stimulus. These representations are found by using affine-connected and/or Riemannian differential geometry to compute the "inner" (sensor-independent) properties of the time series of sensory data. As the "front end" of an intelligent sensory device, this kind of representation "engine" could pass rescaled stimulus representations to the device's pattern analysis module. Because the effects of many extraneous observational conditions have been "filtered out" of these representations, it would not be necessary to recalibrate the device's detectors or to retrain its pattern analysis module in order to account for these factors.

Index terms: sensor, channel, distortion, calibration, pattern recognition, computer vision, speech recognition

Send correspondence to:
David N. Levin
Department of Radiology, MC2026
University of Chicago
5841 S. Maryland Ave.
Chicago, IL 60637

Tel: 773-702-6511
Fax: 773-834-7610
Email: d-levin@uchicago.edu
Web: http://www.radiology.uchicago.edu/Levin.htm

1. INTRODUCTION

Most intelligent sensory devices contain pattern recognition software for analyzing the state of the sensors that detect stimuli in the device's environment. This software is usually "trained" to classify a set of sensor states that are representative of the "unknown" sensor states to be subsequently encountered. After these devices have been trained, their performance may be degraded if the correspondence between the stimuli and sensor states is altered by factors extrinsic to the stimuli of interest (Davies, 1990; Ponting, 1999). For example, this could happen if the response characteristics of the device's detectors are altered or if the channel between the stimulus and the detector is changed. These processes systematically deform the sensor states elicited by stimuli and thereby define a mapping of sensor states onto one another. If such transformations map one of the sensor states in the training set onto another one, the pattern recognition software may misclassify the corresponding stimuli. Likewise, the device may not recognize a stimulus in the training set if it's original sensor state has been transformed into one outside of the training set. These problems can be addressed by periodically recalibrating the device to account for sensor state transformations due to changed conditions. For example, the device can be exposed to a test pattern that should produce a known sensor state under "normal" conditions. Then, the observed differences between the actual and ideal sensor states for this test stimulus can be used to correct subsequently encountered sensor states. This procedure must be implemented after each change in observational conditions in order to keep track of time-dependent distortions. Because the device may not be able to detect the presence of such a change, it may be necessary to recalibrate it at short fixed intervals. However, this will tend to decrease the device's duty cycle by frequently taking it "off-line". Furthermore, the recalibration process may be logistically impractical in some situations (e.g., computer vision and speech recognition devices at remote locations).

In contrast, humans form percepts that are remarkably independent of the nature of their sensors and observational conditions. This was strikingly illustrated by experiments in which subjects wore goggles creating severe geometric distortions of the observed scene (Stratton, 1896, 1897a, and 1897b; Gibson, 1933; Held, 1972). For example, the visual input of some subjects was warped non-linearly,

inverted, and/or reflected from right to left. Although the subjects initially perceived the distortion, their perceptions of the world returned to the pre-experimental baseline after several weeks of constant exposure to familiar stimuli seen through the goggles. For instance, lines reported to be straight before the experiment were initially perceived to be warped, but these lines were once again reported to be straight after several weeks of viewing familiar scenes through the distorting lenses. Similar results were observed when the goggles were removed at the end of the experiment. Namely, the world initially appeared to be distorted in a manner opposite to the distortion due to the lenses, but eventually no distortion was perceived. These experiments suggest that humans utilize recent sensory experiences to "normalize" their perception of subsequent sensory data, in a way that "filters out" the effects of systematic transformations of sensory data. This impression is reinforced by the fact that different persons tend to share similar perceptions of the world, despite obvious differences in their sensory organs and processing pathways. This "universality" of perception may be due to the apparent ability of each individual to "filter out" the effects of systematic sensor state transformations, including the transformations relating his/her sensor states to those of other individuals.

In this paper, we show how to build sensory devices that mimic this sensor-independent characteristic of human perception. *Specifically, we demonstrate how time-dependent sensory data from an evolving stimulus can be rescaled in a non-linear, time-dependent fashion in order to create a time series of stimulus representations that are invariant under any unknown invertible transformation of the sensory data.* As a consequence, the same rescaled representation of an evolving stimulus will be created by two devices that are equipped with significantly different sensors, as long as both devices are sensitive to the same internal degrees of freedom of the stimulus. To see this, consider any sensory device that consistently and sensitively detects the state of the d degrees of freedom of a stimulus. Consistency and sensitivity imply that: 1) each stimulus configuration induces one and only one device "sensor state" (the internal parameters that are produced from the possibly processed output of the device's detectors); 2) each sensor state is induced by one and only one configuration of the stimulus. This correspondence between stimulus configurations and induced sensor states defines a time-independent invertible

transformation between the *d*-dimensional manifold of stimulus configurations and a *d*-dimensional manifold of device sensor states. It follows that the sensor state manifolds of any two such devices must be related by a time-independent invertible transformation, which maps each sensor state of one device onto the sensor state of the other device that is produced by the same stimulus configuration. The existence of such a transformation implies that these devices will create the same stimulus representations if they rescale their time-dependent sensor states by the process demonstrated in this paper. As an illustration, consider computer vision devices that are designed to detect the expressions of a particular face, and suppose that the configurations of that face form a 2D manifold. For instance, this would be the case if each facial expression is defined by the configurations of the mouth and eyes and if these features are controlled by two parameters. Suppose that sensor state x of computer vision system V consists of the amplitudes of two particular coefficients in the 2D Fourier expansion of the face. In order for this sensor state to sensitively and consistently reflect the configuration of the evolving face, there must be a time-independent invertible mapping between x and the manifold of facial expressions (i.e., between x and the two parameters controlling the expressions). Now, consider another computer vision system V', which has a sensor state x' consisting of two other features of the image. For example, V' might be programmed to compute the amplitudes of two specific coefficients in the 2D Bessel expansion of the facial image. Alternatively, V' might be identical to V, except for the fact that its camera observes the face through a warping lens, so that it records two Fourier components of the warped face. If the internal state of V' sensitively reflects the facial configuration, there must be a time-independent invertible mapping between x' and the manifold of facial expressions. It follows that there is a time-independent invertible mapping between sensor state x of system V and sensor state x' of system V' that maps $x(t)$ onto $x'(t)$ as the two devices observe an evolving facial expression. Therefore, if each of these vision systems rescales its sensor states as described in this paper, they will independently derive identical rescaled representations of the evolving facial expression, despite the dramatic differences between their detectors. Thus, the rescaling process enables devices of this kind to "see" the world in the same sensor-independent way.

Notice that any physical process that invertibly transforms the sensor states of a device will not affect the transformation-independent stimulus representations described in this paper. Transformative processes of this kind include: 1) altered performance of the device's detectors (e.g., altered gain curve of a detector circuit or distortion of an electronic image in a camera), 2) certain alterations of observational conditions that are external to the detectors and the stimuli (e.g., different intensity of a scene's illumination or different positioning of the detectors with respect to the stimuli), 3) systematic modifications of the presentation of the stimuli themselves (e.g., systematic warping of printed pages or systematic morphing of a voice). Therefore, if the pattern analysis techniques of an intelligent sensory device are applied to rescaled stimulus representations, instead of the "raw" (unrescaled) sensor states, the device will not have to be recalibrated and/or retrained in order to account for these extraneous processes (Davies, 1990).

The method in this paper differs significantly from other techniques for representing data, such as multidimensional scaling or dimensional reduction (Shepard, 1962; Carroll, 1980; Cox, 1994). In each of these approaches, it is necessary to impose an *ad hoc* measure of "distance" between each pair of neighboring data points (e.g., Tenenbaum, 2000) or, at least, to rank the distances between pairs of neighboring points (e.g., Holman, 1978; Roweis, 2000). In each case, the defined distances or distance rankings are not invariant under general, non-linear coordinate transformations. Therefore, the scale values assigned to each data point do not share the transformation-independence of the rescaled representations described in this paper. However, it should also be mentioned that multidimensional scaling methods are applicable to more general data sets than the technique in this paper. Specifically, multidimensional scaling can be applied to data that do not form a time series, unlike the method in this paper.

As mentioned above, an invertible transformation relates the sensor states of two devices that detect the same internal degrees of freedom of a stimulus. Therefore, the task of finding sensor-independent stimulus representations is mathematically equivalent to the task of creating transformation-independent stimulus representations; i.e., representations that are unaffected by invertible

transformations of the sensor states from which the representations are derived. As shown in Section 2, a time series of sensor states defines a "natural" scale or coordinate system on the sensor state manifold, and each sensor state in the time series can be represented by its location on that scale. This rescaled representation of a sensor state is invariant if all of the sensor states in the time series are subjected to the same invertible transformation. This is because the relationship between each untransformed sensor state and the scale derived from the untransformed sensor state time series is the same as the relationship between the corresponding transformed sensor state and the scale derived from the transformed time series. This is analogous to the fact that the physical rotation or translation of a collection of particles in a plane does not alter the coordinates of each particle in the collection's "natural" internal coordinate system (its "center-of-mass" coordinate system), even though the absolute coordinates of each particle are transformed. This is because the rotation/translation transforms the internal coordinate system and each particle's absolute coordinates in the same way, without disturbing their relationship.

