1/16/02 v. 3.3

SENSOR-INDEPENDENT STIMULUS REPRESENTATIONS

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. These representations are invariant because they encode "inner" properties of the time series of stimulus configurations themselves. This means that 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. Such sensor-independent stimulus representations will also be 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. In an intelligent sensory device, this kind of representation "engine" could function as a "front end" that passes rescaled sensor state 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.

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/faculty/Levin.html

1. INTRODUCTION

Humans form percepts that are remarkably independent of the condition of their sensors. 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 example, 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*. This has the following consequence: any two devices, possibly equipped with dramatically 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. 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 the corresponding collection 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 the amplitudes of two specific coefficients in the 2D Bessel expansion of the facial image. If the internal state of this system also 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 face, 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) 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 extraneous processes of the above types (Davies, 1990).

As mentioned above, the methodology of this paper can create invariant representations of the sensor data produced by any device whose sensory states are invertibly related to the underlying stimulus configurations. The embedding theorems of non-linear dynamics (Whitney, 1936; Takens, 1981; Sauer, 1991) suggest that this invertibility requirement will be satisfied by *almost any sensory device with a*

sufficient number of sensors. Specifically, those theorems show that, if the stimulus evolves in a d-dimensional space, the sensor states of almost any device having more than $2d$ sensors will be invertibly related to the underlying stimulus configurations. Thus, the techniques in this paper may be applied to the signals produced by any such device.

The method in this paper differs significantly from techniques for multidimensional scaling or dimensional reduction (Shepard, 1962; Carroll, 1980; Cox, 1994). In each of these methods, 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 rankings are not invariant under general, non-linear coordinate transformations. Therefore, the scale values assigned to each data point are also not transformation-independent, unlike the rescaled representations described in this paper. However, it should also be mentioned that multidimensional scaling methods are applicable to data that do not form a time series, unlike the technique 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 show how tensor calculus can be used to find a transformation-independent representation of each sensor state in a sensor state time series (Levin, 2000b, 2001a, 2001c). The technique is illustrated with an analytical example in Section 2 and with 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. For example, in an imaging system, the processing units may extract the time-dependent locations or intensites 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). This processing may also include linear or non-linear dimensional reduction procedures that map the detector outputs onto sensor states with the same dimensionality as the underlying degrees of freedom of the stimulus (Roweis, 2000; Tenenbaum, 2000; and references therein). 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.

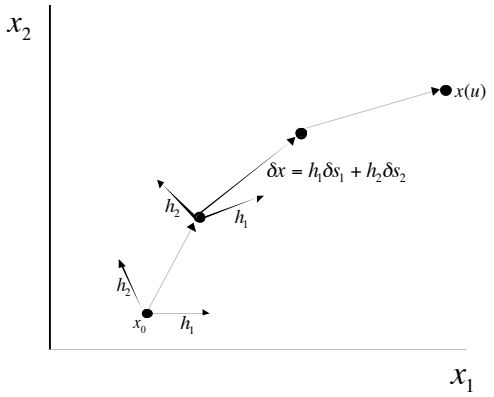

Figure 1. Consider a path $x(u)$ ($0 \leq u \leq 1$) between a reference sensor state x_0 and a sensor state of interest. If vectors h_a can be defined at each point along the path, each line segment δx can be decomposed into its components δs_a along the vectors at that point.

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 (e.g., $x \rightarrow x'$). 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 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 sensor state scale, 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 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, in the same way that a global "center-of-mass" coordinate system is determined by the coordinates of the particles in a collection. Because s denotes the "location" of a sensor state with respect to these local coordinate systems, it will be invariant under any invertible transformation (linear or non-linear) of the entire sensor state time series, in analogy to the manner in which the location of a particle in the center-of-mass coordinate system is invariant under global rotations/translations of the whole particle collection.

In the following, we show how the information required for this type of description (a reference state, paths connecting it to other sensor states, and the vectors h_a) can be derived from the local structure of a collection of sensor states in a time series. For the sake of simplicity, this is first illustrated for one-dimensional manifolds of sensor states ($N=1$). Then, we show how a similar procedure can be used to handle manifolds of any dimension.

2.B. One-Dimensional Sensor State Manifolds

Suppose that the processed output of the detectors consists of a single signal x. For example, x could represent the intensity of a pixel at a certain location in a digital image of a scene, or it could represent the amplitude of a microphone's output. Suppose that the device has been exposed to a time-dependent series of stimuli, which produce sensor states $x(t)$, where t denotes time, and let X be the sensor signal at time T. In this paragraph, we show how to rescale the signal level at this particular time point. The exact same procedure can be used to rescale the signal level at other times, thereby deriving a representation of the entire signal time series. Suppose that $x(t)$ passes through all of the signal levels in $[0, X]$ at one or more times during a chosen time interval of length ΔT (e.g. $T - \Delta T \le t < T$). Here, ΔT is a parameter that can be chosen freely, although it influences the adaptivity and noise sensitivity of the method (see Discussion). At each $y \in [0, X]$, define the value of the function $h(y)$ to be

$$h(y) = \left\langle \frac{dx}{dt} \right\rangle_y \qquad \text{(Eq. 3)}$$

where the right side denotes the derivative averaged over those times in $T - \Delta T \le t < T$ when $x(t)$ passes through the value *y*. If $h(y)$ is non-vanishing for all $y \in [0, X]$, it can be used to compute the scale function $s(x)$ on this interval

$$s(x) = \int_0^x \frac{dy}{h(y)} \qquad \text{(Eq. 4)}$$

The quantity $S = s(X)$ can be considered to represent the level of the signal *X* at time *T,* after it has been non-linearly rescaled by means of the function $s(x)$. Now, now consider the signal produced by the time-independent transformation $x \to x' = x'(x)$. Furthermore, suppose that $x \to x'$ is invertible (i.e., $x'(x)$ is monotonic), and suppose that it preserves the null signal (i.e., $x'(0) = 0$). The transformed signal $x'(t) = x'[x(t)]$ has the value $X' = x'(X)$ at *t=T*. During $T - \Delta T \le t < T$, $x'(t)$ passes through each of the values in $[0, X']$, because of our assumption that $x(t)$ attains all of the values in $[0, X]$ during that time interval. Therefore, for each $y' \in [0, X']$, the process in Eq.(3) can be applied to the transformed signal in order to define the function $h'(y')$ at time *T*

