Channel-Independent and Sensor-Independent Stimulus Representations

David N. Levin
University of Chicago (d-levin@uchicago.edu)

Abstract

This paper shows how to design intelligent sensory devices that represent stimulus relationships in a channel-independent and sensor-independent manner. First, we demonstrate that a device's channel and sensor define a coordinate system that the device imposes on the space of stimulus states. Then, differential geometry is applied to derive coordinate-system-independent stimulus relationships from statistical characteristics of the time series of previously-encountered stimuli. This procedure is illustrated with analytic examples. In an intelligent sensory device, this kind of representation "engine" could function as a "front end" that passes channel/sensor-independent stimulus representations to a pattern recognition module. After a pattern recognizer has been trained in one of these devices, it could be used without change in other devices having different channels and sensors.

1. Introduction

A method of designing channel-independent and sensor-independent sensory devices is suggested by the following discussion of human perception. 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 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, B, C, and D differ by small stimulus transformations. 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 [3]. Figure 2 shows an example in which an observer describes stimulus E as being related to stimuli A, B, and C by a succession of analogous stimulus transformations: "E is the stimulus that is produced by starting with stimulus A and performing three transformations perceptually equivalent to the $A \rightarrow C$ transformation, followed by two transformations perceptually equivalent to the $A \rightarrow B$ transformation". In principle, an observer, who perceives local stimulus analogies at each point in stimulus space, can use this technique to describe the paths connecting any two stimuli of interest. This knowledge of relative stimulus locations can be used to navigate through stimulus space, and it comprises a kind of representation of the stimulus space (a kind of "map of the world") in the same sense that a geographic

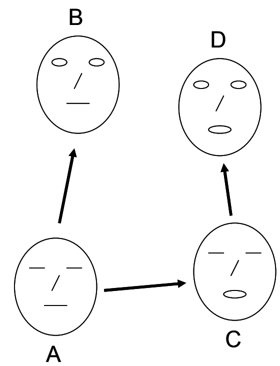

Figure 1. Consider a two dimensional space populated by faces characterized by the extent to which the mouth and eyes are open. Given two faces (*A* and *B*) that are related by a small transformation ($A \rightarrow B$, called an "eye-opening" transformation) and given another face (*C*) that differs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

representation of a city is provided by the knowledge of how to navigate between any two points in it.

If the brain uses stimulus analogies to impose order on the "world" of stimuli, it seems likely that these analogies are derived from past sensory history, rather than being "hard-wired" from birth. For instance, there is evidence that young children "learn to see" and adults with no previous visual experience usually do not sense meaningful relationships between visual stimuli. The latter phenomenon is illustrated by many neurological case histories in which congenital cataracts were excised from the eyes of a blind adult with the intention of allowing the patient to see for the first time. In each case, when the post-operative bandages were removed, the patient invariably experienced an inchoate mixture of meaningless visual patterns, presumably because he/she never learned how to impose order on visual information. Normal individuals with similar past experiences learn to perceive the world in approximately the same way, and this is remarkable, given the fact that different individuals have different "sensors" (e.g., eyes, ears, and primary sensory cortices). To illustrate, suppose that stimuli *A*, *B*, *C*, and *D* create complex spatial distributions of neuronal electrical discharges (denoted x_A, x_B, x_C, and x_D) in the primary sensory cortex of observer Ob. The same stimuli will produce different three-dimensional electrical current distributions (x'_A, x'_B, x'_C, and x'_D) in the sensory cortex of a second observer Ob', who has a different sensory system Yet, despite these differences, if the two observers have had similar life histories and Ob senses $x_A : x_B = x_C : x_D$, then Ob' tends to independently conclude $x'_A : x'_B = x'_C : x'_D$. A related fact is that each individual's perceptions of stimuli tend to be independent of the channel through which he/she views those stimuli. This fact was strikingly illustrated by experiments in which subjects wore goggles that created severe geometric distortions and inversions of the observed scenes [2]. 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. These observations suggest the following generalization. Consider an individual who has a certain history of stimulus exposure and "learns" to perceive relationships (e.g., $x_A : x_B = x_C : x_D$) among certain stimulus-induced sensory cortex states. If that person is exposed to a similar history of stimuli through a new observational channel, he/she will relearn relationships among sensory cortex states with the result that the altered sensory cortex states induced by stimuli through the new channel will be perceived to be related to one another in the same way (e.g., $x'_A : x'_B = x'_C : x'_D$) as the corresponding unaltered cortex states were related before the channel changed. *In summary: the phenomena discussed in this Section suggest that the brain uses its past experience to derive stimulus relationships (e.g., stimulus analogies and relative stimulus locations) that are invariant in the presence of channel-induced and sensor-induced transformations of sensory cortex states.*