In Section 2, we demonstrate two methods of finding a transformation-independent representation of each sensor state in a sensor state time series (Levin, 2000b and 2001a). The technique in Section 2.B utilizes affine-connected differential geometry, and the method in Section 2.C is based on Riemannian differential geometry. The first of these techniques is illustrated with analytical and numerical examples in Section 3. The implications of these results are discussed in Section 4.

2. THEORY

2.A. Transformation-Independent Descriptions of Sensor States

Consider any sensory system having detectors that are sensitive to various features of an evolving stimulus. These detectors may send their outputs to a processing unit that combines them in a linear or non-linear fashion. This processing may also include dimensional reduction procedures that map the detector outputs onto a smaller number of quantities (Roweis, 2000; Tenenbaum, 2000). For example, in an imaging system, the processing units may extract the time-dependent locations or intensities of particular image features. In a speech recognition system, the processing units could compute parameters

of the signal's short-term Fourier spectrum, such as peak frequencies or amplitudes, cepstral parameters, etc (Rabiner, 1993). Let the device's "sensor state" x denote the array of numbers x_k ($k = 1, \ldots, N, N \geq 1$) that form the output of the processing unit, and let $x(t)$ denote the time series of sensor states produced by an evolving stimulus.

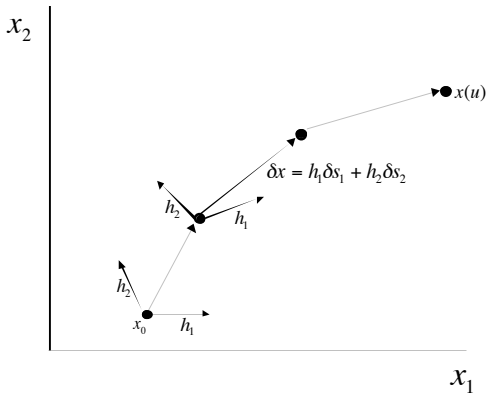

Figure 1. Consider a path $x(u)$ ($0 \leq u \leq 1$) between a reference sensor state x_0 and a sensor state of interest. If vectors h_a can be defined at each point along the path, each line segment δx can be decomposed into its components δs_a along the vectors at that point.

Our goal is to create a representation of each of the observed sensor states that is unaffected by invertible sensor state transformations. Such a transformation relabels each sensor state in a way that is mathematically equivalent to a change of coordinates on the sensor state manifold. Therefore, our task is to create a *coordinate-independent* representation of each sensor state in the time series; i.e., a representation of the sensor states that is independent of the nature of the x coordinate system, which we happen to be using to label them. Such a coordinate-independent description can be created with the help of coordinate-independent ways of identifying: 1) a reference sensor state (x_0) that serves as the origin of the scale for representing sensor states, 2) a path $x(u)$ ($0 \leq u \leq 1$) through the manifold of sensor states that connects the reference sensor state to a sensor state of interest x ($x(0) = x_0$, $x(1) = x$), 3) N linearly-independent contravariant vectors h_a ($a = 1, \ldots, N$) at each point along the path. Here, a vector h is said to be contravariant if it transforms as $h \rightarrow h' = \frac{\partial x'}{\partial x} h$ under the change of coordinate systems $x \rightarrow x'$ (Schrodinger, 1963; Weinberg, 1972). If the foregoing conditions are met, each infinitesimal segment δx along the path can be decomposed into its components δs_a along the vectors h_a (Figure 1):

$$\delta x = \sum_{a=1,\ldots,N} h_a \, \delta s_a \qquad \text{(Eq. 1)}$$

Note that δs is a coordinate-independent (scalar) quantity because δx and h_a are contravariant vectors. Therefore, if the components δs are integrated over the specified path connecting x_0 and x, the result is a coordinate-independent description of the sensor state x (Levin, 2000a):

$$s = \int_{x_0}^{x} \delta s \qquad \text{(Eq. 2)}$$

Essentially, the vectors h_a describe a local coordinate system that is determined by the intrinsic structure of the sensor state time series. When these local coordinate systems are "stitched together", they define the function $s(x)$, which constitutes a global coordinate system or sensor state scale that is determined by the sensor state time series, in the same way that a global "center-of-mass" coordinate system is determined by the coordinates of the particles in a collection. This scale is coordinate-independent in the sense that it is unaffected by linear or non-linear transformations of the extrinsic coordinate system that is used to label the sensor states in the time series. This is analogous to the fact that the center-of-mass coordinate system of a particle collection is unaffected by rotations or translations of the extrinsic coordinate system used to label the particles.

In the following two Sections, the sensor states in a chosen time interval are used to derive a coordinate-independent "parallel transport" operation on the sensor state manifold (i.e., a way of moving vectors across the manifold). Then, as long as contravariant vectors h_{0a} (called reference vectors) can be derived at the reference sensor state x_0, these vectors can be moved across the manifold in a coordinate-independent manner in order to create: 1) "canonical" paths between x_0 and all other points; 2) vectors h_a at all other points. Equations 1-2 can then be used to create coordinate-independent descriptions of sensor states. Therefore, once a parallel transport operation has been derived from the sensor state history of the device, a sensor state scale can be derived from coordinate-independent knowledge of a reference sensor state x_0 and reference vectors h_{0a} at x_0. As discussed below, this reference information may be deduced from coordinate-independent features of the sensor state time series, or it may be derived from prior knowledge.

The parallel transport operation can be derived from the local structure of the sensor state time series in any chosen time interval of sufficient length. For instance, parallel transport can be derived dynamically (i.e., at each time point) from the sensor states encountered in the most recent time interval of length ΔT. Here, ΔT is a user-determined parameter that influences the adaptivity and noise sensitivity of the method. Specifically, in Section 2.B, the sensor state time series is used to define an affine connection on the sensor state manifold. In Section 2.C the sensor state time series is shown to impose a Riemannian metric on the manifold, and an affine connection is then derived from this metric. In both approaches, the resulting affine connection is used to parallel transport the reference vectors across the manifold in a coordinate-independent way.

2.B. Sensor State Manifolds That Support an Affine Connection

The time series of sensor states $x(t)$ describes a trajectory that crosses the sensor state manifold. According to the methods of affine-connected differential geometry (Schrodinger, 1963; Weinberg, 1972), any vector can be moved across this manifold in a coordinate-independent manner if one can define a local affine connection $\Gamma^{k}_{lm}(x)$, which is a quantity transforming as:

$$\Gamma'^{\,k}_{lm} = \sum_{r,s,t=1,\ldots,N} \frac{\partial x'_{\,k}}{\partial x_r} \frac{\partial x_s}{\partial x'_{\,l}} \frac{\partial x_t}{\partial x'_{\,m}} \Gamma^{r}_{st} + \sum_{n=1,\ldots,N} \frac{\partial x'_{\,k}}{\partial x_n} \frac{\partial^2 x_n}{\partial x'_{\,l}\, \partial x'_{\,m}} \qquad (3)$$

Specifically, given any contravariant vector *V* at *x*, consider the array of numbers $V + \delta V$ where:

$$\delta V^k = - \sum_{l,m=1,\ldots,N} \Gamma^{k}_{lm} V^l \, \delta x_m \qquad (4)$$

It can be shown that $V + \delta V$ transforms as a contravariant vector at the point $x + \delta x$, as long as the affine connection transforms as shown in Eq.(3). The vector $V + \delta V$ at $x + \delta x$ is said to be the result of parallel transporting *V* along δx. Our task is to use the past history of sensor states *x(t)* to derive an affine connection on the sensor state manifold. Then, given a set of vectors h_a at just one point on the manifold (e.g., at the reference state x_0), we will be able to use the affine connection to populate the entire

manifold with parallel-transported versions of those vectors. These can be used in Eqs.(1-2) to derive a coordinate-independent representation of any sensor state.