$$h'(y') = \left\langle \frac{dx'}{dt} \right\rangle_{y'} \qquad \text{(Eq. 5)}$$

where the right side denotes the derivative averaged over those times in $T - \Delta T \le t < T$ when $x'(t)$ passes through the value y'. By substituting $x'(t) = x'[x(t)]$ in Eq.(5), using the chain rule of differentiation, and noting that $x(t)$ passes through the value *y* when $x'(t)$ passes through the value $y' = x'(y)$, we find $h'(y') = \left.\frac{dx'}{dx}\right|_y h(y)$. The function $h'(y')$ is non-vanishing for $y' \in [0, X']$ because the monotonicity of $x'(x)$ implies $dx'/dx \neq 0$. This means that the process in Eq.(4) can be used to compute a scale function $s'(x')$ on this interval

$$s'(x') = \int_0^{x'} \frac{dy'}{h'(y')} \qquad \text{(Eq. 6)}$$

The quantity $S' = s'(X')$ represents the level of the transformed signal X' at time *T,* after it has been rescaled by means of a function $s'(x')$, which was derived from $x'(t)$ just as $s(x)$ was derived from $x(t)$. Because of our assumption that $x = 0$ transforms into $x' = 0$, a change of variables ($y \to y'$) in Eq.(4) implies $s'(x') = s(x)$ and, therefore, $S' = S$. This means that the rescaled value of a signal is invariant under the signal transformation $x \to x'$. In other words, the rescaled value *S* of the untransformed signal level at time *T*, computed from recently encountered untransformed signal levels,

will be the same as the rescaled value S' of the transformed signal level at time T, computed from recently encountered transformed signal levels. Now, the above procedure can be followed in order to rescale the signal levels at times other than T. The resulting time series of rescaled signal levels $S(t)$, which the sensory device derives from the untransformed signal $x(t)$ in this way, will be identical to the time series of rescaled signal levels $S'(t)$, which the sensory device derives from the transformed signal $x'(t)$. Note that the scale function defined by Eq.(4) is the same as that defined by Eqs.(1, 2) in the special case of a one-dimensional sensor state manifold. From this more general perspective, $h(y)$ is the contravariant vector identified at each point on the one-dimensional sensor state manifold, and the null signal is the reference sensor state in each relevant coordinate system

Notice that the forms of the scale functions $s(x)$ and $s'(x')$ (and of $h(y)$ and $h'(y')$) will usually be time-dependent because they are computed from the time course of previously encountered signals. At some times, the sensory device may be unable to compute a rescaled signal level. This will happen if the scale function in Eq.(4) does not exist because the quantity $h(y)$ vanishes for some $y \in [0, X]$ or if the function $h(y)$ cannot even be computed at some values of y because these signal levels were not encountered recently. Because of the monotonicity of $x'(x)$, a signal invariant at such times cannot be computed from either the untransformed or transformed signals. The inability to compute signal invariants at some time points means that the number of independent signal invariants (i.e., the number of time points at which $S(t)$ can be computed) may be less than the number of degrees of freedom in the raw signal from which the invariants were computed (i.e., the number of time points at which the signal $x(t)$ is measured).

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: $a = 0.1$ and $b = -0.5$ (measured in inverse time units). If we choose ΔT to be an integral number of periods of $x(t)$, it is easy to see from Eqs.(3, 4) that the untransformed signal implies $h(y) = (a + b)/2$ and $S(t) = s[x(t)] = 2x(t)/(a + b)$ at each point in time. Figure 2b shows $S(t)$, which is the untransformed signal after it has been rescaled at each time point as dictated by its earlier time course. Now, consider the transformed signal that is related to the untransformed signal by any of the following non-linear functions: $x'(x) = g_1 \ln(1 + g_2 x)$ where $g_2 > 0$.

a **b** **c**

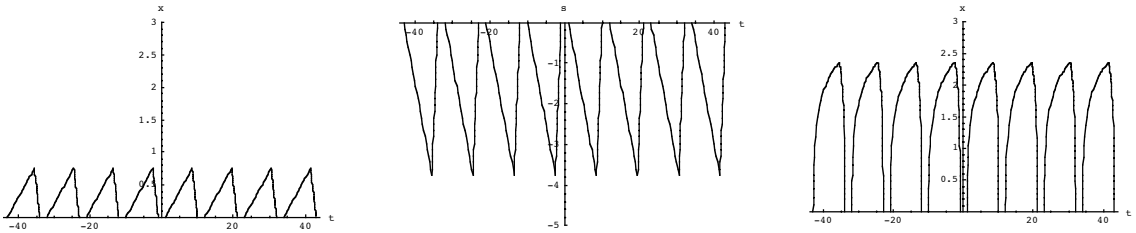

Figure 2. a) An untransformed signal $x(t)$ describing a long succession of identical pulses that are uniformly spaced in time. b) The signal representation $S(t)$ that results from applying the rescaling method in Section 2.B either to the signal in *a* or to the transformed version of that signal in *c*. c) The signal obtained by subjecting the signal in *a* to the transformation: $x'(x) = g_1 \ln(1 + g_2 x)$ where $g_1 = 0.5$ and $g_2 = 150$.

For example, if $g_1 = 0.5$ and $g_2 = 150$, 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.(5) is used to compute $h'(y')$ from the transformed signal , the result is:

$$h'(y') = \frac{1}{2}(a+b) g_1 g_2 \, e^{-y'/g_1} \qquad \text{(Eq. 7)}$$

at each point in time. Then, Eq.(6) shows that the rescaled version of the transformed signal is

$$S'(t) = s'[x'(t)] = \frac{2(e^{x'(t)/g_1} - 1)}{g_2(a+b)}, \qquad \text{(Eq. 8)}$$

Substituting $x'(t) = x'[x(t)]$ into Eq.(8) shows that $S'(t) = S(t)$. 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 $h(y)$ and $h'(y')$) 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. This is because, in order to simplify the calculation, $x(t)$ was chosen to be periodic and ΔT was chosen to be an integral number of these periods. 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 at the beginning of this Section.

