Channel-Independent and Sensor-Independent Stimulus Representations

David N. Levin*, Department of Radiology, University of Chicago

ABSTRACT

It is a remarkable fact that different individuals perceive similar relationships among stimuli even though they observe those stimuli through different channels and with different sensory organs and sensory cortices. This paper shows how to design intelligent sensory devices that represent stimulus relationships in a similarly channel-independent and sensor-independent manner. First, we demonstrate that a device's channel and sensors define a coordinate system that the device imposes on the space of stimulus states. Then, differential geometry is applied to find coordinate-system-independent statistical characteristics of the stimulus trajectory, and these are used to derive coordinate-system-independent relationships among stimuli. Devices, which are built on these principles but have different channels and sensors, will represent stimulus relationships in the same way, as long as they have statistically similar histories of previously-encountered stimuli. In an intelligent sensory device, this kind of representation "engine" could function as a "front end" that passes invariant stimulus representations to a pattern recognition 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 effects.

I. INTRODUCTION

It is difficult to describe a perceptual experience in an *absolute* manner (i.e., without making reference to perceptions of other stimuli), and, in fact, the very notion of absolute perceptual experience may be ill-defined. Instead, people usually describe their perceptions in a *relative* manner; i.e., in relation to their perceptions of other stimuli. Thus, if an individual is asked to describe a light or a shape or a sound, he/she is likely to compare it to his/her perception of other lights or shapes or sounds. People may make certain "primitive" perceptual judgments that are used to construct

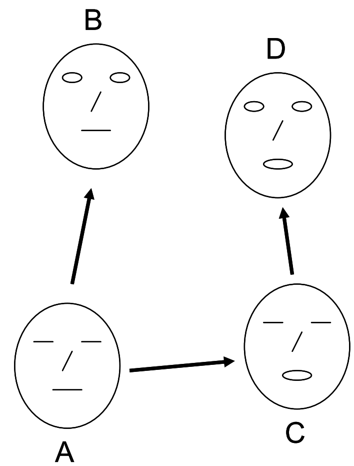

Figure 1. Consider a two dimensional space populated by faces characterized by the extent to which the mouth and eyes are open. Given two faces (*A* and *B*) that are related by a small transformation ($A \rightarrow B$, called an "eye-opening" transformation) and given another face (*C*) that differs from *A* by another small transformation ($A \rightarrow C$, called a "mouth-opening" transformation), observers will often judge a particular transformation of *C*, namely $C \rightarrow D$, to be equivalent to $A \rightarrow B$. In other words, they will perceive the stimuli as being related by the analogy: *A:B = C:D*.

the perceived relationships between stimuli. For example, in most circumstances, observers have a strong sense about whether two stimuli are the same or not; i.e., whether stimulus *A* = *B* or not. Likewise, observers often feel that stimulus "analogies" are true or false. For instance (Fig. 1), suppose that stimuli *A* and *B* differ by one small stimulus transformation, stimuli *C* and *D* differ by another small transformation, and stimuli *A* and *C* differ by a third small transformation. Observers will often have a definite sense about whether the relationship between *A* and *B* is the same (or is not the same) as the relationship between *C* and *D*; i.e., whether *A:B = C:D* or not. A series of such perceived analogies can be concatenated in order to describe relationships between two "distant" stimuli [4]. Figure 2 shows an example in which an observer describes stimulus *E* as being related to stimuli *A*, *B*, *C*, and *D* by a succession of stimulus transformations: "*E* is the stimulus that is produced by starting with stimulus *A* and performing three transformations perceptually equivalent to the $A \to C$ transformation, followed by two transformations perceptually equivalent to the $A \to B$ transformation". In general, an observer, who perceives local stimulus analogies at each point in stimulus space, can describe the relationship between two stimuli in terms of various paths connecting them. Geographic navigation without the help of measuring instruments is the special case in which the observer makes basic perceptual judgments about directions and distance scales at each spatial point. These can then be used to describe relationships between distant points in terms of the paths connecting them (i.e., "The Sears Tower is the building that is reached by starting at the John Hancock building, then traveling 1000 feet west, and finally traveling 2500 feet south").