Consider a point x that is on at least $N(N+1)/2$ trajectory segments. Each of these segments can be divided into infinitesimal line elements that correspond to equal infinitesimal time intervals. These line elements transform as contravariant vectors. Therefore, we can look for affine connections that parallel transport a given line element along itself into the next line element on the same trajectory segment. In other words, we can look for affine connections for which a given trajectory segment is locally geodesic (Schrodinger, 1963; Weinberg, 1972). Equation 4 implies that such an affine connection $\hat{\Gamma}^k_{lm}$ must satisfy the following N constraints:

$$\delta dx^k = -\sum_{l,m=1,\ldots,N} \hat{\Gamma}^k_{lm} dx_l dx_m \qquad (5)$$

where $dx + \delta dx$ represents the trajectory's line element at $x + dx$. Now consider any collection of $N(N+1)/2$ of the trajectory segments at x. An affine connection that makes all of these trajectory segments locally geodesic must satisfy $N^2(N+1)/2$ linear constraints like those in Eq.(5). Because a symmetric affine connection ($\Gamma^k_{lm} = \Gamma^k_{ml}$) has $N^2(N+1)/2$ components, one and only symmetric connection satisfies these equations unless they happen to be inconsistent (no solutions) or redundant (multiple solutions). Notice that if $\hat{\Gamma}^k_{lm}$ is a solution of these equations in one coordinate system, then $\hat{\Gamma}'^k_{lm}$ is a solution of the corresponding equations in any other coordinate system, where $\hat{\Gamma}^k_{lm}$ and $\hat{\Gamma}'^k_{lm}$ are related by Eq.(3). Therefore, if these equations have a unique solution in one coordinate system, there is a unique solution of the corresponding equations in any other coordinate system, and these solutions are related by Eq.(3). Now, consider all collections of $N(N+1)/2$ trajectory segments that have a unique solution to these equations; i.e. all collections that are locally geodesic with respect to one and only one symmetric affine connection at x. Let Γ^k_{lm} be the average of the affine connections computed from these subsets of trajectory segments:

$$\Gamma^k_{lm} = \frac{1}{N_T} \sum_{i=1,\ldots,N_T} \hat{\Gamma}^k_{lm}(i). \qquad (6)$$

where $\hat{\Gamma}^k_{lm}(i)$ is the symmetric affine connection that makes the i^{th} collection of trajectory segments locally geodesic and N_T is the number of such collections. The quantity Γ^k_{lm} transforms as shown by Eq.(3) because each contribution to the right side of Eq.(6) transforms in that way. Therefore, Γ^k_{lm} can be defined to be *the* affine connection at point *x* on the sensor state manifold.

Now, suppose that *N* linearly independent reference vectors h_{0a} can be defined at a point x_0 on the manifold by one of the methods discussed below. The above-described affine connection can be used to parallel transport these vectors to any other point *x* on the manifold. The resulting vectors at *x* will depend on the path that was used to create them if the manifold has non-zero curvature; i.e., if the curvature tensor B^k_{lmn} is non-zero at some points, where:

$$B^k_{lmn} = -\frac{\partial \Gamma^k_{lm}}{\partial x_n} + \frac{\partial \Gamma^k_{\ln}}{\partial x_m} + \sum_{i=1,\ldots,N} \left(\Gamma^k_{im} \Gamma^i_{\ln} - \Gamma^k_{in} \Gamma^i_{lm} \right) \qquad (7)$$

Because this tensor will not vanish in many cases, the path connecting x_0 and *x* must be completely specified in order to unambiguously define vectors at *x*. Such a path can be prescribed in the following coordinate-independent fashion. Generate a trajectory through x_0 by repeatedly parallel transporting the vector h_{01} along itself, and call this trajectory a type 1 geodesic. Next, parallel-transport all of the vectors h_{0a} along this trajectory in order to define vectors h_a there. Now, generate a type 2 geodesic through each point of this geodesic by repeatedly parallel transporting the vector h_2 along itself. Then, parallel-transport all of the vectors h_a along each of these geodesics, and generate a type 3 geodesic through each point on each type 2 geodesic by repeatedly parallel transporting the vector h_3 along itself. Continue in this manner until type *N* geodesics have been generated through each point on each type *N-1* geodesic. Because of the linear independence of the vectors h_{0a} at x_0, the parallel transported h_a will also be linearly independent. It follows that the collection of points on all trajectories of type *n* comprises

an *n*-dimensional subspace of the manifold, and the type *N* trajectories will reach every point on the manifold. This means that any point *x* can be reached from x_0 by following a "canonical" path consisting of a segment of the type 1 geodesic, followed by a segment of a type 2 geodesic, … followed by a segment of a type *N* geodesic. This path specification is coordinate-independent because it is defined in terms of a coordinate-independent operation: namely, the parallel transport of vectors. After the h_a have been "spread" to the rest of the manifold along these paths, a coordinate-independent representation s_a of any point *x* can be generated by integrating Eq.(2) along the "canonical" path between x_0 and *x*.

In order to visualize the entire procedure described above, consider a sensor state manifold containing a single point x_0 at which vectors h_{0a} are specified; e.g., a plane or a small patch of a sphere that is intrinsically "marked" at x_0 with two "pointers" (reference vectors) oriented in preferred directions (called the "north" and "east" directions). If the time series of recently encountered sensor states covers the manifold sufficiently densely, Equation (6) can used to derive the affine connection at each point from the observed evolution of sensor states through it. For instance, if the sensor states move at constant speed along straight lines in the plane or along great circles on the sphere, Eq.(6) leads to the usual parallel transport rules of Riemannian geometry on a plane or sphere. The resulting affine connection can be used to parallel transport the east pointer along itself in order to create a "east-west" geodesic through x_0 (a straight line or a great circle in the above examples). We can then parallel transport the north pointer to create a new north pointer at each point along this east-west geodesic. Finally, we can parallel transport each north pointer along itself in order to create a north-south geodesic through it. Each point on the manifold can then be represented by s_a, which represents the number of parallel transport operations (east or west, followed by north or south) that are required to reach it from x_0. If the manifold is a plane with the above-described straight sensor trajectories and if the "north"/"east" pointers at x_0 are orthogonal, s_a will represent each point in a Cartesian coordinate system. On the other hand, if the manifold is a sphere with the above-described great circular sensor state trajectories, s_a will represent each point by its longitude and latitude. In each case, the resulting

representation does not depend on which coordinate system was originally used to record sensor states and to derive the affine connection.

Notice that this method requires coordinate-independent knowledge of a special sensor state (the reference state) and special vectors (reference vectors) at that point. These might be identified as coordinate-independent features of the history of sensor states. For example, x_0 could be the sensor state at the maximum of the function defined by the number of times each sensor state has been encountered in a chosen time interval. If the sensor state trajectory has preferred directions as it repeatedly passes through x_0, these features may be used to define reference vectors there, as shown in Levin (2001a). Alternatively, the reference state and reference vectors could also be specified by some prior knowledge. For example, we might know *a priori* that: 1) x represents the intensity of a pixel in a digital image, 2) the device was originally calibrated so that $x=0$ corresponds to no light emission/reflection at the pixel's location, and 3) we are primarily concerned about transformations caused by changes of illumination intensity. Because changes in illumination intensity are expected to map the null sensor state onto itself, we only need to be concerned about coordinate transformations with that fixed point. Therefore, in this case, prior knowledge tells us that the null point corresponds to the same sensor state in every coordinate system of interest, and, therefore, it can be chosen to be the reference state. This might also be the case if the transformations of interest were known to reflect differences in the gain curves of the detectors of different devices. Finally, the reference information may be determined by "calibration" stimuli that the user explicitly "shows" to the device. For example, the reference sensor state may be chosen to be the sensor state produced by a user-determined stimulus. Recall that the reference sensor state serves as the origin of the scale function used to rescale sensor states. Therefore, this last procedure is analogous to having a choir leader play a note on a pitch pipe in order to "show" each singer the origin of the desired musical scale. Finally, notice that stimulus representations that are referred to different reference stimuli will reflect different "points of view". For example, suppose that a device is observing a glass of

beverage. It will "perceive" the glass to be half full or half empty if it uses reference sensor states corresponding to an empty glass or a full glass, respectively.

Strictly speaking, the affine connection Γ^k_{lm} must be computed from sensor data at every point on the path used in Eq.(2). This means that the sensor state trajectory $x(t)$ must cover the manifold densely so that it passes through each of these points at least $N(N+1)/2$ times. However, this requirement can be relaxed for most applications. Specifically, suppose that Γ^k_{lm} is only computed at a finite collection of sample points on the manifold, and suppose that it is computed from trajectory segments passing through a very small neighborhood of each sample point (not necessarily through the point itself). Furthermore, suppose that values of Γ^k_{lm} at intervening points are estimated by parametric or non-parametric interpolation (e.g., splines or neural nets, respectively). This method of computation will be accurate as long as the spacing between the sample points and the size of the small neighborhoods around them are small relative to the distance over which the manifold's affine connection varies. In order to satisfy these conditions, the sensor state trajectory must cover the manifold with a sufficiently high density. In other words, the device must have sufficient "experience" in order to derive coordinate-independent representations of sensor states.

Finally, in the above discussion, the sensor state transformation was assumed to be time-independent; e.g., it was assumed that the sensor states in two devices were related to one another in a time-independent fashion. Now, consider the effects of the sudden onset of a sensor state transformation, and suppose that the rescaling of the signal is determined by signal levels encountered in the most recent period of length ΔT. During a transitional period of length ΔT after the transformation's onset, the sensory device will record a mixture of untransformed and transformed signal levels. During this transition, the device's scale function will evolve from the form derived from untransformed signals to the form derived from transformed signals, and during this transitional period the transformed sensor states may be represented differently than the signals at corresponding times in the untransformed time series. However, once ΔT time units have elapsed since the transformation's onset, the device's scale function

will be wholly derived from a time series of transformed sensor states. Thereafter, transformed signal levels will again be represented in the same way as the signals at corresponding times in the untransformed time series. This phenomenon has been explicitly demonstrated in numerical experiments reported elsewhere (Levin, 2001b). Thus, like a human, the system adapts to the presence of the transformation after a period of adjustment.