In the above discussion, the null signal was taken to be the reference sensor state x_0, and the signal transformation was assumed to preserve the null signal. In general, any sensor state can be taken to be the reference sensor state, as long as the reference sensor state x_0' used to rescale a transformed signal time series is the transformed version of the reference state x_0 used to rescale the untransformed signal

time series: i.e., as long as $x_0' = x'(x_0)$. In mathematical terms, this means that the reference state must be chosen in a coordinate-independent manner. For example, the reference sensor state could be chosen to be the sensor state that is the local maximum of a function defined to be the number of times each sensor state is encountered in a chosen time interval. This may be particularly useful for multidimensional sensor state manifolds (see next Section). Alternatively, prior knowledge may be used to choose the reference state. For instance, as described above, we may know that the null sensor state always corresponds to the same stimulus, and, therefore, it can be chosen to be the reference state. For example, this might be the case if the transformations of interest reflect differences in the gain curves of the detectors of different devices. Finally, the reference sensor state may be chosen to be the sensor state produced by a user-determined stimulus that is "shown" to the sensory device. Recall that the reference sensor state serves as the origin of the scale function used to rescale other 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. 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.

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 (e.g., a mixture of the shapes in Figures 2a and 2c). During this transition, the device’s scale function will evolve from the form derived from untransformed signals to the form derived from transformed signals (e.g., from *s(x)* to *s’(x)*), 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 demonstrated in numerical experiments reported elsewhere (Levin, 2001b and 2001c). Thus, like a human, the system adapts to the presence of the transformation after a period of adjustment.

2.C. Multidimensional Sensor State Manifolds

In this section, we generalize the above method to sensory devices that have multidimensional sensor states. Let the device's sensor state be represented by an array of numbers x (x_k, $k = 1,...,N$ where $N \geq 1$), and let $x(t)$ be the time series of sensor states encountered in a chosen time interval (e.g., the most recent time interval of length ΔT). This function describes a trajectory that crosses the sensor state manifold. We now show how these data can be used to define local vectors $h_a(x)$ in a manner that is independent of the coordinate system. Consider a point *x* that has multiple trajectory segments passing through it in at least *N* different directions, where *N* is the manifold's dimension. The time derivatives of the segments passing through *x* form a collection of contravariant vectors $\hat{h}_i$ at *x*:

$$\hat{h}_i = \left.\frac{dx}{dt}\right|_{t_i} \qquad \text{(Eq. 9)}$$

where t_i denotes the i^{th} time at which the trajectory passes through *x*. To verify that this transforms as a vector, note that in any other coordinate system ($x' = x'(x)$) the corresponding quantity is $\hat{h}_i' = \frac{dx'}{dt} = \frac{\partial x'}{\partial x}\frac{dx}{dt} = \frac{\partial x'}{\partial x}\hat{h}_i$. These quantities can be used to define *N* vectors at *x* if they tend to fall into clusters oriented along different directions in the manifold. To see this, pick an integer $C \geq N$ and partition the indices *i* into *C* non-empty sets labeled S_c where *c=1, …, C* . Next, compute the *N x N* covariance matrix M_c of the vectors corresponding to each set of indices:

$$M_c = \frac{1}{N_c}\sum_{i \in S_c} \hat{h}_i \hat{h}_i \qquad \text{(Eq. 10)}$$

where N_c is the number of indices in S_c. Each of these matrices transforms as a tensor with two contravariant indices, and the determinant of each matrix $|M_c|$ transforms as a scalar density of weight equal to minus two (Schrodinger, 1963; Weinberg, 1972); namely, if coordinates on the manifold are transformed as $x \rightarrow x'$, then

$$|M_c| \rightarrow |M_c'| = \left|\frac{\partial x'}{\partial x}\right|^2 |M_c| \qquad \text{(Eq. 11)}$$

This follows from the facts: 1) $M_c \rightarrow M_c' = \frac{dx'}{dx} M_c \frac{dx'}{dx}^T$ where *T* denotes transpose, 2) a matrix and its transpose have the same determinant. Next, compute *E*, which is defined to be the sum of powers of these determinants:

$$E = \sum_c |M_c|^p \qquad \text{(Eq. 12)}$$

where p is some real positive number. Equation 11 implies that E transforms as a scalar density of weight $-2p$. Now tabulate the values of E for all possible ways of partitioning the set of vectors $\hat{h}_i$ into C non-empty sets, and find the partition that results in the smallest value of E. This partition will tend to group the vectors into subsets with minimal matrix determinants. Therefore, the vectors in each group will tend to be linearly dependent or nearly linearly dependent, and they will tend to form a cluster that is oriented in one direction. Next, compute the vectors h_c at x by finding the average vector in each part of the optimal partition:

$$h_c = \frac{1}{N_c} \sum_{i \in S_c} \hat{h}_i \qquad \text{(Eq. 13)}$$

Because the $\hat{h}_i$ are contravariant vectors, the h_c will also transform as contravariant vectors as long as they are partitioned in the same manner in any coordinate system. However, because E transforms by a positive multiplicative factor, the same partition minimizes it in any coordinate system. Therefore, the optimal partition is independent of the coordinate system, and the h_c are indeed contravariant vectors. Finally, the indices of the h_c can be relabeled so that the corresponding determinants $|M_c|$ are in order of ascending magnitude. This ordering is also coordinate-independent because these determinants transform by a positive multiplicative factor (Eq.(11)). As a result, if the foregoing computations are done in any coordinate system, the same vectors h_c will be created, and these vectors provide a coordinate-independent characterization of the directionality of the trajectories passing through x.