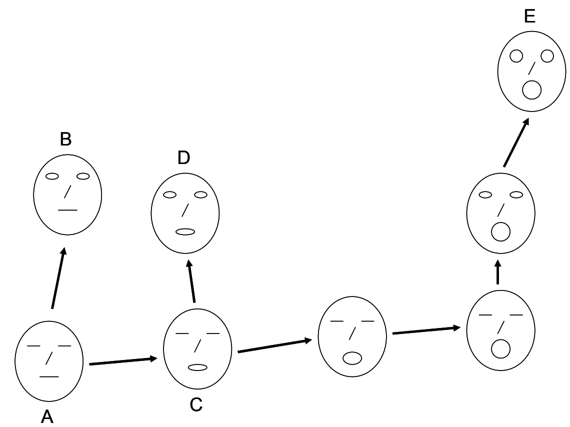

Figure 2. "Distant" stimuli can be related by describing a path connecting them as a sequence of perceptually defined stimulus transformations. For example, an observer may describe *E* as the stimulus that is produced by starting at *A* and performing three transformations perceptually equivalent to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

In this paper, we propose how to design sensory machines that also represent relative stimulus locations in a channel-independent and sensor-independent manner [4]. Consider a sensory device having the general architecture shown in Figure 3, and suppose that the device is exposed to a "world" consisting of an evolving stimulus. Energy from the stimulus traverses a channel, before being detected and processed by a sensor. This processing defines a sensor state (*x*), which is a function of the stimulus state and is analogous to the state of the brain's primary sensory cortex. In a conventional sensory device, the sensor state would be sent directly to a pattern recognition module, and, if the channel or sensor changes, the device must be taken "off-line" for retraining of the recognizer or recalibration of the detector. In the devices proposed in this paper, the sensor state time series is first processed in order to learn how to represent each stimulus state by its channel-independent and sensor-independent location relative to previously-encountered stimulus states. Because

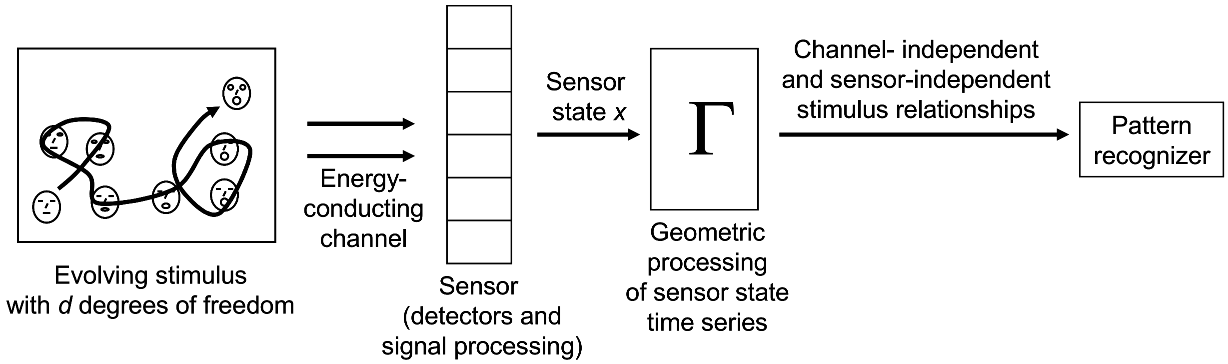

Figure 3. The architecture of an intelligent sensory device. The energy from the stimulus traverses a channel, before being detected and processed by a sensor. The resulting collection of measurements is used to define a sensor state. The time series of sensor states is processed in order to find channel-independent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

pattern recognition is applied to this information, it is not channel-dependent or sensor-dependent.