Given a space of stimuli, an arbitrarily-chosen set of perceptual analogies could be used to navigate among them. What principle might guide a human brain when it chooses a particular set of analogies in order to "navigate" among the stimuli of the world? A clue may be provided by the fact that similar perceptions are often experienced by different individuals with similar past experiences. This approximate "universality" of perception is remarkable in the following way. Each person only has direct access to the state of his/her brain's sensory cortices, and, in the presence of a single stimulus, these states differ from person to person, due to differences in their sensory organs and neuronal connections. For example, stimuli *A*, *B*, *C*, and *D* may create complex spatial distributions of neuronal electrical discharges in the sensory cortex of observer Ob (denoted x_A, x_B, x_C, and x_D), and the same stimuli may produce different electrical current distributions in the sensory cortex of observer Ob' (x'_A, x'_B, x'_C, and x'_D). Yet, somehow, as long as Ob and Ob' have similar histories of previous stimulus exposure, they tend to independently conclude that these brain states satisfy $x_A : x_B = x_C : x_D$ and $x'_A : x'_B = x'_C : x'_D$, respectively. A related fact is that each individual's perceptions of stimuli are remarkably independent of the channel through which he/she views those stimuli. This fact was strikingly illustrated by experiments in

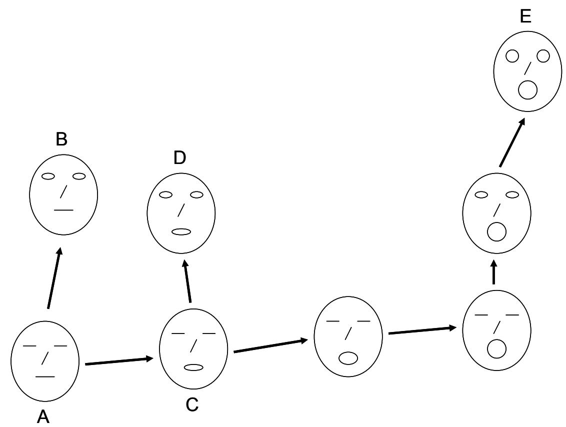

Figure 2. "Distant" stimuli can be related by describing a path connecting them as a sequence of perceptually defined stimulus transformations. For example, an observer may describe *E* as the stimulus that is produced by starting at *A* and performing three transformations perceptually equivalent to the mouth-opening transformation $A \to C$, followed by two transformations perceptually equivalent to the eye-opening transformation $A \to B$.

which subjects wore goggles that created severe geometric distortions and inversions of the observed scenes [2, 14]. Although each subject initially perceived the distortion of the new channel, his/her perceptions of the world tended to return to the pre-experimental baseline after several weeks of constant exposure to familiar stimuli seen through the goggles. There are similar auditory phenomena: if an individual hears a musical composition performed through different acoustic channels (e.g., in different rooms), the notes in the first rendition are perceived to be related to one another in the same way as the notes in the second performance; i.e., the melody is perceived to be the same even though the two performances produce different acoustic waveforms and different auditory cortical states. These observations suggest the following generalization. Consider an individual who has a certain history of stimulus exposure and "learns" to perceive a relationship $x_A : x_B = x_C : x_D$ among certain stimulus-induced brain states. If that person is exposed to a similar history of stimuli through a new observational channel, he/she will perceive the same relationships ($x'_A : x'_B = x'_C : x'_D$) among the transformed brain states. In other words, the above-mentioned phenomena suggest that perceived relationships among the states of brain sensory cortex are invariant under

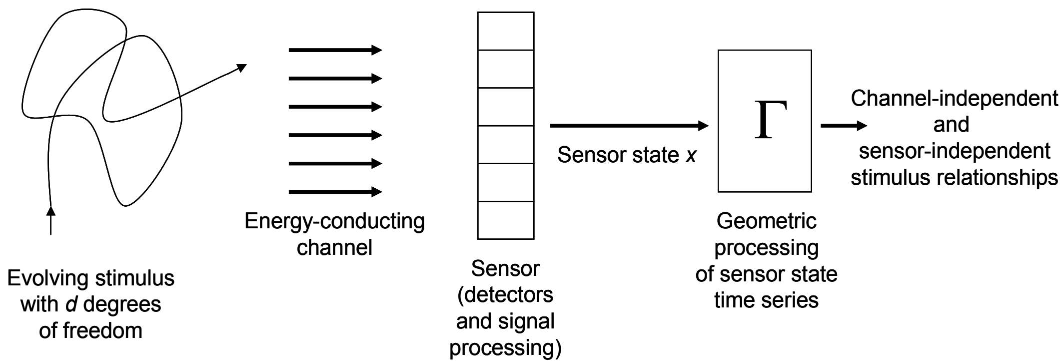

Figure 3. The architecture of an intelligent sensory device. The energy from the stimulus traverses a channel, before being detected and processed by a sensor. The resulting collection of measurements defines a sensor state. The time series of sensor states is processed in order to find channel-independent and sensor-independent relationships among stimulus states.

systematic transformations of those states $x \to x'$.