2.C. Sensor State Manifolds That Support Metrics

In this section, we show how the sensor state trajectory can impose a Riemannian metric on the manifold, which can then be used to define a parallel transport operation. Parallel transport makes it possible to move the reference vectors from x_0 in order to define paths on the manifold and in order to create vectors h_a at other points. Then, Eqs.(1-2) can be used to create coordinate-independent descriptions of sensor states.

As before, let the sensor state be an array of numbers x_k ($k = 1, \ldots, N$, $N \geq 1$), and let $x(t)$ be the sensor state trajectory during a chosen time interval (e.g., the most recent time interval of length ΔT). Consider a point x that is on at least $N(N+1)/2$ trajectory segments. Each of these segments defines an infinitesimal line element $dx = x(t + dt) - x(t)$, where t is the time at which the trajectory segment passed through x and dt is an infinitesimal time interval. Now consider one of these line elements, and look for metrics that assign unit length to it. Such a metric $\hat{g}_{kl}$ must satisfy the following constraint:

$$\sum_{k,l=1,\ldots,N} \hat{g}_{kl} dx_k dx_l = 1 \qquad (8)$$

Next, consider any collection containing $N(N+1)/2$ of the line elements at x. A metric that assigns unit length to all of these line elements must satisfy $N(N+1)/2$ linear constraints like the one in Eq.(8). Because a metric has $N(N+1)/2$ components, one and only one metric satisfies these equations unless they happen to be inconsistent (no solutions) or redundant (multiple solutions). If these equations have a unique solution in one coordinate system, there is a unique solution of the corresponding equations in any

other coordinate system, and these solutions define the same covariant tensor in the different coordinate systems. This is a consequence of the fact that each line element *dx* transforms as a contravariant vector. Now, consider all collections of $N(N+1)/2$ line elements that have a unique solution to these equations. Let g_{kl} be the average of the metrics computed from these subsets of line elements:

$$g_{kl} = \frac{1}{N_L} \sum_{i=1,\ldots,N_L} \hat{g}_{kl}(i). \qquad (9)$$

where $\hat{g}_{kl}(i)$ is the metric that assigns unit length to the i^{th} collection of line elements and N_L is the number of such collections. Note that sets of line elements for which Eq.(8) has no solution or multiple solutions do not contribute to Eq.(9). The quantity g_{kl} transforms as a covariant tensor because each contribution to the right side of Eq.(9) transforms in that way. Therefore, g_{kl} can be defined to be *the* metric at point *x* on the sensor state manifold.

There are several ways of using the above-defined metric to define parallel transport on the sensor state manifold. For example, the following quantity is a symmetric affine connection that preserves the metrically computed lengths of vectors during parallel transport:

$$\Gamma^k_{lm} = \frac{1}{2} \sum_{n=1,\ldots,N} g^{kn} \left(\frac{\partial g_{mn}}{\partial x_l} + \frac{\partial g_{nl}}{\partial x_m} - \frac{\partial g_{lm}}{\partial x_n} \right) \qquad (10)$$

where g^{kl} is the contravariant tensor that is the inverse of g_{kl}. Other definitions of the affine connection are also possible (Schrodinger, 1963). Now, suppose that *N* linearly independent vectors h_{0a} can be defined at a point x_0 on the manifold. For example, the method in Levin (2001a) can be used to derive these vectors from the sensor state trajectory in the vicinity of x_0 if it has intrinsic directionality there. The above-described affine connection can be used to parallel transport these vectors to any other point *x* on the manifold. As in Section 2.B, the resulting vectors will depend on the path that was used to create them if the manifold has non-zero curvature. Therefore, in general, the path connecting x_0 and *x* must be completely specified in order to unambiguously define vectors at *x*. Such a path can be prescribed exactly as it was in Section 2.B. Namely, we can define a "canonical" path to *x* that follows a specific sequence

of geodesics, which are created by parallel transport of the vectors h_{0a} at x_0. Then, a coordinate-independent representation s_a of sensor state x can be generated by integrating Eq.(2) along the canonical path between x_0 and x.

As in Section 2.B, the affine connection (and, therefore, the metric) must be computed from sensor state data at every point on the path used in Eq.(2). This means that the sensor state trajectory $x(t)$ must cover the manifold densely so that it passes through each of these points at least $N(N+1)/2$ times. However, this requirement can usually be relaxed by computing the metric at a finite collection of sample points from trajectory segments passing through a very small neighborhood of each sample point (not necessarily through the point itself). Then, the values of g_{kl} at intervening points can be estimated by parametric or non-parametric interpolation (e.g., splines or neural nets, respectively). As before, this method of computation will be accurate as long as the distance between sample points and the size of the small neighborhoods around them are small relative to the distance over which the metric changes.

3. ANALYTICAL AND NUMERICAL EXAMPLES

3.A. An Analytical Example on a One-Dimensional Sensor State Manifold

It is useful to illustrate these results with a simple example. Suppose the untransformed signal $x(t)$ is a long periodic sequence of triangular shapes, like those in Fig. 2a. For example, if the sensor state represents the intensity of a pixel in a digital image of a scene, Figure 2a might be its response to a series of identical objects passing through the scene at a constant rate. Alternatively, if the sensor state represents the amplitude of a microphone's output, Figure 2a might be its response to a series of uniformly spaced identical pulses. Let a and b be the slopes of the lines on the left and right sides, respectively, of each shape; Fig. 2a shows the special case: $a = 0.6$ and $b = -0.3$ (measured in inverse time units). Our task is to use the method of Section 2.B to rescale this signal at each time point t with respect to a coordinate-independent sensor state scale, derived from the sensor state history during the

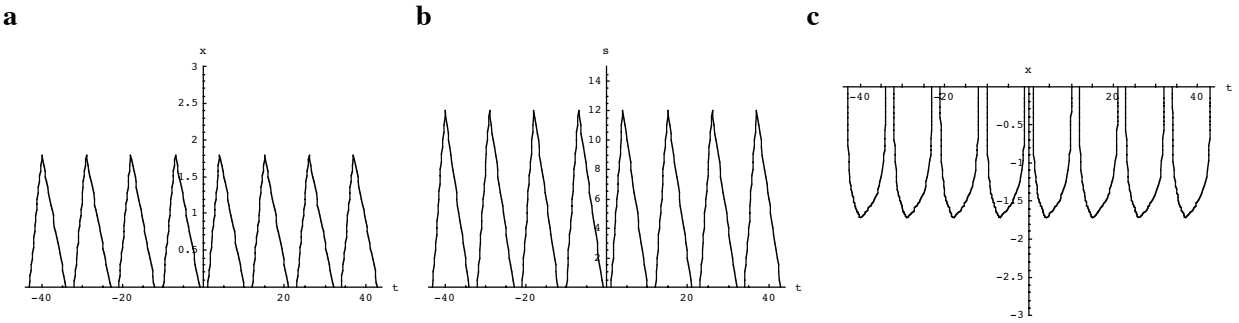

Figure 2. a) An untransformed signal $x(t)$ describing a long succession of identical pulses that are uniformly spaced in time. b) The signal representation $S(t)$ that results from applying the rescaling method in Section 2.B either to the signal in *a* or to the transformed version of that signal in *c*. c) The signal obtained by subjecting the signal in *a* to the transformation: $x'(x) = g_1 \ln(1 + g_2 x)$ where $g_1 = -0.3$ and $g_2 = 170$.

most recent time interval of length ΔT. First, notice that Eqs.(5, 6) show that the affine connection at a point *y* on a one-dimensional sensor state manifold is found by averaging the values of:

$$\hat{\Gamma}(i) = -\frac{\frac{d^2x}{d^2t}}{\left(\frac{dx}{dt}\right)^2} \qquad \text{Eq.(11)}$$

where the right side is evaluated at times $t_i \varepsilon [t - \Delta T, t]$ at which $x(t_i) = y$. Because the second derivative of the time series in Fig. 2a vanishes at all relevant times, $\Gamma = 0$ at each *y*. Now suppose that the reference sensor state is taken to be the null signal ($x_0 = 0$) and the reference vector is chosen to be $h_0 = (a + b)/2$. As mentioned previously, this reference information could have been determined by prior knowledge. For example, the user could have "shown" the device a stimulus that produces a null sensor state having a time derivative equal to h_0. The derived affine connection can be used to move this reference vector across the manifold in order to define a vector $h(y)$ at any other point *y*. This is done by means of Eq.(4), which has the form on one-dimensional manifolds:

$$\frac{dh}{dy} = -\Gamma(y)h \qquad \text{Eq.(12)}$$

In this example, it follows that $h(y) = (a + b)/2$. In the case of a one-dimensional manifold, Eqs.(1, 2) imply:

$$s = \int_{x_0}^{x} \frac{dy}{h(y)} \qquad \text{Eq.(13)}$$

Therefore, in this example, $s(x) = 2x/(a + b)$, and the rescaled signal at each time is $S(t) = s[x(t)] = 2x(t)/(a + b)$, which is shown in Fig. 2b. Now, consider the transformed signal that is related to the Fig. 2a by any of the following non-linear functions: $x'(x) = g_1 \ln(1 + g_2 x)$ where $g_2 > 0$. For example, if $g_1 = -0.3$ and $g_2 = 170$, the transformed signal $x'(t)$ looks like Figure 2c. For instance, in the above-mentioned examples, this could represent the pixel or sound intensity in a second sensory device that has a gain curve non-linearly related to the gain curve of the first device. When Eq.(11) is used to compute the affine connection from the transformed signal, we get $\Gamma' = 1/g_1$ at each point y' on the sensor state manifold. The transformed (coordinate-independent) forms of the reference sensor state and reference vector are $x_0' = 0$ and $h_0' = g_1 g_2 (a + b)/2$, respectively. As before, these are assumed to be determined by prior knowledge (e.g., by the above-mentioned specification of the user). Equation 12 shows that the affine connection parallel transports the reference vector into the vector $h'(y') = \frac{1}{2} g_1 g_2 (a + b) e^{-y'/g_1}$ at point y'. Then, Eq.(13) gives the scale function that is inherent in the transformed signal: $s'(x') = \frac{2(e^{x'/g_1} - 1)}{g_2 (a + b)}$. By substitution, it is easy to see that the rescaled version of the transformed signal is the same as the rescaled version of the untransformed signal; i.e., $S'(t) = s'[x'(t)] = s[x(t)] = S(t)$. Thus, both signals rescale to the form in Fig. 2b. In other words, the rescaled signal $S'(t)$, which is derived from the transformed signal $x'(t)$, is the same as the rescaled signal $S(t)$, which is derived from the untransformed signal $x(t)$. This is because the effect of the invertible signal transformation on the signal level at any given time ($x(t) \rightarrow x'(t)$) is compensated by its effect on the form of the scale function at that time ($s(x) \rightarrow s'(x')$). Notice that $s(x)$ and $s'(x')$ (as

well as $h(y)$, $h'(y')$, Γ, and Γ') happen to be time-independent in this particular example, and this implies that $x(t)$ and $x'(t)$ are rescaled in a time-independent fashion. In the general case, the scale functions depend on time in a manner dictated by the earlier time course of the signal. However, identical self-scaled signals (i.e., $S(t) = S'(t)$) will still be derived from the untransformed and transformed signals, as demonstrated by the proof in Section 2.B.

3.B. A Numerical Example on a Two-Dimensional Sensor State Manifold

Let $x = (x_1, x_2)$ represent the sensor state of a device, and suppose that Fig. 3a shows the trajectory of sensor states during the most recent period of length ΔT. For example, these numbers might be the coordinates of a specific feature being tracked in a time series of digital images, or they could be the amplitudes or frequencies of peaks in the short-term Fourier spectrum of an audio signal. Notice that these trajectory segments happen to be straight lines that are traversed at constant speed. Equations 5-6 were used to compute the affine connection on a uniform grid of sample points that was centered on the origin and had spacing equal to two units. To do this, we considered a small square neighborhood of each sample point. Each trajectory segment that traversed the neighborhood was divided into line elements traversed in equal time intervals. Next, we considered any three pairs of such line elements, where each pair consisted of two adjacent line elements on a trajectory segment. Then, we asked if there was a unique affine connection $\hat{\Gamma}^k_{lm}$ that parallel transported each line element into the other line element of the same pair. The affine connection at the sample point was set equal to the average of the quantities $\hat{\Gamma}^k_{lm}$ that were derived from all possible triplets of paired line elements in the neighborhood (Eq.(6)). Triplets of paired line elements for which there was no unique solution (e.g., multiple solutions or no solution) did not contribute to this average. In this way, we derived an affine connection for which the neighboring trajectory segments were geodesic in an average sense. In this particular case, all components of the resulting affine connection were computed to be zero; i.e., the *x* coordinate system is a geodesic coordinate system of a flat manifold. This result is expected because a vanishing affine connection is the

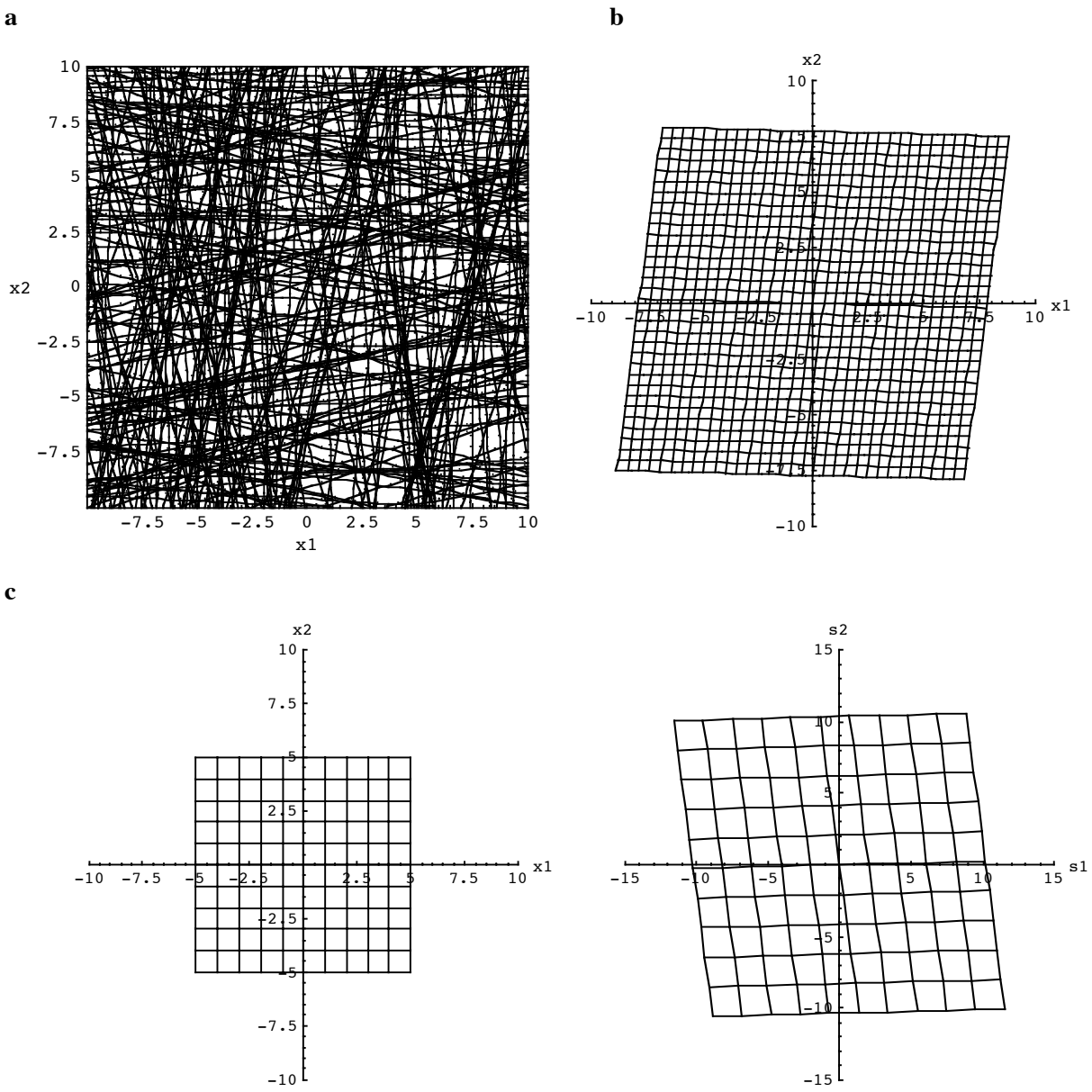

Figure 3. a) The simulated trajectory of recently encountered sensor states $x(t)$. The speed of traversal of each trajectory segment is indicated by the dots, which are separated by equal time intervals. The nearly horizontal and vertical segments are traversed in the left-to-right and bottom-to-top directions, respectively. b) The level sets of $s_a(x)$, which show the intrinsic coordinate system or scale derived by applying the method in Section 2.B to the data in *a*. The nearly vertical curves are loci of constant s_1 for evenly spaced values between –16 (left) and 16 (right); the nearly horizontal curves are loci of constant s_2 for evenly spaced values between –16 (bottom) and 16 (top). c) The grid-like array of sensor states on the left has the coordinate-independent representation on the right, obtained by using the scale in *b* to rescale those sensor states.

only one that parallel transports equally long line elements of straight lines into one another.