The first N vectors that are linearly independent can be defined to be the h_a in Eq.(1). These can be used to compute δs, the coordinate-independent representation of any line element passing through x. Once we have specified a path connecting a reference state x_0 to any sensor state x, Eq.(2) can be integrated to create a coordinate-independent representation s of that state. The path must be completely specified because the integral in Eq.(2) may be path-dependent. To see this, note that Eq.(1) can be inverted to form:

$$\delta s_a = \tilde{h}_a \cdot \delta x \qquad \text{(Eq. 14)}$$

where the covariant vectors $\tilde{h}_a$ are found by solving $\sum_{a=1,\ldots,N} \tilde{h}_{ak} h_a^l = \delta_k^l$, δ_k^l being the Kronecker delta function. It follows from Eq.(2) that each component of s is a line integral of $\tilde{h}_a$ for $a=1, \ldots, N$. Stoke's theorem shows that these line integrals will be path-dependent unless the "curl" of $\tilde{h}_a$ vanishes:

$$\frac{\partial \tilde{h}_{ak}}{\partial x_l} - \frac{\partial \tilde{h}_{al}}{\partial x_k} = 0 \qquad \text{(Eq. 15)}$$

Because this may not be true for some sensor state manifolds, we must create a coordinate-independent way of specifying a path from x_0 to any point *x* on the manifold. Such a path can be determined in the following manner: first, generate a "type 1" trajectory through x_0 by moving along the local h_1 direction at x_0 and then moving along the h_1 direction at each subsequently encountered point. Next, generate a "type 2" trajectory through each point on the type 1 trajectory by moving along the local h_2 direction at that point and at each subsequently-encountered point. Continue in this fashion until a type *N* trajectory has been generated through each point on every trajectory of type *N-1*. Because of the linear independence of the h_a at each point, the collection of points on type *n* trajectories ($1 \le n \le N$) comprises an *n*-dimensional subspace of the manifold. Therefore, each point on the manifold lies on a type *N* trajectory and can be reached from x_0 by traversing the following type of path: a segment of the type 1 trajectory, followed by a segment of a type 2 trajectory, …, followed by a segment of a type *N* trajectory. This path specification is coordinate-independent because the quantities h_a transform as contravariant vectors. Therefore, if Eq.(2) is integrated along this "canonical" path, the resulting value of *s* provides a coordinate-independent description of the sensor state *x* in terms of the recently-encountered sensor states; i.e., a description that is invariant in the presence of processes that remap sensor states.

Strictly speaking, the vectors h_a must be computed in the above-described manner at every point on each path in Eq.(2). This means that the trajectory of previously encountered sensor states $x(t)$ must cover the manifold very densely so that it passes through every point at least *N* times. However, this requirement can be relaxed for most applications. Specifically, suppose that the h_a are only computed at a finite collection of sample points on the manifold, and suppose that these vectors are computed from derivatives of trajectories passing through a very small neighborhood of each sample point (not necessarily passing through the sample point itself). Furthermore, suppose that values of h_a between the sample 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 directionality of the manifold varies. This must be true in all relevant coordinate systems; i.e., in coordinate systems corresponding to the transformative effects of all interesting processes that remap the device's sensor states. Some circumstances may prevent the derivation of vectors h_a at a sufficiently dense set of sample points on the manifold. For example, suppose that there is no unique way of partitioning the $\hat{h}_i$ at a point in order to minimize *E*, or suppose that the h_c (Eq.(13)) associated with a minimal value of *E* do not contain *N* linearly independent members. These results would indicate that the temporal course of sensor states *x(t)* does not endow the manifold with sufficient directionality.

However, in this situation, it may still be possible to create coordinate-independent representations of stimuli by means of other methods (based on affine-connected and/or Riemannian differential geometry), which only require that the manifold have intrinsic directionality at a *single* point. The vectors at that point can then be moved (parallel-transported) to other points on the manifold. These differential geometric techniques are described and illustrated with numerical examples in Levin, 2001a.

3. EXPERIMENTS WITH SIMULATED DATA

3.A. 1D Sensor State Manifolds: Human Speech Waveforms

In this Section, the mathematical properties of the rescaling process are illustrated by applying it to acoustic waveforms of human speech. An adult male American uttered English words with speed and loudness that were characteristic of normal conversation. These sounds were digitized with 16 bits of depth at a sample rate of 11.025 kHz. Figure 3a shows a 40 ms segment of the 406 ms signal $x(t)$ corresponding to the word "open". Figure 3b shows the "*s* representation": i.e., the dynamically rescaled signal $S(t)$ that was derived from Fig. 3a by the method of Section 2. The value of *S* was determined at each time point by a scale function $s(x)$, which was derived from the previous 10 ms of signal (i.e., ΔT = 10 ms). These scale functions are shown by the horizontal lines in Fig. 3a, which denote values of *x* corresponding to $s = \pm 50n$ for *n=1, 2, … .* Figure 3d shows the signal that was derived from Fig. 3a by means of the non-linear transformation ($x'(x)$) shown in Fig. 3c. Figure 3e is the dynamically rescaled signal that was derived from Fig. 3d with the parameter ΔT chosen to be 10 ms. Although there are significant differences between the "raw" (unrescaled) signals in Figs 3a and 3d, their *s* representations (Figs. 3b and 3e) are almost identical, except for a few small discrepancies that can be attributed to the discrete methods used to compute derivatives. Thus, the *s* representation was invariant under a non-linear signal distortion, as expected from the derivation in Section 2. It is interesting to note that this result is apparent when one listens to the sounds represented in Fig. 3. Although all four signals in Figs. 3a, b, d, e sound like the word "open", there is a clear difference between the sounds of the two raw signals, and there is no perceptible difference between the sounds of their rescaled representations. In general, the rescaled signals sound like the word "open", uttered by a voice degraded by slight "static".

Some comments should be made about technical aspects of the example in Fig. 3. The dynamically rescaled signals in Figs 3b and 3e were computed by a minor variant of the method in Section 2.B. Specifically, we assumed that all signal distortions were *monotonically positive*, and we restricted the contributions to Eq.(3) and Eq.(5) to those time points at which the signal had a *positive* time derivative as it passed through the values y and y', respectively. The rescaled signal is still invariant because monotonically positive transformations do not change the sign of the signal's time