In this paper, we examine the more well-defined engineering problem of designing sensory machines that represent stimuli in a channel-independent and sensor-independent manner [5]. Consider a sensory device with the general architecture shown in Figure 3. The device is exposed to a "world" that consists of an evolving stimulus with *d* degrees of freedom. Energy from the stimulus traverses a channel and is detected by a sensor, which includes detectors and a processor of the detectors' outputs. This processing, which may be nonlinear, defines a sensor state (*x*) that consists of a collection of measurements and is analogous to the state of the brain's sensory cortices. Finally, the time series of sensor states is processed to make statements about relationships among stimuli. *As in the human case, we want these relationships to be invariant with respect to the nature of the channel and sensor that were used to make observations.* As shown below, this invariance provides a strong constraint on the possible nature of those relationships.

Previously-reported methods of data representation do not have the desired invariance property. For example, in multidimensional scaling [1] and Isomap [15], the representation of stimulus states depends on the assignment of distances between each pair of sensor states, and these distances change if the stimulus measurements are subjected to a nonlinear transformation, caused by a change of channel or sensors. Likewise, in locally-linear embedding [11], stimulus states are represented by coordinates that are not invariant under general nonlinear transformations of the stimulus measurements. And, of course, even less powerful methods of data representation, such as principal components analysis, have the same drawback.

In the next section, we show that the problem of finding channel-independent and sensor-independent stimulus representations can be mapped onto the geometric problem of finding coordinate-system-independent properties of the trajectory of stimulus configurations. In Section III, we explicitly demonstrate the existence of such invariants for a large class of trajectories. The implications and applications of this approach are discussed in the last section.

II. THEORETICAL FRAMEWORK

As in Figure 3, consider some physical system that has *d* degrees of freedom, and suppose that the observing machine's sensor state consists of measurements that are sensitive to some of the degrees of freedom. For the time being, we assume that the sensor state consists of a set of time-independent functions of the stimulus state. This means that the channel and sensor are assumed to be stationary, although this assumption can be weakened if the device is run in an adaptive mode (Section IV). We make two other assumptions about the channel and sensor: 1) the sensor state consists of more than *2d* measurements that are different functions of the stimulus configuration, and 2) the sensor state *x* moves through a *d*-dimensional subspace of the measurement space as the stimulus state moves through the *d*-dimensional stimulus space. We can then make use of the embedding theorem that is well known in the field of nonlinear dynamics [12]. This theorem states that almost every mapping from a d-dimensional space into a space of more than $2d$ dimensions is invertible (i.e., one-to-one). Essentially, this is because so much "room" is provided by the "extra" dimensions of the higher dimensional space that the d-dimensional subspace, which is the range of the mapping, is very unlikely to self-intersect. The embedding theorem virtually guarantees that there is a time-independent one-to-one mapping between the space of all possible stimulus configurations and the sensor states of the above-described machines. Note that, if the sensor states are *not* invertibly related to the stimulus states, the machine must be insensitive to differences between some pairs of stimuli; i.e., some information is lost during the propagation and detection of the signal. In effect, the embedding theorem states that this is exceedingly unlikely to happen as long as the sensor state is comprised of more than $2d$ measurements. The sensor states of a machine that makes fewer than $2d+1$ sensor measurements may or may not be invertibly related to the stimulus states. The embedding theorem simply assures us that the stimulus-sensor mapping is invertible in almost every situation in which the machine's sensor state is comprised of more than $2d$ measurements.

This result can be understood intuitively by considering a stimulus with one degree of freedom, in which each stimulus configuration corresponds to a point on the z axis. Suppose the sensor measurements are comprised of three or more randomly-chosen functions of z, which map the z axis into a higher dimensional measurement space. The embedding theorem asserts that the resulting mapping is very likely to be invertible; i.e., it describes a non-intersecting curve in the higher dimensional space. For example, suppose that the sensor state is comprised of just three measurements, and the measurements are the functions of z defined by executing a short random movement in the three-dimensional measurement space for each short step taken in one direction along the z axis. Such a random walk in 3D space will almost never self-intersect, indicating that the resulting mapping is one-to-one. On the other hand, if the sensor state consists of only one or two measurements, the corresponding random walk in 1D or 2D space is likely to self-intersect, implying that the mapping is unlikely to be invertible.

