Vision Research 137 (2017) 1–23

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

Two representations of a high-dimensional perceptual space

Jonathan D. Victor *, Syed M. Rizvi, Mary M. Conte

Feil Family Brain and Mind Research Institute, Weill Cornell Medical College, 1300 York Avenue, New York, NY 10065, United States

ARTICLE INFO

Article history:
Received 12 July 2016
Received in revised form 27 April 2017
Accepted 3 May 2017

Keywords:
Local features
Visual textures
Multipoint correlations
Intermediate vision
Border salience

ABSTRACT

A perceptual space is a mental workspace of points in a sensory domain that supports similarity and difference judgments and enables further processing such as classification and naming. Perceptual spaces are present across sensory modalities; examples include colors, faces, auditory textures, and odors. Color is perhaps the best-studied perceptual space, but it is atypical in two respects. First, the dimensions of color space are directly linked to the three cone absorption spectra, but the dimensions of generic perceptual spaces are not as readily traceable to single-neuron properties. Second, generic perceptual spaces have more than three dimensions. This is important because representing each distinguishable point in a high-dimensional space by a separate neuron or population is unwieldy; combinatorial strategies may be needed to overcome this hurdle.

To study the representation of a complex perceptual space, we focused on a well-characterized 10-dimensional domain of visual textures. Within this domain, we determine perceptual distances in a threshold task (segmentation) and a suprathreshold task (border salience comparison). In N = 4 human observers, we find both quantitative and qualitative differences between these sets of measurements. Quantitatively, observers' segmentation thresholds were inconsistent with their uncertainty determined from border salience comparisons. Qualitatively, segmentation thresholds suggested that distances are determined by a coordinate representation with Euclidean geometry. Border salience comparisons, in contrast, indicated a global curvature of the space, and that distances are determined by activity patterns across broadly tuned elements. Thus, our results indicate two representations of this perceptual space, and suggest that they use differing combinatorial strategies.

Significance Statement: To move from sensory signals to decisions and actions, the brain carries out a sequence of transformations. An important stage in this process is the construction of a "perceptual space" – an internal workspace of sensory information that captures similarities and differences, and enables further processing, such as classification and naming. Perceptual spaces for color, faces, visual and haptic textures and shapes, sounds, and odors (among others) are known to exist. How such spaces are represented is at present unknown. Here, using visual textures as a model, we investigate this. Psychophysical measurements suggest roles for two combinatorial strategies: one based on projections onto coordinate-like axes, and one based on patterns of activity across broadly tuned elements scattered throughout the space.

1. Introduction

Perceptual spaces are internal workspaces within a sensory modality. By providing a representation that captures similarities and differences, perceptual spaces form a stage of sensory processing that not only supports simple discrimination judgments but also enables higher levels of processing, such as classification and naming. Our goal here is to understand the nature of this representation, using the perceptual space of image statistics (Victor et al., 2015) as a model. Along with (Edelman, 1998), our use of the term "representation" refers not only to the points of the perceptual space (i.e., to individual stimuli), but also, to similarity judgments (i.e., to how distances between stimuli are computed).

Among perceptual spaces, the space of human trichromatic color vision is the oldest and best known example (Maxwell, 1860). However, many other perceptual spaces have been identified: not only in vision (for faces (Catz, Kampf, Nachson, & Babkoff, 2009; Freiwald, Tsao, & Livingstone, 2009; Tanaka, Meixner, & Kantner, 2011; Valentine, 1991; Wallraven, 2014) and other objects (Wallraven, 2014)) but also in other sensory modalities (Bushdid, Magnasco, Vosshall, & Keller, 2014; Gaissert,

* Corresponding author.
E-mail address: jdvicto@med.cornell.edu (J.D. Victor).

http://dx.doi.org/10.1016/j.visres.2017.05.003
0042-6989/

Wallraven, & Bulthoff, 2010; Geffen, Gervain, Werker, & Magnasco, 2011; Koulakov, Kolterman, Enikolopov, & Rinberg, 2011; McDermott, Schemitsch, & Simoncelli, 2013; McDermott & Simoncelli, 2011; Yoshioka, Bensmaia, Craig, & Hsiao, 2007; Zaidi et al., 2013).

While color space is perhaps the most widely studied, many of its characteristics are not generic. For primate color vision, the properties of the three cone classes determine the dimensions of the space (Baylor, Nunn, & Schnapf, 1987), provide it with a coordinate system, and enable construction of stimuli that modulate each coordinate independently (Derrington, Krauskopf, & Lennie, 1984). For other perceptual spaces, the dimensionality is much larger, and these perceptual dimensions do not map in a straightforward way to the physics of the stimulus (Bushdid et al., 2014; Freiwald et al., 2009; Koulakov et al., 2011; Victor, Thengone, Rizvi, & Conte, 2015; Cho, Yang, & Hallett, 2000; Portilla & Simoncelli, 2000). Thus, it is not even guaranteed that generic perceptual spaces have a coordinate system, or that it is possible to find a set of independent perceptual dimensions. Nevertheless, these more complex perceptual spaces also support threshold and suprathreshold judgments.