The reference state was chosen to be the origin of the *x* coordinate system ($x_0 = 0$), and the reference vectors were chosen to be $h_{01} = (0.488, -0.013)$ and $h_{02} = (0.064, 0.482)$. For the purposes of this paper, this information can be considered to be known from prior knowledge. For example, they could have been specified by having the user "show" the device two stimuli that produce a null sensor state with time derivatives equal to h_{01} and h_{02}. However, in actual fact, the above reference vectors were derived from the intrinsic directionality of the sensor state trajectory near x_0, by using the coordinate-independent technique in Levin (2001a). From this perspective, these vectors express the fact that the trajectory segments in Fig. 3a tend to be oriented in nearly horizontal and vertical directions. The affine connection was used to "spread" these vectors throughout the manifold by parallel transporting them along type 1 and type 2 geodesics. Then, Eqs.(1-2) were used to compute the value of s_a that comprises the coordinate-independent representation of each sensor state *x*. The results are shown in Fig. 3b, which depicts the level sets of $s_a(x)$. Because of the flat nature of this particular manifold, the coordinate system s_a is related to the *x* coordinate system by an affine transformation. Figures 3c shows how some sensor states in the *x* coordinate system are represented in the *s* coordinate system.

Next, we considered what would have happened if the same device had "experienced" sensor states shown in Fig. 4a. These trajectories are related to those in Figure 3a by the following non-linear transformation:

$$x_1 \rightarrow 2.5 + x_1 + 0.01x_1^2 - 0.02x_2^2 - 0.01x_1x_2 \qquad (14)$$

$$x_2 \rightarrow x_2 - 0.01x_1^2 + 0.02x_2^2 + 0.01x_1x_2$$

For example, suppose that *x* is the location of a feature in a digital image. Equation (14) could represent the way the sensor states are transformed by a distortion of the optical/electronic path within the camera (e.g., a camera "wearing" goggles) or by a distortion of the surface on which the camera is focused (e.g.,

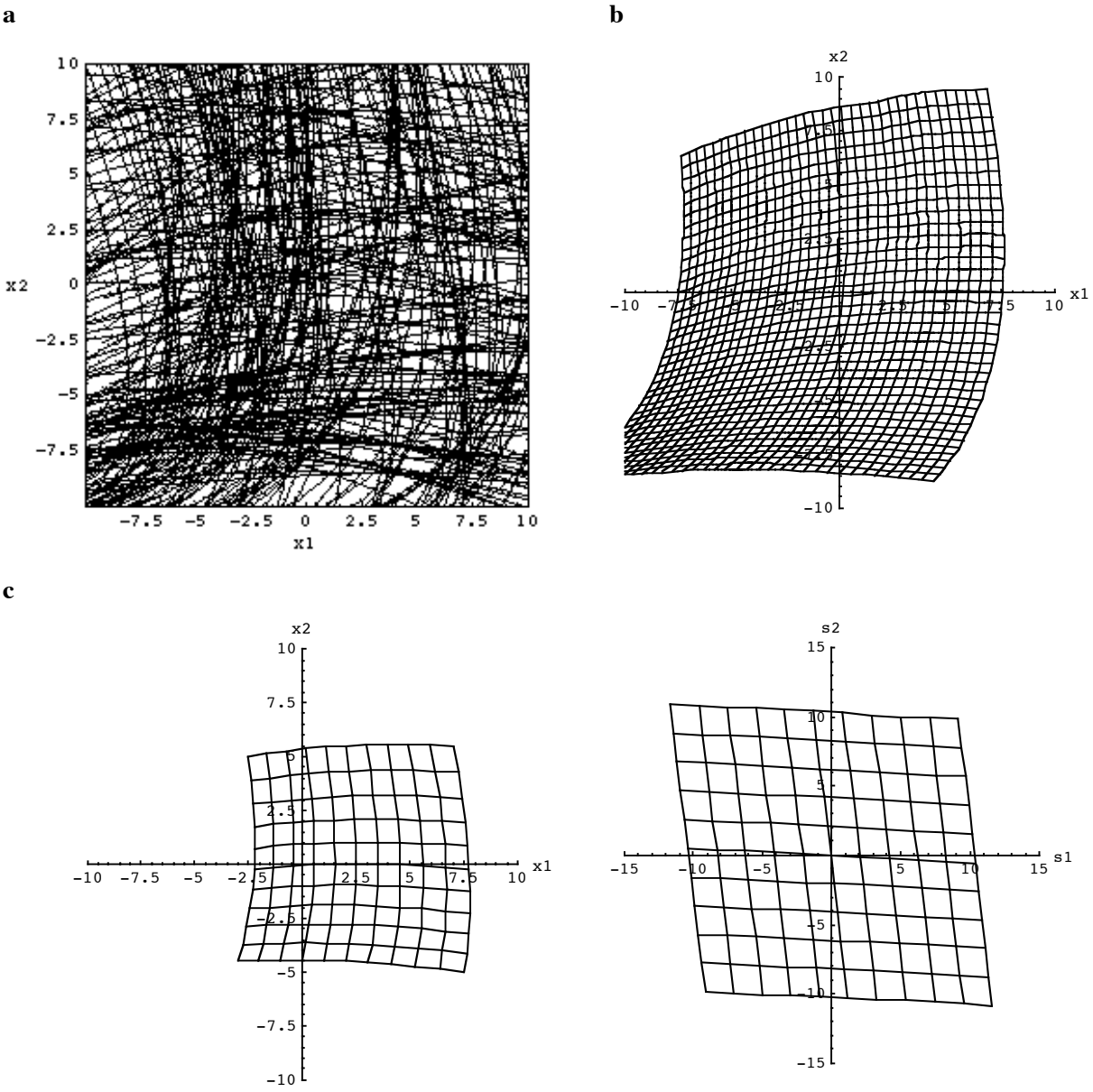

Figure 4. a) The simulated trajectory of recently encountered sensor states $x(t)$ that are related to those in Fig. 3a by the coordinate transformation in Eq.(14). The speed of traversal of each trajectory segment is indicated by the dots, which are separated by equal time intervals. The nearly horizontal and vertical segments are traversed in the left-to-right and bottom-to-top directions, respectively. b) The level sets of $s_a(x)$, which show the intrinsic coordinate system or scale that was derived by applying the method in Section 2.B to the data in *a*. The vertical curves are loci of constant s_1 for evenly spaced values between –24 (left) and 10 (right); the horizontal curves are loci of constant s_2 for evenly spaced values between –22 (bottom) and 16 (top). c) The array of sensor states on the left has the coordinate-independent representation on the right, obtained by rescaling the sensor states by means of the scale in *b*. The panel on the left was created by subjecting the left panel in Fig. 3c to the coordinate transformation in Eq.(14). Notice that the right panel is nearly identical to the right panel in Fig. 3c, thereby confirming the fact that these representations are invariant under the coordinate transformation.

distortion of a printed page). Notice that this transformation is invertible over the domain in Figure 3a to which it was applied. The procedure outlined above was used to compute the affine connection on a uniform grid of sample points. The resulting affine connection was non-vanishing at each sampled point, and smooth interpolation was used to estimate its values at intervening points. The reference sensor state and reference vectors were taken to be the transformed versions of the above-described reference data. As before, these quantities could have been determined by prior knowledge (i.e., by having the user "show" the device the stimuli mentioned above). The transformed reference vectors could also be derived from the intrinsic directionality of the transformed sensor state trajectory in Fig. 4a by applying the coordinate-independent method in Levin (2001a). The affine connection was used to "spread" these vectors throughout the manifold along type 1 and type 2 geodesics, and the *s* representation of each point in the manifold was computed by means of Eqs.(1-2). This is depicted by Fig. 4b, which shows the level sets of the resulting function $s_a(x)$. Notice that the warping of the sensor states has led to the warping of the s_a coordinate system, which is the natural scale imposed on the manifold by those sensor states. The function $s_a(x)$ was used to compute the *s* representation of the transformed version of the "image" in the left panel of Figure 3c. The transformed image and its representation are shown in Figure 4c. Comparison of Figs. 3 and 4 shows that the *s* representations of the untransformed and transformed images are nearly identical. In other words, these representations are invariant with respect to the process that transforms the sensor states by Eq.(14), even though they were computed without any knowledge of that transformation. The tiny discrepancies between Fig. 3c and Fig. 4c can be attributed to errors in the interpolation of the affine connection, which is due to the coarseness of the grid on which the affine connection was sampled. This error can be reduced if the distance between sample points can be decreased. This is possible if the device is allowed to experience a denser set of sensor states (i.e., more trajectory segments than shown in Fig. 3a) so that even tiny neighborhoods contain enough data to compute the affine connection.