In a more physical example, the stimulus might consist of a particle that is moving on a two-dimensional surface in a laboratory, the optical channel may consist of lenses and mirrors, the detectors may be three or more cameras in various locations and positions, and the sensor measurements are derived from their images (e.g., their pixel intensities and/or edge intensities and/or Fourier coefficients and/or wavelet coefficients and/or …). The embedding theorem asserts that, for almost all camera positions/orientations and for almost all choices of image-derived measurements, there must be an invertible transformation between the particle's positions on the two-dimensional surface and the corresponding sensor states, which lie in a two-dimensional subspace of the higher-dimensional space of camera measurements. In another example, the stimulus could be utterances produced by a vocal tract with four degrees of freedom, the acoustic channel could be a room, the detectors may be one or more microphones with various gain functions, and the sensor state could be a collection of nine or more acoustic spectral parameters (such as mel frequency cepstral coefficients). Reverberations can be present, as long as they are shorter than the spectral frame length [3], and channel noise can be present, as long as it affects the spectral parameters in a time-independent manner. This is approximately true if the noise is additive and stationary, and the measured spectral parameters average the sound power spectrum over suitably broad filterbank elements [3, 6]. The embedding theorem states that, for almost all microphone positions and for almost all choices of spectral parameters, there will be a one-to-one mapping between the vocal tract's configurations and the locations of the corresponding spectral parameters in a four-dimensional subspace of spectral parameter space.

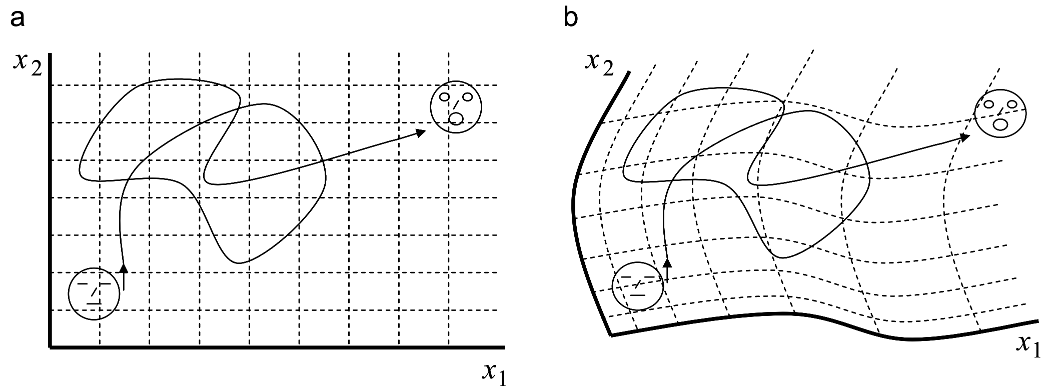

Figure 4. Because of the embedding theorem, devices with different channels and sensors can be considered to impose different coordinate systems (panels *a* and *b*) on the stimulus state space.

Now, in the general case, suppose that a machine uses dimensionally-reduced coordinates, x_k (*k = 1, 2, …, d*) to label each sensor state on the *d*-dimensional manifold of sensor states (for examples of dimensional reduction techniques, see references [11, 15] and references therein). Because the mapping between the stimulus configurations and sensor states is invertible, the machine will assign a different label (a different value of *x*) to each stimulus state, and, therefore, it can be considered to impose the *x* coordinate system on the stimulus space. *Thus, when machines with different channels and sensors observe the same stimulus time series, they are simply using different coordinate systems to view the same trajectory in the stimulus space* (Fig. 4). Therefore, the problem of deriving channel-independent and sensor-independent stimulus relationships is the same as the problem of finding a coordinate-system-independent way of defining stimulus relationships. Presumably, these relationships should be derived from the trajectory of previously-encountered stimuli, which is the only source of information about the stimulus space. Recall that Euclidean geometry provides methods of making statements about a geometric figure that are true in all rotated or translated coordinate systems. Similarly, differential geometry provides the mathematical machinery for making statements about a geometric figure that are true in all non-linearly transformed coordinate systems [13, 16]. In other words, differential geometry seeks to find properties of a geometric figure that are "intrinsic" to it in the sense that they don't depend on extrinsic factors, such as the coordinate system in which the observer chooses to "view" it (i.e., the choice of numerical labels assigned to the figure's points). This suggests the following strategy: 1) we use differential geometry to find coordinate-system-independent properties of the trajectory of previously-encountered stimuli; 2) then, we use those properties to derive coordinate-system-independent perceptual analogies at each point of the traversed part of the stimulus space; 3) these can be used to construct statements about stimulus relationships that are coordinate-system-independent (and, therefore, channel-independent and sensor-independent).

