Sensor-independent stimulus representations

David N. Levin*

Department of Radiology, University of Chicago, Chicago, IL 60637

Communicated by Paul C. Lauterbur, University of Illinois at Urbana–Champaign, Urbana, IL, March 22, 2002 (received for review October 8, 2001)

This paper shows how time-dependent sensory data from an evolving stimulus can be blindly rescaled in a nonlinear time-dependent fashion 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: (*i*) altered performance of a device's detector; (*ii*) changes in the observational environment external to the sensory device and the stimulus; and (*iii*) 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.

Humans form percepts that are remarkably independent of the condition of their sensors. This fact was strikingly illustrated by experiments in which subjects wore goggles, creating severe geometric distortions of the observed scene (1). 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. 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.

This paper shows how to build sensory devices that mimic this sensor-independent characteristic of human perception. Specifically, the paper demonstrates how time-dependent sensory data from an evolving stimulus can be rescaled in a nonlinear time-dependent fashion to create a time series of stimulus representations that are invariant under any unknown invertible transformation of the sensory data. This result 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. As an illustration, 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: (*i*) 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); and (*ii*) 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. Thus, the rescaling process enables devices of this kind to ''see'' the world in the same sensor-independent way.

In the above, it was assumed that the sensor states of each device are invertibly related to the underlying stimulus configurations. The embedding theorems of nonlinear dynamics (2) 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 sensor states of almost all such devices will have the same rescaled representations.

The method in this paper differs significantly from techniques for multidimensional scaling or dimensional reduction (3). In each of these methods, it is necessary to impose an *ad hoc* measure of ''distance'' between each pair of neighboring data points (4) or, at least, to rank the distances between neighboring points (5). In each case, the defined distances or rankings are not invariant under general nonlinear 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 stated that the above-mentioned 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 *Theory*, 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. The invariance reflects the fact that 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. As an illustration, consider the following analogy: the physical rotation or translation of a collection of particles in a

*E-mail: d-levin@uchicago.edu.

www.pnas.org/cgi/doi/10.1073/pnas.102170499

plane does not alter the coordinates of each particle in the collection's "natural" internal coordinate system (the coordinate system originating at the collection's center of mass and oriented along the collection's principal axes), even though the absolute coordinates of each particle are transformed. In this situation, the rotation/translation transforms the internal coordinate system and each particle's absolute coordinates in the same way, without disturbing their relationship.

In *Theory*, we show how tensor calculus can be used to find a transformation-independent representation of each sensor state in a time series (ref. 6; also see http://www.geocities.com/dlevin2001/reprint1.html, http://lanl.arXiv.org/abs/cs.CL/0204003, and http://www.geocities.com/dlevin2001/preprint3.html). The theoretical framework is developed with sufficient generality to handle sensor states having any number of components. In *Experiments with Simulated Data*, the method is then illustrated by applying it to simulated sensor state data having two dimensions, and the implications of these results are addressed in *Discussion*. The simpler special case of one-dimensional sensor states was discussed in other reports (refs. 7 and 8; also see http://www.geocities.com/dlevin2001/reprint2.html and http://www.geocities.com/dlevin2001/preprint1.html), where it was applied to analytic examples, acoustic waveforms of human speech, spectral time series of a bird song, and spectral time series of speech-like sounds.

Theory

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 nonlinear fashion. 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. This processing may also include linear or nonlinear dimensional reduction procedures that map the detector outputs onto sensor states with the same dimensionality as the underlying degrees of freedom of the stimulus (refs. 4 and 5 and refs. 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. This function defines a trajectory on the manifold of all possible sensor states.