Previously-reported methods of data representation do not have the desired channel-independence and sensor-independence. For example, in multidimensional scaling [1] and Isomap [12], 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 [10], stimulus states are represented by coordinates that are not invariant under all of the stimulus measurement transformations, which are induced by channel and sensor changes. And 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 relative stimulus locations can be mapped onto the differential geometric problem of finding coordinate-system-independent properties of the trajectory of previously-encountered stimulus configurations. In Section 3, we describe large classes of stimulus trajectories for which the desired invariant stimulus representations can be analytically derived. The implications and applications of this approach are discussed in the last section.

2. Theory

Suppose that the stimulus is some physical system that has d degrees of freedom, and suppose that the observing machine's sensor makes at least $2d+1$ measurements, which are time-independent functions of the stimulus configuration. 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 4). The measurement functions define a mapping from the d-dimensional stimulus space to a d-dimensional subspace of the higher-dimensional ($\geq 2d+1$) measurement space. The embedding theorem, which is well-known in the field of nonlinear dynamics [11], states that this mapping is invertible for almost all choices of measurement functions. Essentially, invertibility is likely 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. Suppose that the stimulus evolves so that the stimulus trajectory densely covers a patch of stimulus space and the corresponding sensor measurements densely cover a patch of the d-dimensional measurement subspace. Then, the machine can use dimensional reduction methods [10, 12] to learn the location and shape of the measurement subspace, and it can impose an arbitrarily-chosen x (x_k , $k = 1, 2, \ldots, d$) coordinate system on it. The quantity x is defined to be the sensor state of the machine, and, according to the embedding theorem, it is invertibly related to the stimulus state. For example, suppose that the stimulus consists of a particle that is moving on a lumpy transparent two-dimensional surface in a laboratory ($d = 2$). Furthermore, suppose that its state is being monitored by three video cameras in various locations and positions, each of which derives two measurements from the imaged scene (e.g., the two pixel coordinates of the particle or two Fourier coefficients of the image or two wavelet coefficients of the image or two …). Notice that the mapping between the particle's position and the measurements of any one camera may not be invertible because of the lumpiness of the surface. As the particle moves over the surface in the laboratory, the six camera measurements move in a two-dimensional subspace within the six-dimensional

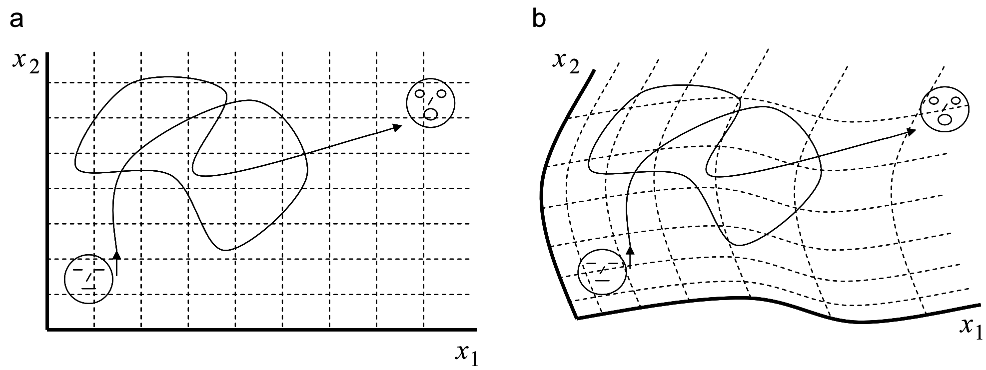

Figure 4. Because of the embedding theorem, devices with different channels and/or sensors can be considered to be viewing the same stimulus trajectory in different coordinate systems of the stimulus state space (panels *a* and *b*).

space of possible measurements, and the location and shape of this subspace can be learned by dimensional reduction of a sufficiently dense set of measurements. Each stimulus configuration (each position of the particle in the laboratory) produces a sensor state x, which is defined to be the location of the corresponding measurements in some two-dimensional coordinate system on the measurement subspace. The embedding theorem asserts that the mapping between the stimulus and sensor states is one-to-one, for almost all choices of camera positions and orientations and for almost all types of image-derived measurements.