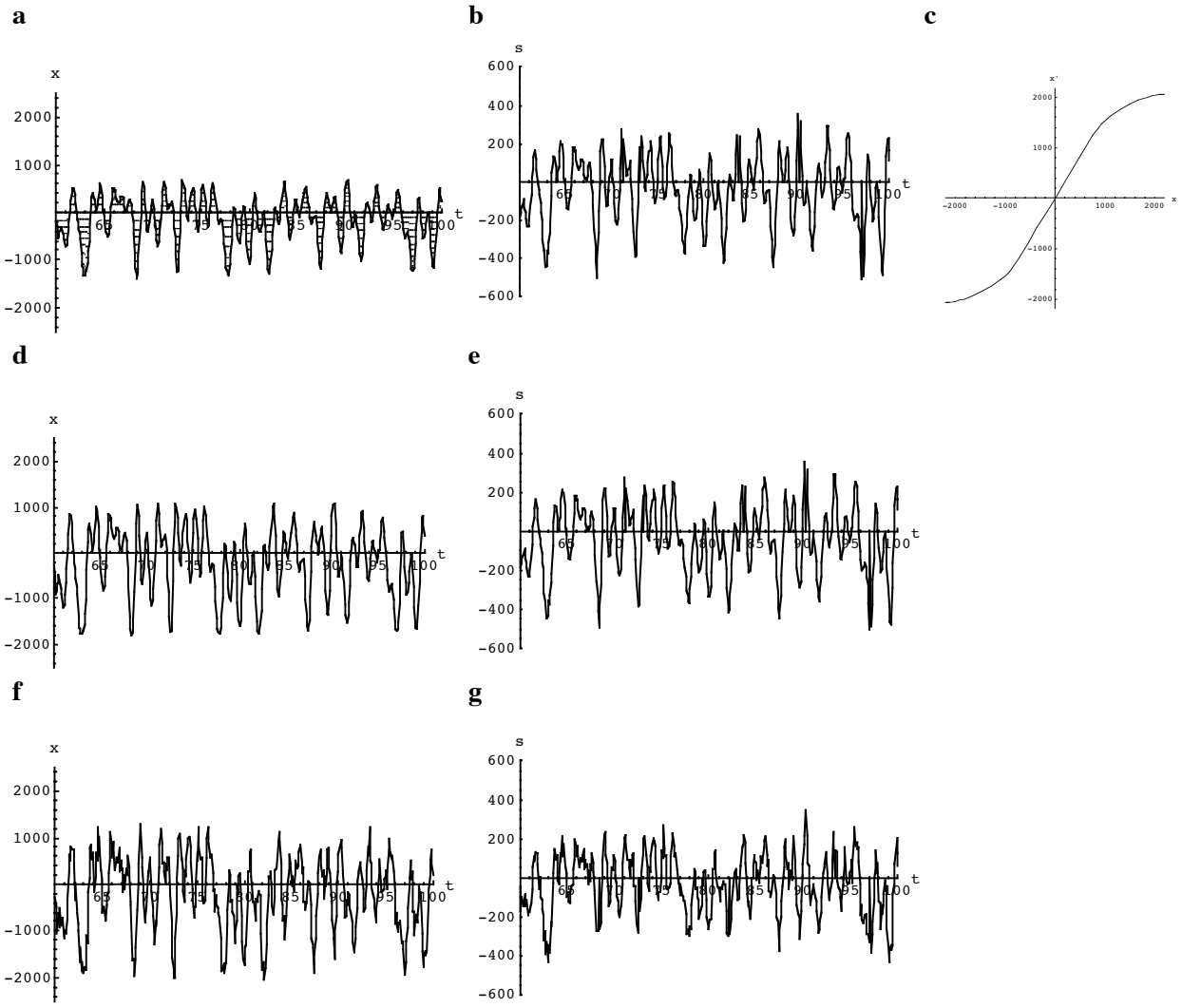

Figure 3. a) A signal obtained by digitizing the acoustic signal of the word "open", uttered by a male speaker of American English. A 40 ms segment of the 406 ms signal is shown, with time given in ms. The horizontal lines show signal amplitudes that have dynamically rescaled values equal to $s = \pm 50n$ for $n=1, 2, \ldots$. b) The signal $S(t)$ (in units of μs) obtained by dynamically rescaling the signal in panel *a*, with the parameter ΔT = 10 ms. c) The non-linear function $x'(x)$ that was used to transform the signal in panel *a* into the one in panel *d*. d) A distorted version of the signal in panel *a*, obtained by applying the non-linear transformation in panel *c*. e) The signal obtained by dynamically rescaling the signal in panel *d* with the parameter ΔT = 10 ms. f) A signal derived from the signal in *d* by adding white noise so that the signal-to-noise ratio is 3.5. g) The signal obtained by dynamically rescaling the signal in panel *f* with ΔT = 10 ms.

derivative, and, therefore, the functions $h(y)$ and $h'(y')$ were still constructed from time derivatives at identical collections of time points. At each time point, we attempted to compute the rescaled signal from the signal time derivatives encountered during the most recent 10 ms (ΔT = 10 ms). At some times, the signal could not be rescaled because the signal level at that time was not attained during the previous 10 ms, and, therefore, there were no contributions to the right side of Eq.(3) for some values of *y*. For example, this happened at $t \sim 71$, 73, 84, 90, and 97 ms in Fig. 3. At such times, a signal invariant could not be computed. As mentioned in Section 2, this occurs at identical time points when dynamic rescaling

is applied to the "untransformed" signal (e.g., Fig. 3a) and to any transformed version of it (e.g., Fig. 3d). This means that the *s* representations of these signals are non-existent at identical time points and that at all other times they exist and have the same values. Therefore, this phenomenon does not corrupt the invariance of the signal's *s* representation, although it does reduce its information content. In this experiment, the *s* representation could be computed at more than 90% of all time points.

Figures 3f and 3g illustrate the effect of noise on dynamic rescaling. Figure 3f was derived from Fig. 3d by adding white noise so that the signal-to-noise ratio is 3.5. This causes a pronounced hiss to be superposed on the word "open" when one plays the entire 406 ms sound exemplified by Fig. 3f. Figure 3g is the *s* representation, derived by dynamically rescaling Fig. 3f with ΔT = 10 ms. Comparison of Figs. 3g, 3e, and 3b shows that the noise has caused some degradation of the invariance of the *s* representation. This is expected because additive noise ruins the invertibility of the transformations relating Fig. 3f to Figs. 3d and 3a, thereby violating the proof of the invariance of *S* in Section 2. The noise sensitivity of the *s* representation can be decreased by increasing ΔT, because this increases the number of contributions to the right side of Eq.(3), which tends to "average out" the effects of noise. However, such an increase in ΔT means that more time is required for the dynamic rescaling process to adapt to a sudden change in signal transformation, as mentioned at the end of Section 2.B.