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: (*i*) a reference sensor state (x_0) that serves as the origin of the sensor state scale; (*ii*) 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 ($x(0) = x_0, x(1) = x$); and (*iii*) 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 $h \rightarrow h' = (\partial x'/\partial x)h$ under the change of coordinate systems $x \rightarrow x'$ (9). 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 (Fig. 1):

$$\delta x = \sum_{a = 1, \ldots, N} h_a \delta s_a. \qquad [1]$$

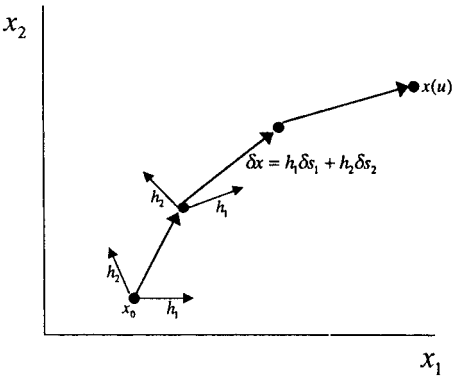

Fig. 1. Consider a path $x(u)$ ($0 \leq u \leq 1$) between a reference sensor state x_0 and a sensor state of interest. If vectors h_a can be defined at each point along the path, each line segment δx can be decomposed into its components δs_a along the vectors at that point.

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 (6, 10):

$$s = \int_{x_0}^{x} \delta s. \qquad [2]$$

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 nonlinear) 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 this paragraph, we show how to derive a coordinate-independent representation of the sensor state at any particular time from the sensor states encountered in a chosen time interval (e.g., the most recent time interval of length ΔT). 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. Here, ΔT is a parameter that can be chosen freely, although it influences the adaptivity and noise sensitivity of the method (see *Discussion*). The first step is to use the sensor state data in the chosen time interval to define local vectors $h_a(x)$ in a coordinate-independent manner. 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: $\hat{h}_i = (dx/dt)|_{t_i}$, where t_i denotes the ith 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 $\hat{h}'_i = (dx'/dt) = (\partial x'/\partial x)(dx/dt) = (\partial x'/\partial x)\hat{h}_i$. 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. The first step is to pick an integer $C \geq N$ and to partition the indices i into C nonempty sets labeled S_c where $c = 1, \ldots, C$. Next, compute the $N \times N$ covariance matrix M_c of the vectors corresponding to each set of indices:

$$M_c = \frac{1}{N_c} \sum_{i \in S_c} \hat{h}_i \hat{h}_i, \quad [3]$$

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 (9); namely, if coordinates on the manifold are transformed as $x \rightarrow x'$, then $|M_c| \rightarrow |M'_c| = |\partial x'/\partial x|^2 |M_c|$. This result follows from the facts: (*i*) $M_c \rightarrow M'_c = (\partial x'/\partial x) M_c (\partial x'/\partial x)^T$, where T denotes transpose, and (*ii*) a matrix and its transpose have the same determinant. Next, compute E, which is defined to be the sum of powers of these determinants:

$$E = \sum_c |M_c|^p, \quad [4]$$

where p is some real positive number. The transformation property of $|M_c|$ 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 nonempty 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:

$$h_c = \frac{1}{N_c} \sum_{i \in S_c} \hat{h}_i. \quad [5]$$

Because the $\hat{h}_i$ are contravariant vectors, the h_c will also transform as contravariant vectors as long as they are optimally 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. 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 understand this property of Eq. **2**, note that Eq. **1** can be inverted to form $\delta s_a = \tilde{h}_a \cdot \delta x$, where the covariant vectors $\tilde{h}_a$ are found by solving $\Sigma_{a=1,\ldots,N} \tilde{h}_{ak} h^l_a = \delta^l_k$, δ^l_k, 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: $(\partial \tilde{h}_{ak}/\partial x_l) - (\partial \tilde{h}_{al}/\partial x_k) = 0$. 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 (6, 10): 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 ($1 \leq n \leq N$) 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; 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**. It follows 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 computed only 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 nonparametric interpolation (e.g., splines or neural nets, respectively). The resulting vectors will be correct as long as the spacing between the sample points and the sizes of the small neighborhoods around them are small relative to the distance over which the directionality of the manifold varies. If the data are transformed into any other coordinate system by a transformation that is smooth enough to be nearly linear over these neighborhoods, they will lead to the same vectors, and, therefore, to the same rescaled representation of each sensor state. Thus, it is desirable for the sensor state trajectory to cover the manifold as densely as possible so that vectors can be computed in the smallest possible neighborhoods, making the derived rescaled representations invariant under the largest possible class of coordinate transformations. However, even if the manifold has been covered sufficiently densely, special circumstances may prevent the derivation of vectors h_a at some locations. For example, suppose that there is no unique way of partitioning the $\hat{h}_i$ at a point to minimize E, or suppose that the h_c (Eq. **5**) 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 at every point. 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 require only 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 ref. 6. (also see the web sites cited in the Introduction)