In general, our objective is to represent the locations of stimulus states relative to one another in a channel-independent and sensor-independent manner. The discussion in Section 1 suggests that the machine should learn these relationships from its "life experience", i.e., from the sequence of previously-observed stimuli. In any event, the trajectory of previously-encountered stimuli provides the only possible source of such "structure" on the stimulus space. Because the mapping between the stimulus and sensor states is invertible, the sensor state can be considered to be the location of the stimulus in a particular (x) coordinate system on the stimulus space. Therefore, 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). Hence, the problem of defining channel-independent and sensor-independent relative stimulus locations is the same as the problem of finding coordinate-system-independent stimulus locations. 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 (such as the stimulus trajectory) that are true in all non-linearly transformed coordinate systems [13]. 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) first, use differential geometry to find coordinate-system-independent properties of the trajectory of previously-encountered stimuli; 2) then, use those properties to derive coordinate-system-independent stimulus analogies at each point of the traversed part of the stimulus space; 3) finally, use these analogies to derive relative stimulus locations (e.g., Fig. 2) that are coordinate-system-independent (and, therefore, channel-independent and sensor-independent).

From a geometric point of view, a local stimulus analogy specifies a coordinate-system-independent way of moving one short line segment along a second short line segment in order to create a third short line segment that is said to be "equivalent" to the first one. For example, in Fig. 1, $C \to D$ is the line segment at C that is perceptually equivalent to the line segment $A \to B$ at A, after it has been moved along the line segment $A \to C$. This is exactly the operation that is performed by the parallel transfer procedure of differential geometry, 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). Therefore, if a parallel transfer operation can be derived from the stimulus trajectory, it can be used to define stimulus analogies, and these can be concatenated to define relative stimulus locations in a coordinate-system-independent manner. There are a few methods of using the stimulus trajectory to define a parallel transfer operation [4, 5], and in this paper, we focus on one of them. First, we use the stimulus trajectory to derive a metric on stimulus space (i.e., a coordinate-system-independent method of defining the lengths of line segments and other vectors). Then, we derive a parallel transfer mechanism from that metric by using the following fact: there is a unique parallel transfer procedure that preserves the lengths of transferred vectors with respect to a given metric without producing torsion (without producing reentrant geodesics, [13]). In any coordinate system (x), this particular 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 transferred along a line segment δx into the vector $V+\delta V$ at $x+\delta x$, where

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

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

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

and where g^{kl} is the inverse of g_{kl}. Therefore, 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 symmetric covariant tensor). Once that metric has been found, Eqs. 1-2 can be used to define a parallel transfer operation, which makes it possible to define stimulus analogies and relative stimulus locations in a coordinate-system-independent manner (and, therefore, in a channel-independent and sensor-independent manner).

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

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

where $\dot{x} = \frac{dx}{dt}$ and 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 symmetric contravariant tensor. Furthermore, as long as the trajectory has d linearly-independent segments passing near y, this tensor can be inverted to form a symmetric covariant tensor, $g_{kl}(y) = (c^{-1})_{kl}$. Thus, Eq. 3 can be used to derive a metric from any stimulus trajectory for which the above-described limit exists.

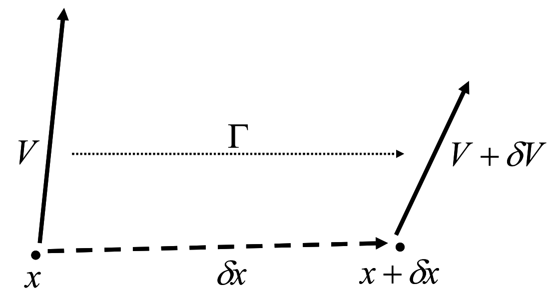

Figure 5. The parallel transfer procedure Γ moves a vector *V* at *x* along the short path segment δx to produce an "equivalent" vector $V+\delta V$ at $x+\delta x$.