3.B. Experiments with 2D Sensor State Manifolds

In this section, we demonstrate the method in Section 2.C by applying it to simulated data on a two-dimensional sensor state manifold. Let $x = (x_1, x_2)$ represent the sensor state of the device. 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. Suppose that Figure 4a represents the trajectories of the sensor states that were previously encountered by the system. Notice that these lines tend to be oriented in nearly horizontal or vertical directions, thereby endowing the manifold with directionality at each point. We used these data to compute the local vectors h_a 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 neighborhood of each sample point, and the time derivative of each trajectory segment traversing the neighborhood was computed at equal time intervals. Then, Eqs.(9-13) with $p=1$ were applied in order to derive local vectors from the collection of time derivatives at each sample point. The resulting vectors h_a, shown in Figure 4b, were then interpolated in order to estimate the vectors at intervening points. As expected, these vectors reflect the horizontal and vertical orientations of the trajectories from which they were derived. Finally, Eqs.(1-2) were applied to these h_a in order to compute the coordinate-independent representation s_a of each

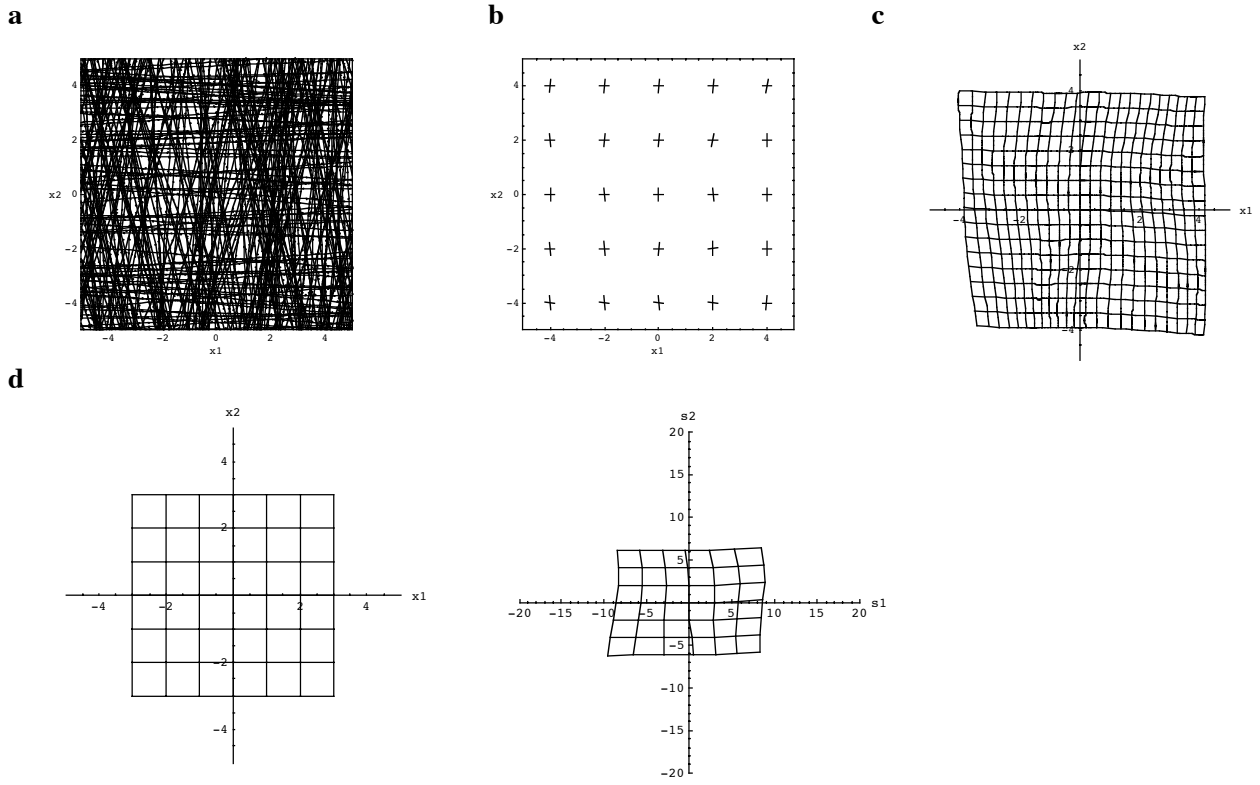

Figure 4. a) The simulated trajectory of recently encountered sensor states $x(t)$. The speed of traversal of each trajectory segment is indicated by the dots, which are separated by equal time intervals. The nearly horizontal and vertical segments are traversed in the left-to-right and bottom-to-top directions, respectively. b) The local preferred vectors h_a that were derived from the data in *a* by means of the method in Section 2.C. The nearly horizontal and vertical lines denote vectors that are oriented to the right and upward, respectively. c) The level sets of $s(x)$, which show the intrinsic coordinate system or scale derived by applying the method in Section 2.C to the data in *a*. The nearly vertical curves are loci of constant s_1 for evenly spaced values between –11 (left) and 12 (right); the nearly horizontal curves are loci of constant s_2 for evenly spaced values between –8 (bottom) and 8 (top). d) The coordinate-independent representation (right figure) of a grid-like array of sensor states (left figure), obtained by using the scale in *c* to rescale those sensor states.

sensor state on the manifold, relative to the reference state which was chosen to be $x_0 = (0,0)$. The result is shown in Figure 4c, which depicts the level sets of the scale function $s_a(x)$ that is intrinsic to the sensor state history in Figure 4a. Figure 4d shows how an "image" of sensor states in the *x* coordinate system is represented in the *s* coordinate system.