In the above discussion, it was assumed that the reference state was chosen in a coordinate-independent manner; i.e., the sensor state data in the x and x' coordinate systems were rescaled with respect to reference states related by $x'_0 = x'(x_0)$. For example, this condition could be satisfied by choosing the reference state to be the sensor state that is the local maximum of a function x

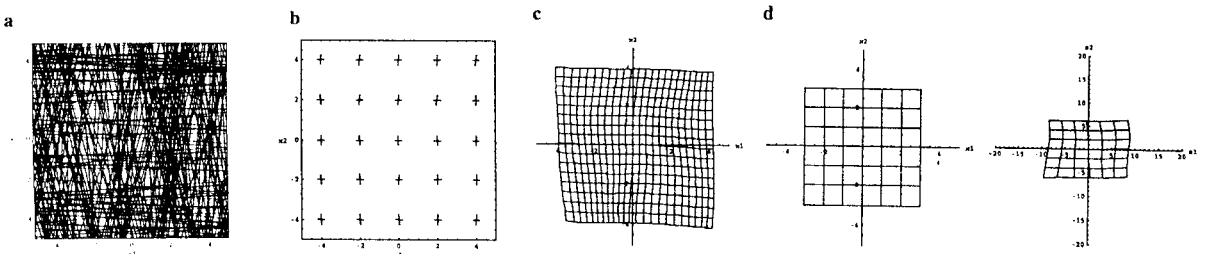

Fig. 2. (*a*) The trajectory of simulated sensor states $x(t)$ to be used to derive the rescaling of a current sensor state. The speed of traversal of each trajectory segment is indicated by the dots, which are separated by equal time intervals. The nearly horizontal and vertical segments are traversed in the left-to-right and bottom-to-top directions, respectively. (*b*) The local vectors h_a that were derived from the data in *a* by means of the method in *Theory*. 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 *Theory* 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*) of a grid-like array of sensor states (*Left*), obtained by using the scale in *c* to rescale those sensor states.

defined to be the number of times x is encountered in a chosen time interval. Alternatively, prior knowledge may be used to choose the reference state. For instance, 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 instance, 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 to "show" each singer the origin of the desired musical scale. Notice that stimulus representations that are referred to different reference stimuli may 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.

Experiments with Simulated Data

In this section, the method in *Theory* is demonstrated 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 Fig. 2*a* depicts the sensor states to be used to derive the rescaling of a current sensor state. Notice that these trajectory segments 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. Specifically, we considered a small neighborhood of each sample point, and the time derivatives of each trajectory segment traversing the neighborhood were computed at equal time intervals. Then, Eqs. **3–5** with $p = 1$ were applied to derive local vectors from the collection of time derivatives at each sample point. The resulting vectors h_a, shown in Fig. 2*b*, were then interpolated to estimate the vectors at intervening points. As expected, these vectors reflect the horizontal and vertical orientations of the trajectory segments from which they were derived. Finally, Eqs. **1** and **2** were applied to these h_a to compute the coordinate-independent representation s of each sensor state on the manifold, relative to the reference state that was chosen to be $x_0 = (0, 0)$. The result is shown in Fig. 2*c*, which depicts the level sets of the scale function $s(x)$ that is intrinsic to the sensor state history in Fig. 2*a*. Fig. 2*d* 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 Fig. 3*a*. These trajectory segments are related to those in Fig. 2*a* by the following nonlinear transformation:

$$x_1 \rightarrow 0.1 + x_1 + 0.1x_2 + 0.01x_1^2 - 0.02x_2^2 - 0.01x_1x_2$$

$$x_2 \rightarrow 0.2 - 0.2x_1 + x_2 - 0.01x_1^2 + 0.02x_2^2 + 0.01x_1x_2. \quad \textbf{[6]}$$