3. Examples

In this Section, we demonstrate large classes of stimulus trajectories for which the local velocity covariance matrix (and metric) are well-defined and can be computed analytically. Specifically, we construct these trajectories from the behavior of physical systems that can be realized in the laboratory. First, consider a system with d degrees of freedom, and suppose that the system's energy is given by

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

where μ_{kl} and V are some functions and x is the stimulus' location in some coordinate system. Equation 4 is the energy function of a large variety of physical systems in which the degrees of freedom can be spatial and/or internal and in which d may be large or small. 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 physical metric μ_{kl}. For example, if the particle is moving on a frictionless spherical, cylindrical, or planar surface in the laboratory, μ_{kl} is the metric induced on the surface by the Euclidean metric in the laboratory's coordinate system. Returning to the more general case, suppose that the system intermittently exchanges energy with a thermal "bath" at temperature T. In other words, suppose that the system evolves along one trajectory from the Boltzmann distribution at that temperature and periodically jumps to another randomly-chosen trajectory from that distribution. After a sufficient number of jumps, the amount of time the particle will have spent in a small neighborhood dx of x and a small neighborhood $d\dot{x}$ of $\dot{x}$ is proportional to the Boltzmann distribution [9]

$$\mu(x)\exp\left[-E(x,\dot{x})/kT\right]dx\,d\dot{x} \qquad (5)$$

where k is the Boltzmann constant and μ is the determinant of μ_{kl}. If this expression is substituted into the right side of Eq. 3 and the resulting Gaussian integrals are evaluated, the velocity covariance matrix is found to be well-defined and given by

$$c^{kl}(x) = kT\mu^{kl}(x) \qquad (6)$$

where μ^{kl} is a contravariant tensor equal to the inverse of μ_{kl}. It follows that the trajectory-induced metric on the stimulus space is $g_{kl}(x) = \mu_{kl}(x)/kT$; i.e., it is proportional to the physical metric on the surface. Thus, a metric is induced on the stimulus state space by this large class of systems and trajectories. It follows that Eqs. 1-2 define a coordinate-system-independent method of moving line segments across the stimulus manifold, and this procedure can be used to define channel-independent and sensor-independent stimulus analogies and relative stimulus locations. As an illustration, reconsider the case (Section 2) of the particle that is moving on a surface and is being observed by a machine equipped with multiple cameras. Two observing machines of this type will agree on all statements about relative particle locations (i.e., all statements about parallel transfer operations, like those in Fig. 2), even though their sensors and channels may be dramatically different (e.g., different numbers of cameras, different camera positions and orientations, camera lenses producing different distortions of the observed scene, different image-derived measurements such as particle pixel coordinates vs. image Fourier coefficients vs. …).

As another example, consider a car that is being driven through a city, which is laid out on a flat surface, and suppose that the car is equipped with a video camera that is always pointing west. Then, the stimulus space is two-dimensional ($d = 2$), being comprised of the collection of western views associated with all locations in the city. Suppose that the camera's output consists of five measurements that are derived from each image encountered by the car along its route. The time series of these measurements sweeps out a trajectory within a two-dimensional subspace of the five-dimensional space of possible measurements. If the car's path through the city is sufficiently dense, the associated measurement trajectory will densely cover this two-dimensional subspace, and the machine can use dimensional reduction methods in order to learn the subspace's location and shape. The machine's sensor state is defined to be the measurement's location in some two-dimensional coordinate system on this subspace. Because the stimulus and sensor states are invertibly related according to the embedding theorem, the sensor state can be taken to define the location of the stimulus in the machine-defined coordinate system on the stimulus space. Now, suppose that the car is being driven at a constant speed s along randomly-chosen directions in the Euclidean coordinate system of the city. Then, in this coordinate system, the time average in Eq. 3 is well-defined and equal to $c^{kl}(x) = \frac{s^2}{2}\delta_{kl}$ where δ_{kl} is the Kronecker delta, and the trajectory-derived metric on the stimulus space is proportional to the Euclidean metric, namely $g_{kl}(x) = \frac{2}{s^2}\delta_{kl}$. In any other coordinate system (e.g., in the coordinate system defined by the machine's sensor state), the trajectory-derived metric will be similarly proportional to the form of the Euclidean metric in that coordinate system. This implies that two different machines will derive the same metric (each in its machine-defined coordinate system). Therefore, they will agree on all statements about relative stimulus locations (i.e., all statements about the geography of the city), despite the fact that they may be equipped with different cameras (e.g., cameras with different lenses or "goggles"), they may use different image-derived measurements, etc. It is interesting to note that, if the car drives more slowly in one part of the city than in others, the trajectory-derived metric will be larger there, and the machine will "perceive" that part of the city to be relatively larger.