One of the basic notions of differential geometry is parallel transfer, which comprises a coordinate-system-independent way of designating vectors along a path that are said to be equivalent to a given vector at the path's origin (Fig. 5). However, local perceptual analogies also specify a way of choosing equivalent vectors along short path segments. For example, suppose that the line segments $A \rightarrow B$ and $C \rightarrow D$ in Fig. 1 are very short and, therefore, define vectors. Then, $C \rightarrow D$ is the vector at *C* that is perceptually equivalent to the vector $A \rightarrow B$ at *A*, after it has been moved along the infinitesimal line segment $A \rightarrow C$. So, if we can derive a parallel transfer operation on the stimulus space, we can use it as a coordinate-independent way of defining perceptual analogies. There aren't many ways of using the trajectory of previously-encountered stimuli to derive a parallel transfer operation, and some of these possibilities are discussed in the author's previous papers [5, 6, 8]. Here we focus on one of those strategies. First, we use the stimulus trajectory to derive a metric tensor on stimulus space (i.e., a coordinate-system-independent method of defining the lengths of vectors). Then, we derive a parallel transfer mechanism from that metric by using the following fact: there is a unique parallel transfer

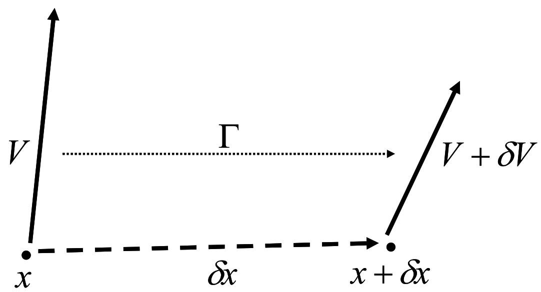

Figure 5. The parallel transfer procedure Γ moves a vector V at x along the short path segment δx to produce an "equivalent" vector $V + \delta V$ at $x + \delta x$.

procedure that preserves the lengths of transferred vectors with respect to a given metric without producing torsion (without producing reentrant geodesics, [13, 16]). In any particular coordinate system, that parallel transfer operation can be computed in the following manner (Fig. 5): given the metric tensor $g_{kl}(x)$, a vector V at x is parallel transported along a line segment δx into the vector $V + \delta V$ at $x + \delta x$, where

$$\delta V^k = - \sum_{l,m=1,\ldots,d} \Gamma^k_{lm}(x) V^l \delta x_m \,, \tag{1}$$

where $\Gamma^k_{lm}(x)$ (the affine connection at x) is given by

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

and where g^{kl} is the inverse of g_{kl}. So, the whole problem of deriving channel-independent and sensor-independent stimulus relationships has been reduced to the task of using the stimulus trajectory to derive a non-singular metric (i.e., a non-singular second rank covariant tensor).

Let $x(t)$ describe the evolving stimulus state in the coordinate system used by a particular observing machine. At each point y, consider the local covariance matrix of the trajectory's "velocity"

$$c^{kl}(y) = \langle \dot{x}_k \dot{x}_l \rangle_{x(t) \sim y} \tag{3}$$

where the bracket denotes the time average over the trajectory's segments in a small neighborhood of y. If this quantity approaches a definite limit as the neighborhood shrinks around y, it will certainly transform as a contravariant tensor. Furthermore, as long as the trajectory has d linearly-independent segments passing through y, this tensor can be inverted to form a covariant tensor, $g_{kl}(y) = (c^{-1})_{kl}$. So, Eq. 3 can be used to derive a metric from most trajectories, as long as the above-described limit exists.

III. EXAMPLES

There is a large class of trajectories that can be proven to induce metrics on the stimulus manifold. We demonstrate this by explicitly constructing such trajectories from physical systems that can be realized in the laboratory. First, consider a system with d degrees of freedom given by x_k ($k = 1, 2, \ldots, d$) in some coordinate system, and suppose that the system's energy is given by

$$E(\dot{x}, x) = \frac{1}{2} \sum_{k,l=1,\ldots,d} \mu_{kl}(x) \dot{x}_k \dot{x}_l + V(x) \tag{4}$$

for some functions μ_{kl} and V, where $\dot{x} = \frac{\partial x}{\partial t}$. This is the energy function of a large variety of physical systems in which the degrees of freedom are not necessarily spatial and in which d may be large. However, the simplest system with this energy is a single particle with unit mass that is moving in potential $V(x)$ on a possibly-curved two-dimensional frictionless surface with metric μ_{kl}. For example, the particle may be moving on a frictionless spherical, cylindrical, or planar surface in a laboratory. Returning to the more general case, consider a large non-interacting collection of such systems that are in thermal equilibrium with some external "bath" at temperature T. According to statistical mechanics [10], the probability of having a system with "velocity" $\dot{x}$ at point x is proportional to the Boltzmann distribution

$$P(\dot{x}, x) \propto \exp\left[-E(\dot{x}, x)/kT \right] \tag{5}$$

where k is the Boltzmann constant. For the whole collection of systems, it follows that the average velocity dyadic at y is

$$\langle \dot{x}_k \dot{x}_l \rangle_{x=y} = kT\mu^{kl}(y) \tag{6}$$

where μ^{kl} is a contravariant tensor equal to the inverse of μ_{kl}. This result can be confirmed by explicitly doing the Gaussian integrals that result from using Eqs. 4-5 to compute the left side of Eq. 6.