For example, suppose that the sensor state is the location of a feature in an evolving digital image. Eq. **6** could represent the way the sensor states are transformed by a distortion in the

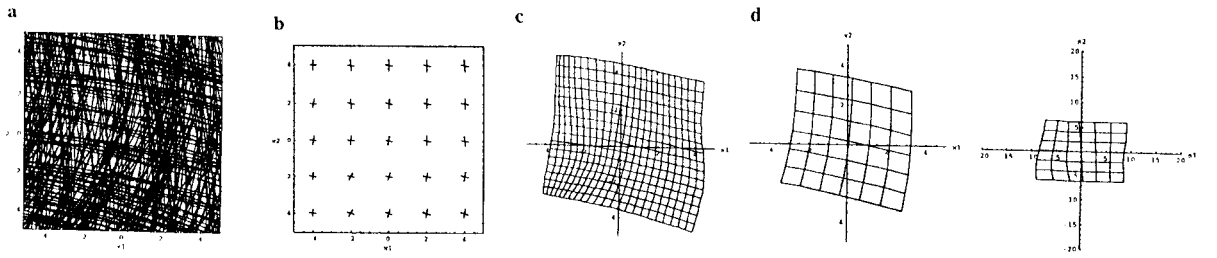

Fig. 3. (*a*) The trajectory of simulated sensor states $x(t)$ that are related to those in Fig. 2*a* by the coordinate transformation in Eq. **6**. (*b*) The local preferred vectors h_a that were derived from the data in *a* by means of the method in *Theory*. (*c*) The level sets of $s(x)$, which show the intrinsic coordinate system or scale that was derived by applying the method in *Theory* to the data in *a*. 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*) of the array of sensor states (*Left*), obtained by rescaling the sensor states on the left by means of the scale in *c*. *Left* was created by subjecting the corresponding panel in Fig. 2*d Left* to the coordinate transformation in Eq. **6**.

optical/electronic path within the camera, by distortions in the optical channel between the imaged object and the camera, or by a distortion of the object itself (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. Fig. 3*b* shows the resulting vectors, which are oriented along the principal directions apparent in Fig. 3*a*. Next, interpolation was used to estimate the h_a at intervening points, and Eqs. **1** and **2** were used to compute the coordinate-independent representation s of each sensor state on the manifold, relative to the reference sensor state that was chosen to be $x_0 = (0.1, 0.2)$. Notice that we have assumed prior knowledge of the transformed coordinates 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 Fig. 3*c*, which depicts the level sets of the function $s(x)$, the scale function inherent to the sensor state data in Fig. 3*a*. This function was used to compute the s representation of the transformed version of the "image" in Fig. 2*d Left*. The transformed image and its s representation are shown in Fig. 3*d*. Comparison of Figs. 2*d* and 3*d* shows that the s representations of the untransformed and transformed images are nearly identical. The s representations are nearly the same because the unrescaled images are the same image viewed in different coordinate systems, and the rescaling process creates a coordinate-independent description of the image's relationship to the intrinsic structure (h_a vectors) of the manifold. The tiny discrepancies between the right sides of Figs. 2*d* and 3*d* can be attributed to errors in the interpolation of the h_a, which are 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 improvement is possible if the device is allowed to experience a denser set of sensor states (i.e., more trajectory segments than shown in Figs. 2*a* and 3*a*), so that even tiny neighborhoods contain enough data to compute the h_a.

Discussion

In this paper, we demonstrated how time-dependent sensory data from an evolving stimulus could be rescaled in a nonlinear 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. In the Introduction, 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 invertibly related. Thus, the rescaled sensor state time series can be considered to reflect an "inner" property of the time series of stimulus configurations. In a sense, this is *why* 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.