In this Section, we demonstrated large classes of trajectories that endowed the stimulus manifold with a metric (and, consequently, channel-independent and sensor-independent stimulus relationships). It is likely that metrics can be derived from many other trajectories, especially those with local velocity distributions that vary sufficiently smoothly across stimulus space. However, it is an experimental question whether this is the case for the trajectories of specific physical stimuli. For example, it has been shown that a metric can be derived from the stimulus trajectories produced by spoken English. In these experiments, the metric was used to derive a channel-independent representation of speech trajectories, with the ultimate goal of achieving channel-independent automatic speech recognition. A complete description of these experiments is available in reference [5].

4. Discussion

We have shown how a machine, which observes stimuli through a specific channel and with specific sensors, can glean channel-independent and sensor-independent information about the locations of stimuli relative to one another. This is possible if two requirements are satisfied: 1) the observed stimulus trajectory has well-defined local velocity covariance matrices; 2) the device measures more than *2d* properties of the stimulus, where *d* is the number of stimulus degrees of freedom. Like a human, such a machine's representation of the world depends on the statistical properties of its past sensory "experience". Two such machines, using different channels and sensors, will represent stimulus relationships in the same way, as long as their past trajectories through stimulus space impose the same geometry on that space, and this will be true as long as the velocity covariance matrices of their trajectories are identical when expressed in the same stimulus coordinate system. On the other hand, if their stimulus trajectories have different velocity covariance matrices, the two machines will not agree on all statements about relative stimulus locations. For example, there will be striking disagreements if the trajectory-derived metric of one machine is flat and the trajectory-derived metric of the other one is curved [3]. The methodology in this paper should not be confused with conventional blind source estimation techniques. The latter attempt to estimate the form of the stimulus trajectory when it is measured in a specific manner (e.g., in a specific coordinate system). In this paper, we seek to determine properties of the stimulus trajectory that are independent of the way it is measured (i.e., coordinate-system-independent).

It is interesting to consider the kinds of "memories" that one of these devices could have. A neural network or parametric method could be used to store the location and shape of the *d*-dimensional measurement subspace, once the machine has learned that information by observing a sufficiently dense stimulus trajectory. This subspace contains all sensor state measurements encountered in the "world". It could also store the affine connection, which encodes the relative locations of sensor states in that subspace. For example, the weights of a neural network could be adjusted so that it comprises a kind of "connection engine" for mapping each sensor state x onto $\Gamma^{k}_{lm}(x)$. If the machine subsequently re-experiences a sensor state (e.g., the sensor state corresponding to stimulus *E* in Fig. 2), it can use the stored affine connection to deduce it's location relative to other nearby sensor states in the measurement subspace (e.g., *A*, *B*, and *C* in Fig. 2). Thus, previously-experienced stimulus relationships can be recomputed at any time, as long as the machine's memory contains the previously-experienced affine connection. On the other hand, suppose that the parallel transfer mechanism has changed since a stimulus was first encountered, because the machine has experienced a statistically different stimulus trajectory in the intervening time. Then, the machine will describe the relative location of a stimulus differently than it did before. Likewise, the trajectory segment of a recalled train of events may be described differently with the new affine connection than it was with the affine connection that existed at the time those events first transpired. Despite these differences, each description of the events is valid at the time it was derived, in the sense that it is coordinate-system-independent (and, therefore, independent of factors extrinsic to the stimulus trajectory, such as the nature of the machine's channels and sensors).