4. DISCUSSION

In this paper, we demonstrated how time-dependent sensory data from an evolving stimulus could be rescaled in a non-linear, time-dependent fashion in order to create a time series of stimulus representations that are invariant under any unknown invertible transformation of the sensory data. Conventional methods of multidimensional scaling or dimensional reduction do not produce data representations that are transformation-independent in this way. In Section 1, we argued that this result has the following consequence: any two devices that sensitively and consistently detect the same d degrees of freedom of an evolving stimulus will create the same rescaled representation of the stimulus, even though the devices may be equipped with significantly different sensors. This conclusion followed from two related facts: there must be a time-independent invertible mapping between the d-dimensional manifold of stimulus configurations and a corresponding d-dimensional manifold of sensor states of each such device and, therefore, there is a time-independent invertible transformation between the sensor states of any two such devices, as they observe the same evolving stimulus. Notice that the rescaled representation of the sensor state time series in any such device must be identical to the rescaled representation of the time series of stimulus configurations themselves, because the two time series are related by an invertible transformation. In this sense, the rescaled sensor state time series can be considered to reflect an "inner" property of the time series of stimulus configurations. In other words, the rescaled sensor states are not affected by device-dependent "outer" features of the sensory process, such as the nature of the device's raw sensor states or the coordinate system that the device uses to label them.

As mentioned in Section 1, a rescaled sensor state time series is also independent of any extrinsic processes that remap the device's sensor states in an invertible fashion. For example, such representations are unaffected by a variety of observational conditions that are external to the device and the stimulus (e.g., altered intensity of a scene's illumination or altered positioning of the detectors with respect to the stimuli). Furthermore, any such device will produce identical rescaled representations of two *different* stimuli (e.g., S and S') whose time-dependent configurations are related by a time-independent invertible mapping. To see this, recall that there is a time-independent invertible mapping between the time series

of S configurations and the time series of sensor states $x(t)$ produced by S. Likewise, there is an invertible mapping between the time series of S' configurations and the time series of sensor states $x'(t)$ when the device observes S'. It follows that there is a time-independent invertible mapping between $x(t)$ and $x'(t)$, and, therefore, these time series have identical rescaled representations. As an example, suppose that one of the previously-described computer vision systems (e.g., system V in Section 1) was exposed to a time series of expressions of face F, and, on another occasion, it was exposed to a time series of expressions of a different face F'. Further, suppose that the two time series depicted similar sequences of facial expressions in the sense that there was a time-independent invertible mapping between the two parameters controlling F and the two analogous parameters controlling F'. It follows that the vision system would produce identical rescaled representations of the F and F' time series. For this reason, it would be quite natural for such a system to "recognize" the fact that the two faces were making analogous movements.

Strictly speaking, there is more than one way to interpret such an observation: i.e., the fact that two stimulus time series produce different sensor state time series, each of which leads to the same time series of rescaled representations. Without additional information, the device may not be able to determine whether the differences between the two sensor state time series were due to: 1) physical differences in the stimuli themselves; 2) the presence and absence of a process that affected the device's detector or the "channel" between it and the stimulus. For example, suppose the above-described vision system V sequentially observes two evolving facial stimuli that have identical time series of rescaled representations but different time series of raw sensor states, $x(t)$ and $x'(t)$. It may not be able to determine if $x(t)$ and $x'(t)$ were produced by: 1) two different faces that evolved through analogous facial expressions; 2) the same face that underwent the same sequence of expressions, first in the absence and then in the presence of some transformative process (e.g., the absence and presence of an image-warping lens). For instance, if $x(t)$ and $x'(t)$ differed by a scale factor, the device could attribute the sensor state differences to: 1) a change in the complexion of the observed face; 2) a change in the gain of

the device's camera or a change in the illumination of the face. Of course, humans can suffer from illusions due to similar confusions. Like a human, the device could distinguish between these possibilities only if it had additional information about the likelihood of various processes that might cause the transformation between the observed sensor states or if it was able to observe additional degrees of freedom of the stimulus.

In the above discussion it was assumed that the sensor states in a given time series were remapped by a *time-independent* invertible transformation. Now, consider the effects of the *sudden onset* of a process that invertibly transforms the sensor states. Suppose that each sensor state is rescaled by means of a scale derived from the sensor state time series encountered in the most recent ΔT time units. After the onset of the transformative process, there will be a transitional period of length ΔT, during which the device's stimulus representations will not be the same as those derived from the corresponding time series of untransformed sensor states. This is because these representations are referred to a mixture of transformed and untransformed sensor states. However, once the sensor state "database" is dominated by transformed data (i.e., once ΔT time units have elapsed), the representation of each stimulus will return to the form that is derived from the untransformed sensor state time series. This is because the description of each subsequently encountered sensor state will be referred to the properties of a collection of transformed sensor states. This phenomenon has been illustrated by numerical experiments reported elsewhere (Levin, 2001b). The time interval ΔT should be long enough so that the sensor states observed within it populate the sensor state manifold with sufficient density to derive sensor state representations (see the discussion of this issue in Sections 2.B and 2.C). Specifically, there must be enough sensor state trajectory segments near each point to endow the manifold with local structure (an affine connection or a metric). Thus, the device must have sufficient "experience" in order to form invariant stimulus representations, reminiscent of the role of experience in the acquisition of vision by human infants (Hebb, 1949; Sacks, 1995). Increasing ΔT will also tend to decrease the noise sensitivity of the method, because it increases the amount of signal averaging in the determination of the local

structure. Within these limitations, ΔT should be chosen to be as short as possible so that the device rapidly adapts to changing observational conditions.

Notice that, if the stimulus representation at each time point is derived from sensor states encountered in a "sliding time window" (e.g., the most recent time interval of length ΔT), a given sensor state may be represented in different ways at different times. This is because the representations of this sensor state may be computed from different collections of recently encountered sensor states on different occasions. In other words, the representation of an unchanged stimulus may be time-dependent because the representations are derived from the device's recent "experience", which may be time-dependent. Conversely, a given stimulus will be represented in the same way at two different times as long as the two descriptions are referred to collections of stimuli having the same *average* local properties (i.e., the same affine connection or metric). To visualize this, consider the following example. Consider the location of a particle in the center-of-mass coordinate systems of two different clusters of particles in a plane. The two descriptions of the particle's location will be the same, as long as the two collections have the same center-of-mass coordinate systems. In other words, the two representations of the particle's location are identical as long as these descriptions are referred to particle collections with the same average properties. Similarly, the stability of the *average* local properties of recently encountered sensor states will stabilize the representation of individual stimuli. If this type of temporal stability is important, stimulus representations should be derived from collections of sensor states that are sufficiently large to have stable statistical properties. This may put a lower bound on the length of the time period (e.g., ΔT) during which those sensor states are collected. Notice that rescaled stimulus representations have the same type of stability as the percepts of the human subjects of "goggle" experiments (Stratton, 1896, 1897a, and 1897b; Gibson, 1933; Held, 1972). Specifically, each subject's perception of stimuli returned to the pre-goggle baseline, after a period of adjustment during which he/she was exposed to familiar stimuli seen through the goggles. Likewise, the rescaled representation of each stimulus will return to the form that it had before the onset of a transformative process, after a period of adjustment during which the sensory device encounters stimuli with average properties similar to those encountered earlier.

In Sections 2.B and 2.C, an affine connection was directly derived from the sensor state trajectory $x(t)$ or from a metric that was directly related to $x(t)$. In either case, the affine connection was used to populate the entire manifold with vectors, which were parallel transported versions of reference vectors at the reference sensor state. These vectors were then utilized to create coordinate-independent representations of sensor states. These methods have some advantages with respect to the approach in Levin (2001a), in which local "preferred" vectors were directly derived from the intrinsic directionality of $x(t)$ at *every* point on the manifold. First of all, there may be points on the manifold at which such vectors cannot be derived because $x(t)$ may not endow the manifold with directionality there. However, an affine connection (and metric) can still be defined at such points and elsewhere, and it can then be used to transport vectors to those locations from other points where vectors can be defined in a coordinate-independent fashion. Thus, the methods based on affine-connected and Riemannian differential geometry are more generally applicable. The approach in Section 2.B of this paper is particularly advantageous because it tends to represent the most commonly encountered sensor state trajectories as geodesics (the generalization of the straight lines of Euclidean geometry). In other words, the sensor state tends to evolve along the direction created by parallel transporting the most recently observed line element of $x(t)$ along itself. Thus, such a device has a simple rule that provides some “intuition” about the likely evolution of a changing stimulus. In contrast, a device based on the approach in Levin (2001a) has no “intuition” about the future course of stimulus evolution. It “knows” a set of preferred directions at each point on the manifold, but it cannot use the past behavior of the stimulus to predict which direction it will follow in the future. Similarly, a device based on the methodology in Section 2.C of this paper “knows” that the average speed of sensor state evolution is unity because of Eq.(8). However, the device has no way to predict the direction of stimulus evolution because the stimulus trajectories may not resemble the geodesics of the metric-derived affine connection. In this sense, the type of device described in Section 2.B is more “intelligent” than those discussed in Section 2.C and in Levin (2001a).