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 may include: (*i*) 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); (*ii*) 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); (*iii*) systematic modifications of the presentation of the stimuli themselves (e.g., systematic warping of printed pages or of a voice). 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 understand this fact, 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. For example, a computer vision system would create the same rescaled representations of the time series of facial expressions of two different faces, as long as there were a time-independent invertible mapping that related each expression of one face to the corresponding expression of the other one (i.e., as long as the faces consistently mimicked each other).

Strictly speaking, there is more than one way to interpret such an observation; i.e., the observation of two 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: (*i*) intrinsic differences in the stimuli themselves; (*ii*) the presence of a process that affected the device's detector or the "channel" between it and the stimulus. Of course, humans can suffer from illusions because of 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 were 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 difference occurs because the device's 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. A transformation-independent representation will be recovered because the description of each subsequently encountered sensor state will be referred to the properties of a collection of transformed sensor states. Thus, like a human, the system adapts to the presence of the transformation after a period of adjustment. This phenomenon has been illustrated by experiments reported elsewhere (7, 8). 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 *Theory*). Specifically, there must be enough sensor state trajectory segments near each point to endow the manifold with local structure (local vectors) that can serve as "landmarks." Thus, the device must have sufficient "experience" to form stimulus representations, reminiscent of the role of experience in the acquisition of vision by human infants (11). 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 (7, 8). 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. Specifically, two representations may differ because they are referred to different collections of recently encountered sensor states. In other words, the representation of a recurring 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. As an illustration, imagine 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 constraint 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 (1). Specifically, each subject's perception of stimuli returned to the pregoggle 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 nonlinear signal-processing method presented in this paper could be used as a representation "engine" in the "front end" of intelligent sensory devices (D.N.L., patents pending). 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 to account for these factors. Other reports (refs. 6–8; also see web sites cited in the Introduction) discuss possible applications to systems for: (*i*) computer vision, (*ii*) channel- and speaker-independent speech recognition, and (*iii*) channel-independent telecommunication of information.

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 invariance of a stimulus even though its "appearance" is changing because of extraneous factors. These phenomena have been the subject of philosophical discussion since the time of Plato (e.g., the allegory of "The Cave"), and they have also intrigued modern neuroscientists (12). 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 experienced stimuli. A significant evolutionary advantage would accrue to organisms that developed the ability to process sensory information in this way.

1. Kohler, I. (1972) in *Perception: Mechanisms and Models*, eds. Held, R. & Richards, W. (Freeman, San Francisco), pp. 299–309.
2. Sauer, T., Yorke, J. A. & Casdagli, M. (1991) *J. Stat. Phys.* **65,** 579–616.
3. Cox, T. & Cox, M. (1994) *Multidimensional Scaling* (Chapman & Hall, London).
4. Tenenbaum, J. B., de Silva, V. & Langford, J. C. (2000) *Science* **290,** 2319–2323.
5. Roweis, S. T. & Saul, L. K. (2000) *Science* **290,** 2323–2326.
6. Levin, D. N. (2001) *Proc. SPIE Int. Soc. Opt. Eng.* **4322,** 1677–1688.
7. Levin, D. N. (2001) *Proceedings of the International Conference on Advances in Infrastructure for Electronic Business, Science, and Education on the Internet* (Scuola Superiore G. Reiss Romoli, L'Aquila, Italy, August 6–12).
8. Levin, D. N. (2001) *J. Acoust. Soc. Am.*, in press.
9. Schrodinger, E. (1963) *Space–Time Structure* (Cambridge Univ. Press, Cambridge, U.K.).
10. Levin, D. N. (2000) *J. Math. Psychol.* **44,** 241–284.
11. Hebb, D. O. (1949) *The Organization of Behavior* (Wiley, New York).
12. Zeki, S. (1999) *Inner Vision: An Exploration of Art and the Brain* (Oxford Univ. Press, Oxford, U.K.).