The desired stimulus trajectory can now be constructed by randomly selecting systems from this Boltzmann distribution and time translating the chosen systems so that they are traversed during sequential time segments, mimicking the time course of a single system that intermittently changes its energy. For this composite trajectory, the right side of Eq. 3 has a well-defined local limit, and it is equal to the velocity dyadic locally averaged over the ensemble of component trajectories (i.e., Eq. 6). Therefore, in this case, the right side of Eq. 3 is $kT\mu^{kl}(y)$, and its inverse transforms as a covariant tensor, as desired. Thus, for this large class of trajectories, a valid metric is induced on the stimulus state space, and Eqs. 1-2 define a coordinate-system-independent method of moving vectors across the stimulus manifold. It is likely that metrics can be similarly derived from many other mathematical classes of trajectories, especially those with sufficiently smooth spatial variation of their local velocity distributions. However, it is an experimental question whether this is the case for the trajectories of specific physical stimuli. Speech data were studied in reference [6], and the experimental results suggest that a metric can be derived from the stimulus trajectories produced by spoken English.

The above-described process is illustrated by considering a specific example: the stimulus trajectory constructed by concatenating the trajectories of a Boltzmann distribution of particles moving in some potential V on a flat two-dimensional surface (d=2). Suppose that two machines with different channels and/or sensors are each making more than four sensor measurements of the particle's motion. Let $x(t)$ be the dimensionally-reduced trajectory recorded by the first machine, and $x'(t)$ be the dimensionally-reduced trajectory recorded by the other machine. As mentioned above, $x(t)$ and $x'(t)$ describe the particle's position in two different coordinate systems. To make things simple, suppose that the x coordinate system happens to be a Cartesian coordinate system on the surface, so that $\mu_{kl} = \delta_{kl}$ where δ_{kl} is the Kronecker delta symbol. The particle coordinates recorded by the second machine are given by some invertible function $x' = x'(x)$ of the coordinates recorded by the first machine. In effect, the first machine is recording the particle trajectory that would be seen through a normal undistorted optical channel, while the second machine is recording the particle trajectory that would be observed through a distorted optical channel (e.g., a channel that non-linearly warps the image). As described above, the first machine derives the metric $g_{kl} = \mu_{kl} / kT = \delta_{kl} / kT$ from the trajectory of previously-encountered stimuli, so that Eq. 2 implies that the affine connection vanishes $\Gamma^k_{lm} = 0$. Therefore, according to Eq. 1 and Fig. 5, this machine equates numerically-equal vectors at different points in stimulus space, just as in ordinary Euclidean geometry in a Cartesian coordinate system. The second machine uses its previously-recorded observations to derive the metric $g'_{kl} = \mu'_{kl} / kT = \frac{1}{kT} \sum_{n=1,2} \frac{\partial x_n}{\partial x'_k} \frac{\partial x_n}{\partial x'_l}$. Substitution into Eq. 2, followed by algebraic manipulation, implies that the second machine's affine connection is $\Gamma'^k_{lm} = \sum_{n=1,2} \frac{\partial x'_k}{\partial x_n} \frac{\partial^2 x_n}{\partial x'_l \partial x'_m}$. Substituting this into Eq. 1 shows that this affine connection parallel transfers the vector $\sum_{n=1,2} \left. \frac{\partial x'_k}{\partial x_n} \right|_{x'} V^n$ at x' along the short path $\delta x'$ into the vector $\sum_{n=1,2} \left. \frac{\partial x'_k}{\partial x_n} \right|_{x'+\delta x'} V^n$ at $x'+\delta x'$, where V^n is any doublet of numbers. In other words, the second machine equates vectors in the x' coordinate system that are numerically equal to one another, once they have been transformed into the x coordinate system. This shows that the parallel transfer operations of the two machines are exactly the same, although they are implemented in different coordinate systems. Therefore, if one machine perceives the equivalence of vectors defined by two small transformations of nearby stimuli (e.g., in Fig. 1, if $A \to B$ is perceived to be the same as $C \to D$), the other machine will independently come to the same conclusion. In general, the two machines will agree about any sequences of parallel transferred line segments that are required to construct a path from one stimulus to another, and, in this sense, they will represent the "world" in the same way.