. It is philosophically interesting that the channel-independent and sensor-independent information gleaned by these machines reflects the intrinsic properties of the observed *trajectory* of the stimulus, not the absolute or "real" properties of the physical stimulus itself. As an illustration, consider a machine that has been exposed exclusively to a "world" consisting of certain physical stimuli (e.g., a moving particle on a two-dimensional surface) and has used the trajectory of these stimuli to learn their coordinate-system-independent relative locations. Now, consider a second machine, remotely-located in another room or another galaxy, that has been exposed exclusively to different "world" of physical stimuli (e.g., a time series of changing faces, like those in Fig. 4). If the two stimulus spaces are related by a one-to-one mapping that maps the first stimulus trajectory onto the second one, the first machine will describe the relative locations of its stimuli exactly as the second machine describes the relative locations of the corresponding states of its stimulus system. Suppose that each machine has not been exposed to any other stimuli and that it only has access to information about the relative locations of observed stimuli (i.e., it is "unaware" of the raw sensor measurements for each stimulus). Then, the two machines will not be able to discern any differences between their "worlds", even if they freely communicate observed stimulus relationships to one another. These devices will only discover the differences between their stimulus systems if each one measures the locations of its stimuli relative to common "reference" stimuli that both devices are able to observe.

The nonlinear signal processing method presented in this paper could be used as a representation "engine" in the "front end" of an intelligent sensory device [8]. It would produce channel-independent and sensor-independent stimulus locations that could be used by the device's pattern recognition module in order to recognize stimuli. Because the effects of channels and sensors have been "filtered out" of these representations, devices with different channels and sensors could use the same pattern recognition module, without recalibrating their detectors or retraining the recognizer. As described in Section 3, this may make it possible to design detector-independent computer vision devices. Furthermore, 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, there is the possibility of creating a speech-like telecommunications system that is resistant to channel-induced corruption of the transmitted information [5]. In such a system, information is carried by the coordinate-system-independent relationships among the transmitted power spectra.

Up to now, we have assumed that the channel and sensor of a machine are time-independent and that a machine derives a parallel transfer mechanism from the history of all previously-encountered stimuli. However, if a machine derived its affine connection from stimuli encountered in the most recent time interval ΔT, it could adapt to channels and/or sensors that drift over a longer time scale. It would continue to represent relative stimulus locations in the same way, as long as the velocity covariance matrices of the stimulus trajectory were temporally stable. Notice that the trajectories constructed in Section 3 have this kind of temporal stability. This process of adaptation, which is analogous to human adaptation in the goggle experiments, was demonstrated with simple experimental examples in reference [6].

The remarkable channel-independence and sensor-independence 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 varying because of changing observational conditions? 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 way the observer views them (i.e., independent of the choice of sensor and channel). 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 because they 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

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

[2] R. Held and R. Whitman, *Perception: Mechanisms and Models* (Freeman, San Francisco, 1972).

[3] D. N. Levin, "A differential geometric description of the relationships among perceptions", J. Math. Psych. **44**, 241-284 (2000).

[4] D. N. Levin, "Stimulus representations that are invariant under invertible transformations of sensor data", Proc. SPIE **4322**, 1677-1688 (2001).

[5] D. N. Levin, "Blind normalization of speech from different channels", J. Acoust. Soc. Am. **116**, 519-529 (2004) .

[6] D. N. Levin, "Representations of sound that are insensitive to spectral filtering and parameterization procedures", J. Acoust. Soc. Am. **111**, 2257-2271 (2002).

[7] D. N. Levin, "Sensor-independent stimulus representations", Proc. Nat. Acad. Sci. (USA) **99**, 7346-7351 (2002).

[8] D. N. Levin, US Patent #6,687,657 (2004) and pending patents.

[9] F. Reif, *Fundamentals of Statistical and Thermal Physics* (McGraw-Hill, New York, 1965).

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

[11] T. Sauer, J. A. Yorke, and M. Casdagli, "Embedology", J. Stat. Phys. **65**, 579-616 (1991).

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

[13] S. Weinberg, *Gravitation and Cosmology* (John Wiley & Sons, New York, 1972).