Next, we considered what would have happened if the same device had "experienced" the sensor states shown in Figure 5a. These trajectories are related to those in Figure 4a by the following non-linear transformation:

$$x_1 \rightarrow 0.1 + x_1 + 0.1x_2 + 0.01x_1^2 - 0.02x_2^2 - 0.01x_1x_2 \qquad \text{(Eq. 25)}$$

$$x_2 \rightarrow 0.2 - 0.2x_1 + x_2 - 0.01x_1^2 + 0.02x_2^2 + 0.01x_1x_2$$

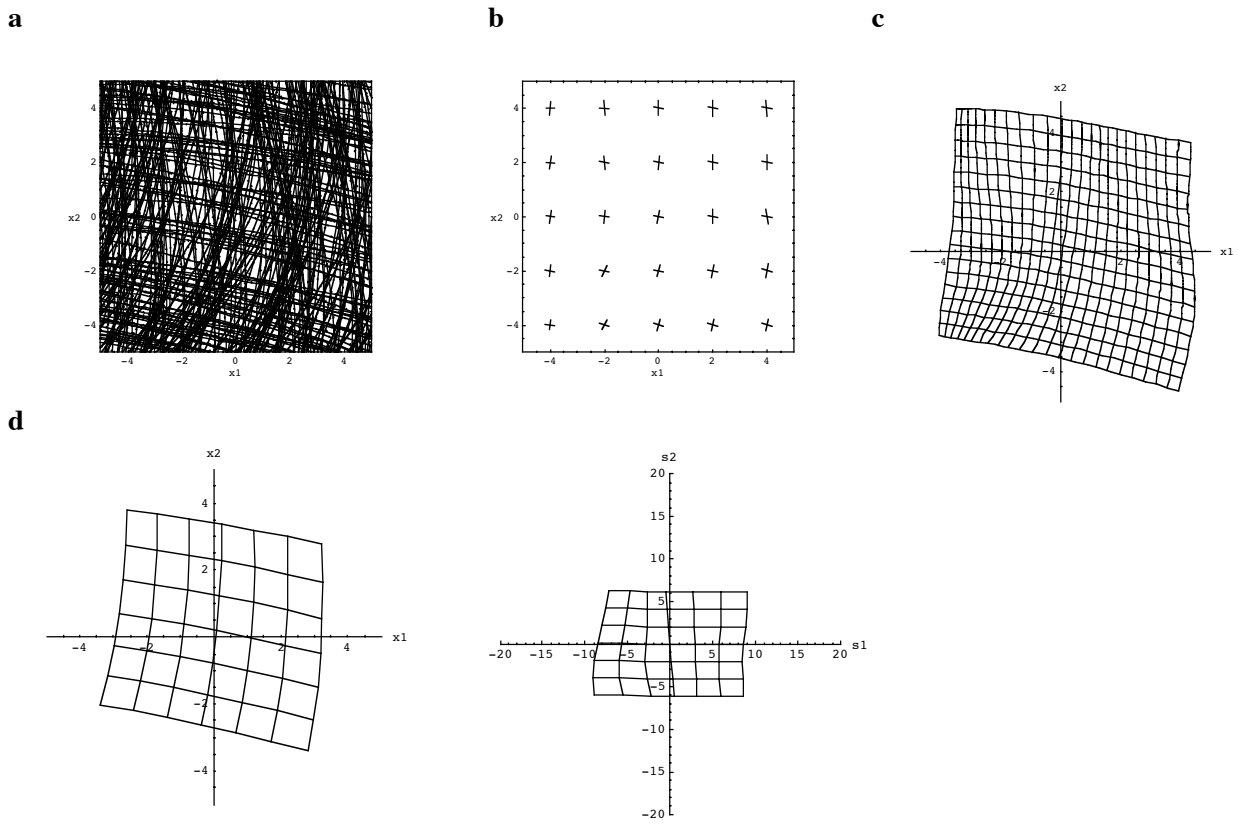

Figure 5. a) The simulated trajectory of recently encountered sensor states $x(t)$ that are related to those in Fig. 4a by the coordinate transformation in Eq.(25). The speed of traversal of each trajectory segment is indicated by the dots, which are separated by equal time intervals. The nearly horizontal and vertical segments are traversed in the left-to-right and bottom-to-top directions, respectively. b) The local preferred vectors h_a that were derived from the data in Fig. 5a by means of the method in Section 2.C. The nearly horizontal and vertical lines denote vectors that are oriented to the right and upward, respectively. c) The level sets of $s(x)$, which shows the intrinsic coordinate system or scale that was derived by applying the method in Section 2.C to the data in Fig. 5a. The vertical curves are loci of constant s_1 for evenly spaced values between –12 (left) and 11 (right); the horizontal curves are loci of constant s_2 for evenly spaced values between –9 (bottom) and 7 (top). d) The coordinate-independent representation (right figure) of an array of sensor states (left figure), obtained by rescaling the sensor states by means of the scale in Fig. 5c. The panel on the left was created by subjecting the corresponding left panel in Fig. 4d to the coordinate transformation in Eq.(25). Notice that the right panel is nearly identical to the one in Fig. 4d, thereby confirming the fact that these representations are invariant under the coordinate transformation.

For example, suppose that the sensor state is the location of a feature in a digital image. Equation (25) could represent the way the sensor states are transformed by a distortion of the optical/electronic path within the camera or by a distortion of the surface on which the camera is focused (e.g., distortion of a printed page). The procedure outlined above was used to compute the local vectors on a uniform grid of sample points. Figure 5b shows the resulting vectors, which are oriented along the principal directions

apparent in Figure 5a. Next, interpolation was used to estimate the h_a at intervening points, and Eqs.(1-2) were used to compute the coordinate-independent representation s_a of each sensor state on the manifold, relative to the reference sensor state which was chosen to be $x_0 = (0.1, 0.2)$. Notice that we have assumed prior knowledge of the transformed position of the reference sensor state. In other words, we have assumed that we have the prior knowledge necessary to identify this state both before and after the onset of the process, which remaps the sensor states. The result of this calculation is shown in Figure 5c, which depicts the level sets of the function $s_a(x)$, the scale function inherent to the sensor state data in Figure 5a. These functions were used to compute the s_a representation of the transformed version of the "image" in the left panel of Figure 4d. The transformed image and its s_a representation are shown in Figure 5d. Comparison of Figure 4d and Figure 5d shows that the s_a representations of the untransformed and transformed images are nearly identical. Thus, the stimulus representations are invariant in the presence of unknown invertible transformations of sensor states, such as the one in Eq.(25). The tiny discrepancies between the right sides of Figure 4d and Figure 5d can be attributed to errors in the interpolation of the h_a, which is due to the coarseness of the grid on which h_a was sampled and to the size of the neighborhoods used to determine the vectors at each sample point. This error can be reduced if the distance between sample points and the size of the neighborhood around each sample point 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 Figures 4a and 5a) so that even tiny neighborhoods contain enough data to compute the h_a.