IV. DISCUSSION

We have shown how a machine can represent stimuli in a sensor-independent and channel-independent manner. In other words, a machine observing stimuli with a specific channel and specific sensors can glean channel-independent and sensor-independent information about the relationships of stimuli to one another. The nature of those relationships reflects the geometry that is induced on the stimulus manifold by the history of previously-encountered stimuli. For example, the relationships among stimuli in an intrinsically curved manifold will be strikingly different than the relationships among stimuli in a flat space [4]. In other words, like humans, a machine's representation of the world depends on the statistical properties of its past experience. Notice that two machines will represent stimulus relationships in the same way, as long as they have encountered trajectories that impose the same geometry on the stimulus space (i.e., geometry derived from velocity covariance matrices that are identical when expressed in the same stimulus coordinate system). Two such machines will agree on the "perceived" relationships among stimuli, even if their previously-encountered stimulus trajectories differ in other respects. On the other hand, if the stimulus trajectories have different velocity covariance matrices, the two machines will "perceive" different relationships between some stimuli.

It is interesting to consider the kinds of "memories" that one of these devices could have. In principle, if the device is exposed to a sufficiently dense collection of stimulus states, it can learn the shape and location of the *d*-dimensional manifold of sensor states within the larger measurement space. Furthermore, it could record the affine connection at each point on that manifold. For example, it could adjust the weights of a neural network so that it comprises a kind of "connection engine" for mapping each point *x* onto $\Gamma^k_{lm}(x)$. If the machine subsequently re-experiences a sensor state *E* on the manifold (e.g., state *E* in Fig. 2), it can use the affine connection to deduce its relationship to other nearby sensor states (e.g., states *A, B, C,* and *D* in Fig. 2). In this way, when a machine re-experiences a stimulus, it can also recall the web of its familiar relationships to other stimuli. In other words, stimulus relationships can be recomputed at any time, as long as the same affine connection is in memory. On the other hand, suppose that the parallel transfer mechanism has changed since the stimulus *E* was first encountered, because the machine has experienced a significantly different stimulus trajectory in the intervening time. Then, the machine may derive relationships between *E* and other stimulus states that differ from the ones it derived before. Likewise, the trajectory segment of a recalled train of events will be described differently with the new affine connection than it was with the affine connection that existed at the time those events first transpired. The "good news" is that both versions of events are coordinate-system-independent and, therefore, they are both independent of factors extrinsic to the stimulus trajectory, such as the nature of the machine's channels and sensors. The "bad news" is that this information is influenced by the dynamical history of previously-encountered stimuli and does not reflect "absolute" properties of the stimuli themselves. The relative nature of the gleaned information can be illustrated in the following way. Consider a machine that has been exposed to a "world" consisting of certain physical stimuli (e.g., a moving particle) and has used the trajectory of these stimuli to learn coordinate-system-independent relationships among them, as described in this paper. Now, consider a second machine, remotely-located in another room or another galaxy, that has been exposed to a different "world" of stimuli (i.e., stimuli from a different physical system, e.g., sounds produced by a vibrating string). If there is a one-to-one mapping between the possible configurations of the two physical systems that maps the first stimulus trajectory into the second one, the first machine will derive parallel transfer relationships among stimulus states (e.g., Fig. 2) that are the same as the relationships that the second machine derives among the corresponding states of its stimulus system. Therefore, the two machines will represent their physically different "worlds" in exactly the same manner. Even if they communicate the observed stimulus relationships to one another, they will not discern any differences between their "worlds", because the communicated relationships are relative, not absolute, properties of experienced stimuli. These devices will only discover the differences between their stimulus systems if they arrange to observe the same physical system.

Up to now, we have assumed that a machine's channel and sensors are time-independent and that each machine derives a parallel transfer mechanism from the history of *all* previously-encountered stimuli. However, a machine could adapt to channels and/or sensors that drift over a time scale longer than ΔT , if its affine connection was derived from stimuli encountered in the most recent time interval ΔT . It would continue to represent stimulus relationships in the same way, as long as the stimulus trajectory's velocity covariance matrices were temporally stable (in any fixed coordinate system). Notice that the trajectories constructed in Section III have this kind of temporal stability. This process of adaptation was demonstrated with simple experimental examples in reference [7].