The non-linear signal processing method presented in this paper could be used as a representation "engine" in the "front end" of intelligent sensory devices (Levin, 2000c). It would produce rescaled sensor state representations that are passed to the device's pattern analysis module. Because the effects of many extraneous observational conditions have been "filtered out" of these representations, it would not be necessary to recalibrate the device's detectors or to retrain its pattern analysis module in order to account for these factors. For example, computer vision devices of this kind could adapt to different intensities of the illumination of a scene, to various spatial relationships between the camera and the subject, and to drifting characteristics of the camera itself (Davies, 1990). Similarly, speech recognition devices with this capability should be able to adapt to microphones that have different response curves and are placed at various spatial locations (Rabiner, 1993; Ponting 1999). Furthermore, such a device may be able to "normalize" the voices of different speakers (Pisoni, 1996; Nygaard, 1998), if there is an invertible transformation relating the speakers' vocal tracts when they make the same utterances. This conjecture is supported by experiments in which the technique was applied to the short-term Fourier spectra of speech-like sounds, generated by a linear prediction mechanism (Levin, 2001b). Finally, it should be mentioned that this rescaling technique could be used to design a speech-like communications system that is resistant to corruption of information due to signal distortions in the transmitter, receiver, or the channel between them. In such a system, information is encoded in the rescaled representation of the signal time series (Levin, 2001b).

Humans tend to have similar perceptions despite significant differences in their sensory organs and processing pathways. Furthermore, each individual has the remarkable ability to perceive the intrinsic constancy of a stimulus even though its “appearance” is changing due to extraneous factors. These phenomena have been the subject of philosophical discussion since the time of Plato, and they have also intrigued modern neuroscientists (Zeki, 1999). This paper shows how to design a sensory device that represents stimuli invariantly in the presence of processes that systematically transform their sensor states. These stimulus representations are invariant because they encode "inner" properties of the time series of the stimulus configurations themselves; i.e., properties that are independent of the nature of the observing

device or the conditions of observation. Perhaps, the approximate universality and constancy of human perception are due to a similar appreciation of the "inner" structure of time series of stimuli. A significant evolutionary advantage would accrue to organisms that developed this ability.

REFERENCES

Carroll, J. D. and Arabie, P. (1980). Multidimensional scaling. *Annual Reviews of Psychology*, **31**, 607-649.

Cox, T. and Cox, M. (1994). *Multidimensional Scaling*. London: Chapman & Hall.

Davies, E. R. (1990). *Machine Vision: Theory, Algorithms, and Practicalities*. New York: Academic Press.

Gibson, J. J. (1933). Adaptation, after-effect, and contrast in the perception of curved lines. *Journal of Experimental Psychology*, **16**, 1-31.

Hebb, D. O. (1949). *The Organization of Behavior*. New York: Wiley.

Held, R. and Whitman, R. (1972). *Perception: Mechanisms and Models*. San Francisco: W. H. Freeman.

Holman, E. W. (1978). Completely nonmetric multidimensional scaling. *Journal of Mathematical Psychology*, **18**, 39-51.

Levin, D. N. (2000a). A differential geometric description of the relationships among perceptions. *Journal of Mathematical Psychology*, **44**, 241-284.

Levin, D. N. (2000b). Time-dependent signal representations that are independent of sensor calibration. *Journal of the Acoustical Society of America*, **108**, 2575. Posted at http://asa.aip.org/newport/information.html.

Levin, D. N. (2000c). Self-referential method and apparatus for creating stimulus representations that are invariant under systematic transformations of sensor states. Patent pending.

Levin, D. N. (2001a). Stimulus representations that are invariant under invertible transformations of sensor data. *Proceedings of the Society of Photoelectronic Instrumentation Engineers*, **4322**, 1677-1688.

Levin, D. N. (2001b). Universal communication among systems with heterogeneous "voices" and "ears". *Proceedings of the International Conference on Advances in Infrastructure for Electronic Business, Science, and Education on the Internet*, Scuola Superiore G. Reiss Romoli S.p.A., L'Aquila, Italy, August 6-12. Posted at http://www.ssgrr.it/en/ssgrr2001/index.htm.

Nygaard, L. C., Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Perception and Psychophysics*, **60**, 355-376.

Pisoni, D. B. (1997). Some thoughts on "normalization" in speech perception. In: Johnson, K. and Mullennix, J. W. (Eds.), *Talker Variability in Speech Processing*, San Diego: Academic Press, pp. 9-32.

Ponting, K. M. (1999). Channel adaptation. In: Ponting, K. (Ed.), *Computational Models of Speech Pattern Processing*. Berlin: Springer.

Rabiner, L. and Juang, B.-H. (1993). *Fundamentals of Speech Recognition*. Englewood Cliffs, N. J.: Prentice Hall.

Roweis, S. T., Saul, L. K. (2000). Nonlinear dimensionality reduction by locally linear embedding. *Science*, **290**, 2323-2326.

Sacks, O. (1995). *An Anthropologist on Mars: Seven Paradoxical Tales*. New York: Knopf, pp. 108-152.

Schrodinger, E. (1963). *Space-Time Structure*. Cambridge, UK: Cambridge University Press.

Shepard, R. N. (1962). The analysis of proximities: multidimensional scaling with an unknown distance function. I. *Psychometrika*, **27**, 125-140 and II. *Psychometrika*, **27**, 219-246.

Stratton, G. M. (1896). Some preliminary experiments on vision without inversion of the retinal image. *The Psychological Review*, **3**, 611-617.

Stratton, G. M. (1897a). Vision without inversion of the retinal image. *The Psychological Review*, **4**, 341-360.

Stratton, G. M. (1897b). Vision without inversion of the retinal image (concluded). *The Psychological Review*, **4**, 463-481.

Tenenbaum, J. B., de Silva, V., and Langford, J. C. (2000). A global geometric framework for nonlinear dimensionality reduction. *Science*, **290**, 2319-2323.

Weinberg, S. (1972). *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity*. New York: Wiley.

Zeki, S. (1999). *Inner Vision: An Exploration of Art and the Brain*. Oxford, UK: Oxford University Press.

FIGURES

Figure 1: Consider a path $x(u)$ $(0 \le u \le 1)$ between a reference sensor state x_0 and a sensor state of interest. If vectors h_a can be defined at each point along the path, each line segment δx can be decomposed into its components δs_a along the vectors at that point.

Figure 2. a) An untransformed signal $x(t)$ describing a long succession of identical pulses that are uniformly spaced in time. b) The signal representation $S(t)$ that results from applying the rescaling method in Section 2.B either to the signal in *a* or to the transformed version of that signal in *c*. c) The signal obtained by subjecting the signal in *a* to the transformation: $x'(x) = g_1 \ln(1 + g_2 x)$ where $g_1 = -0.3$ and $g_2 = 170$.

Figure 3. a) The simulated trajectory of recently encountered sensor states $x(t)$. The speed of traversal of each trajectory segment is indicated by the dots, which are separated by equal time intervals. The nearly horizontal and vertical segments are traversed in the left-to-right and bottom-to-top directions, respectively. b) The level sets of $s_a(x)$, which show the intrinsic coordinate system or scale derived by applying the method in Section 2.B to the data in *a*. The nearly vertical curves are loci of constant s_1 for evenly spaced values between –16 (left) and 16 (right); the nearly horizontal curves are loci of constant s_2 for evenly spaced values between –16 (bottom) and 16 (top). c) The grid-like array of sensor states on the left has the coordinate-independent representation on the right, obtained by using the scale in *b* to rescale those sensor states.

Figure 4. a) The simulated trajectory of recently encountered sensor states $x(t)$ that are related to those in Fig. 3a by the coordinate transformation in Eq.(14). The speed of traversal of each trajectory segment is indicated by the dots, which are separated by equal time intervals. The nearly horizontal and vertical segments are traversed in the left-to-right and bottom-to-top directions, respectively. b) The level sets of $s_a(x)$, which show the intrinsic coordinate system or scale that was derived by applying the method in Section 2.B to the data in *a*. The vertical curves are loci of constant s_1 for evenly spaced values between

–24 (left) and 10 (right); the horizontal curves are loci of constant s_2 for evenly spaced values between –22 (bottom) and 16 (top). c) The array of sensor states on the left has the coordinate-independent representation on the right, obtained by rescaling the sensor states by means of the scale in *b*. The panel on the left was created by subjecting the left panel in Fig. 3c to the coordinate transformation in Eq.(14). Notice that the right panel is nearly identical to the right panel in Fig. 3c, thereby confirming the fact that these representations are invariant under the coordinate transformation.

David N. Levin received his Ph.D. in theoretical physics from Harvard University in 1970 and did research in quantum field theory until 1977, when he entered medical school at the University of Chicago. He joined the faculty after receiving an M.D. and completing radiology residency at the University. During 1987-1999, he was Director of Clinical MRI at the University. He is currently Professor in the Department of Radiology and co-directs the University's Brain Research Imaging Center. This facility is equipped with a 3 T MRI scanner that is dedicated to brain research with functional MRI and MR spectroscopy. His past research interests have included multimodality 3D brain imaging, computer-assisted neurosurgery, and image segmentation. His current research is focused on new methodology for mapping the brain with functional MRI, novel techniques for using prior knowledge to increase the speed of MR image acquisition, and applications of non-linear signal processing to the design of intelligent sensory systems.