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. The scale values assigned to data points by conventional methods of multidimensional scaling or dimensional reduction are not transformation-independent in this way. In Section 1, we argued that this 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 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 observes two evolving facial stimuli, F and F', 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). Similarly, suppose V recorded two sensor state time series that differed by a scale factor but had identical rescaled representations. 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 and 2001c). 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 Section 2.C). Specifically, there must be enough sensor state trajectory segments near each point to endow the manifold with local structure (local vectors). Thus, the device must have sufficient "experience" in order to form 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 two representations of

the same sensor state may be referred to different collections of recently encountered sensor states. In other words, the representation of an unchanged stimulus may be time-dependent because the representations are derived from the device's recent "experience" and that experience 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 h_a). Thus, the stability of the stimulus representation depends on the stationarity of the vectors h_a that are used to create that representation. 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.

The family of all signals $x(t)$ that rescale to a given function $s(t)$ can be considered to form an equivalence class. If such a class includes a given signal, it also includes all invertible transformations of that signal. Signals can be assigned to even larger equivalence classes of all signals that lead to the same result when rescaling is applied N times in succession, where $N \geq 2$. For example, suppose that two signals do not have the same representation after one application of rescaling, but the same function is produced by two applications of rescaling. Then, these two signals can be considered to be equivalent at a deeper level, in the sense that the "inner" properties of their "inner" properties are the same. Successive applications of rescaling may eventually create a function that is not changed by further applications of the procedure (i.e., the serial rescaling process may reach a fixed "point"). For example, it is easy to show that, if the scaling of a one-dimensional signal is time-independent (i.e., if $h(y)$ and $s(x)$ are time-

independent), it will rescale to such a fixed point. Such a signal is loosely analogous to music, in the sense that musical compositions are also based on a time-independent scale (e.g., the equally tempered scale of Western music).

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, 1997; 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 and 2001c). The rescaling technique in this paper could also be used to design a speech-like communications system that is resistant to information loss 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). Finally, it should be mentioned that the analysis of a non-linear dynamical system (e.g., Gouesbet, 1997) may be enhanced by rescaling its trajectories. Such a rescaled trajectory depicts "inner" properties of the system that are independent of the nature of the observed measurements (e.g., delay coordinates vs time derivatives vs multiple scalar measurements). However, a rescaled representation can only be computed in those parts of state space that the trajectory samples with sufficient density.

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 experienced 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.

Gouesbet, G., Le Sceller, L., Letellier, C., Brown, R., Buchler, J. R., Kollath, Z. (1997). Reconstructing a dynamics from a scalar time series. *Annals of the New York Academy of Sciences*, **808**, 25-49.

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.

Levin, D. N. (2001c). Representations of sound that are insensitive to spectral filtering and parameterization procedures. Submitted for publication.

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*, New York: 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.

Sauer, T., Yorke, J. A., Casdagli, M. (1991). Embedology. *Journal of Statistical Physics*, **65**, 579-616.

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.

Takens, F. (1981). Detecting strange attractors in turbulence. In: Rand, D. A. and Young, L.-S. (Eds.), *Lecture Notes in Mathematics, Vol. 898: Dynamical Systems and Turbulence, Warwick, 1980*. Berlin: Springer-Verlag, pp. 368-381.

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.

Whitney, H. (1936). Differentiable manifolds. *Annals of Mathematics*, **37**, 645-680.

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 $g_1 = 0.5$ and $g_2 = 150$;

Figure 3. a) A signal obtained by digitizing the acoustic signal of the word "open", uttered by a male speaker of American English. A 40 ms segment of the 406 ms signal is shown, with time given in ms. The horizontal lines show signal amplitudes that have dynamically rescaled values equal to $s = \pm 50n$ for *n=1, 2, … .* b) The signal $S(t)$ (in units of μs) obtained by dynamically rescaling the signal in panel *a*, with the parameter ΔT = 10 ms. c) The non-linear function $x'(x)$ that was used to transform the signal in panel *a* into the one in panel *d*. d) A distorted version of the signal in panel *a*, obtained by applying the non-linear transformation in panel *c*. e) The signal obtained by dynamically rescaling the signal in panel *d* with the parameter ΔT = 10 ms. f) A signal derived from the signal in *d* by adding white noise so that the signal-to-noise ratio is 3.5. g) The signal obtained by dynamically rescaling the signal in panel *f* with ΔT = 10 ms.

Figure 4. 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 local preferred vectors h_a that were derived from the data in *a* by means of the method in Section 2.C. The nearly horizontal and vertical lines denote vectors that are oriented to the right and upward, respectively. c) The level sets of $s(x)$, which show the intrinsic coordinate system or scale derived by applying the method in Section 2.C to the data in *a*. The nearly vertical curves are loci of constant s_1 for evenly spaced values between –11 (left) and 12 (right); the nearly horizontal curves are loci of constant s_2 for evenly spaced values between –8 (bottom) and 8 (top). d) The coordinate-independent representation (right figure) of a grid-like array of sensor states (left figure), obtained by using the scale in *c* to rescale those sensor states.

Figure 5. a) The simulated trajectory of recently encountered sensor states $x(t)$ that are related to those in Fig. 4a by the coordinate transformation in Eq.(25). 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 local preferred vectors h_a that were derived from the data in Fig. 5a by means of the method in Section 2.C. The nearly horizontal and vertical lines denote vectors that are oriented to the right and upward, respectively. c) The level sets of $s(x)$, which shows the intrinsic coordinate system or scale that was derived by applying the method in Section 2.C to the data in Fig. 5a. The vertical curves are loci of constant s_1 for evenly spaced values between –12 (left) and 11 (right); the horizontal curves are loci of constant s_2 for evenly spaced values between –9 (bottom) and 7 (top). d) The coordinate-independent representation (right figure) of array of sensor states (left figure), obtained by rescaling the sensor states by means of the scale in Fig. 5c. The panel on the left was created by subjecting the corresponding left panel in Fig. 4d to the coordinate transformation in Eq.(25). Notice that the right panel is nearly identical to the one in Fig. 4d, 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.