The nonlinear signal processing method presented in this paper could be used as a representation "engine" in the "front end" of intelligent sensory devices [9]. It would produce channel-independent and sensor-independent stimulus representations that are passed to the device's pattern recognition 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. For example, suppose that a machine has been trained under one set of conditions to recognize *E* by the relationship to *A*, *B*, *C*, and *D* shown in Fig. 2. If its channel or sensors changed, after a period of adaptation, it could continue to use parallel transfer to recognize *E* in this way. References [6] and [7] describe experiments in which this method was used to normalize speech signals with respect to different channels, with the ultimate goal of achieving channel-independent automatic speech recognition. The same general approach may lead to detector-independent computer vision devices, and, because the method does not explicitly depend on the nature of the detectors, it is a natural way to achieve multimodal fusion of audio and optical sensors. Finally, it is worth mentioning the possibility of creating a speech-like telecommunications system that is resistant to channel-induced corruption of the transmitted information [6]. In such a system, information is carried by the coordinate-system-independent relationships among the transmitted signals and among the received signals. As an example, imagine that the transmitter broadcasts a power spectrum that is an invertible function of two time-dependent variables, and suppose that coordinate-system-independent relationships among the points on the variables' trajectory constitute the information to be communicated. Suppose that the signal propagates through a linear channel and is detected by a receiver, which computes the outputs of a filterbank with more than four broad elements. If the channel noise is stationary and reverberations are shorter than the frame length, the embedding theorem almost guarantees that the filterbank output trajectory is invertibly related to the two transmitter variables. Mathematically, this means that the transmitter and receiver are recording the same trajectory in two different coordinate systems. Once the receiver computes a metric from the output time series of the filterbank, it can compute the coordinate-system-independent relationships among points on that trajectory. These must be the same as the coordinate-system-independent relationships among the points on the transmitter variables' trajectory, which is the desired information.

The remarkable constancy and universality of human perception have been the subject of philosophical discussion since the time of Plato (e.g., the allegory of "The Cave"), and these issues have also intrigued modern neuroscientists. How can an individual perceive the intrinsic constancy of a stimulus even though its appearance is changing because of extraneous factors? Why do different people tend to have similar perceptions despite significant differences in their sensory organs and neural processing pathways? This paper shows how to design sensory devices that invariantly represent stimuli in the presence of processes that transform their sensor states. These stimulus representations are invariant because they encode "inner" properties of the time series of stimulus configurations themselves; i.e., properties that are largely independent of the nature of the observing device or the conditions of observation. Significant evolutionary advantages would accrue to organisms that developed the ability to process sensory information in this way. For instance, they would not be confused by the appearance-changing effects of different observational conditions, and communication among different organisms would be facilitated by their tendency to independently represent stimuli in the same way. Biological experimentation is required to determine whether humans and other species achieve these objectives by means of the general approach in this paper. However, it would not be surprising if they did, because there are not many other ways of representing the world in an invariant fashion.

References

*Address correspondence to:
David N. Levin
1720 N. Lasalle Dr., #25
Chicago, IL 60614
Email: d-levin@uchicago.edu
Tel: 847-491-6697
http://www.radiology.uchicago.edu/levin.htm

[1] T. Cox and M. Cox, *Multidimensional Scaling*. London: Chapman & Hall, 1994.
[2] R. Held and R. Whitman, *Perception: Mechanisms and Models*. San Francisco, CA: Freeman, 1972.
[3] X. Huang, A. Acero, and H-W Hon, *Spoken Language Processing*. Upper Saddle River, NJ: Prentice Hall PTR, 2001, p. 483.
[4] D. N. Levin, "A differential geometric description of the relationships among perceptions", *J. Math. Psych.*, vol. 44, pp. 241-284, 2000.
[5] D. N. Levin, "Stimulus representations that are invariant under invertible transformations of sensor data", *Proc. SPIE*, vol. 4322, pp. 1677-1688, 2001.
[6] D. N. Levin, "Blind normalization of speech from different channels", *J. Acoust. Soc. Am.*, vol. 116, pp. 519-529, 2004.
[7] D. N. Levin, "Representations of sound that are insensitive to spectral filtering and parameterization procedures", *J. Acoust. Soc. Am.*, vol 111, pp. 2257-2271, 2002.
[8] D. N. Levin, "Sensor-independent stimulus representations", *Proc. Nat. Acad. Sci. (USA)*, vol. 99, pp. 7346-7351, 2002.
[9] D. N. Levin, US Patent # 6,687,657, 2004 and pending patents.
[10] F. Reif, Fundamentals of Statistical and Thermal Physics. New York: McGraw-Hill, 1965.
[11] S. T. Roweis and L. K. Saul, Nonlinear dimensionality reduction by locally linear embedding, *Science*, vol. 290, pp. 2323-2326, 2000.
[12] T. Sauer, J. A. Yorke, and M. Casdagli, "Embedology", *J. Stat. Phys.*, vol. 65, pp. 579-616, 1991.
[13] E. Schrodinger, *Space-Time Structure*. Cambridge, UK: Cambridge U. Press, 1963.
[14] G. M. Stratton, "Vision without inversion of the retinal image (concluded)", *Psychol. Rev.*, vol. 4, pp. 463-481, 1897.
[15] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction", *Science*, vol. 290, pp. 2319-2323, 2000.
[16] S. Weinberg, *Gravitation and Cosmology*. New York: John Wiley & Sons, 1972.