Because typical perceptual spaces are multi-dimensional, representing them via "brute-force" strategies – in which each discriminable stimulus is represented by a separate neuron (or neural population) – is biologically implausible, because of a dimensional explosion of the resources required. If there are D independent dimensions and N discriminable values on each of the corresponding axes, there would be N^D distinct points in the space. In the case of color ($D = 3$ and $N > 100$), this leads to an estimate of over 10^6 distinct stimuli (colors) that need to be represented. For olfactory stimuli, it is estimated that D is much larger than 10 (Koulakov et al., 2011), and the total number of discriminable stimuli has been estimated at $> 10^{12}$ (Bushdid et al., 2014). The space of visual textures, the present focus, is also high-dimensional; to analyze how it is represented, we study regions within a well-characterized 10-dimensional subset (Victor & Conte, 2012; Victor et al., 2015).

The dimensional explosion in resources required for a brute-force representation can be mitigated by combinatorial strategies. One class of such strategies makes use of coordinates for the space (e.g., the amount of each color primary). By projecting the entire space onto each axis, a high-dimensional space can be efficiently represented in terms of its one-dimensional projections. A second class of strategies does not rely on a coordinate system in the usual sense, but instead postulates that neurons have a diverse set of broadly-tuned sensitivities. Interestingly, theoretical arguments suggest that this strategy becomes efficient for spaces of dimensionality $D \geqslant 3$ (Zhang & Sejnowski, 1999).

While both kinds of strategies are combinatorial, they make contrasting predictions about distances. Consider an experiment that measures perceptual distance between test points that are displaced in opposite directions from a reference point near the center of the space. In this experiment, we measure the perceptual distance as the amount of the displacement increases – that is, as the test points are pulled further and further apart. In a coordinate-based representation, the perceptual distance can only increase – since the distance between the projections of the two test points onto any axis must increase, as the test points move away from the reference. But in a representation based on patterns of activity across broadly-tuned neurons, other outcomes are possible. For example, suppose that most of the neurons are tuned to regions near the center of the space, and very few of them cover its periphery – as would be expected from an efficient deployment of resources (Hermundstad et al., 2014). Then, as the test points move into the periphery, fewer and fewer neurons contribute to their representations, and they therefore become less distinguishable.

These considerations motivate our approach to probing the representation of visual textures. In one experiment, we measure discrimination thresholds; in another, we measure suprathreshold perceptual distances. Our results suggest that both kinds of combinatorial strategies are used to compute distances – a coordinate-based representation that accounts for discrimination thresholds, and a distributed representation that accounts for the global perceptual geometry of the space.

2. Materials and methods

The experiments described here consist of two kinds of psychophysical measurements: threshold judgments, using a texture segmentation paradigm, and suprathreshold judgments, using a border salience paradigm. Both paradigms made use of the same domain of visual textures; we describe this domain first and then describe the specifics of the two paradigms.

2.1. The stimulus space

The stimulus domain is a continuum of visual textures. The parameters that describe the textures – i.e., the coordinates of the space – are a set of image statistics, each of which measures a specific local correlation (described below). Importantly, the texture associated with a particular set of values of the image statistics is a "maximum-entropy" ensemble: a collection of images, or, equivalently, a single infinite image, that are as random as possible, given the specified values of the statistics. This ensures that the image statistics fully determine the information available to the visual system. The stimuli used in the experiments are then random samples of this ensemble. For full details concerning the domain and sampling algorithms, see ((Victor & Conte, 2012); additional background and rationale may be found in other publications that use this domain (Hermundstad et al., 2014; Victor, Thengone, & Conte, 2013; Victor et al., 2015).

Each texture is a binary (black-and-white) coloring of a grid of checks. The parameters associated with a given texture are the probabilities of occurrence of each of the ways that 2×2 neighborhoods can be colored. Although 16 such colorings are possible ($16 = 2^{2\times2}$), there are only 10 degrees of freedom – because the 16 probabilities must sum to 1, and the overlapping portions of adjoining 2×2 blocks necessarily must match. It is natural to recast these 10 degrees of freedom in terms of local correlations, which are the coordinates of the space. Note that here we are referring to the coordinates of the stimuli themselves, which need not correspond to coordinates of a perceptual representation.

This strategy yields four groups of coordinates, corresponding to first-, second-, third-, and fourth-order correlations (Fig. 1A). (An nth-order correlation means that n checks must be simultaneously considered to determine the correlation's value.) Each of these 10 coordinates ranges from -1 to +1; the origin of the space (the texture corresponding to a value of 0 for each coordinate) is a completely random binary image.

Coordinates are designated as follows. The single first-order coordinate, γ, is the difference between the probability of a white check and the probability of a black check. It indicates the luminance bias: $\gamma = +1$ means that all checks are white, $\gamma = -1$ means that all checks are black, and $\gamma = 0$ means that both are equally likely.

The four second-order coordinates, denoted β_{-}, $\beta_{|}$, $\beta_{\backslash}$, and $\beta_{/}$, measure two-point correlations, in the orientations indicated by