Cognition 179 (2018) 178–191

Contents lists available at ScienceDirect

Cognition

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

Original Articles

Continuous to discrete: Ensemble-based segmentation in the perception of multiple feature conjunctions

Igor S. Utochkin*, Vladislav A. Khvostov, Yulia M. Stakina

National Research University Higher School of Economics, Russian Federation

ARTICLE INFO

Keywords:
Ensemble summary statistics
Segmentation
Categorization
Texture discrimination
Features and conjunctions
Sampling

ABSTRACT

Although objects around us vary in a number of continuous dimensions (color, size, orientation, etc.), we tend to perceive the objects using more discrete, categorical descriptions (e.g., berries and leaves). Previously, we described how continuous ensemble statistics of simple features are transformed into categorical classes: The visual system tests whether the feature distribution has one or several peaks, each representing a likely "category". Here, we tested the mechanism of segmentation for more complex conjunctions of features. Observers discriminated between two textures filled with lines of various lengths and orientations, which had same distributions between the textures, but opposite directions of correlations. Critically, feature distributions could be "segmentable" (only extreme feature values and a large gap between them) or "non-segmentable" (both extreme and middle values with smooth transition are present). Segmentable displays yielded steeper psychometric functions indicating better discrimination (Experiment 1). The effect of segmentability arises early in visual processing (Experiment 2) and is likely to be provided by global sampling of the entire field (Experiment 3). Also, rapid segmentation requires both feature dimensions having a "segmentable" distribution supporting division of the textures into categorical classes of conjunctions. We propose that observers select items from one side (peak) of one dimension and sample mean differences along a second dimension within the selected subset. In this scenario, subset selection is a limiting factor (Experiment 4) of texture discrimination. Yet, segmentability provided by the sharp feature distributions seems to facilitate both subset selection and mean comparison.

1. Introduction

Our capacity to attend to objects and store them in the working memory for deep processing is very limited (Cowan, 2001; Luck & Vogel, 1997; Pylyshyn & Storm, 1988; Scholl, 2001). However, in everyday perception we often encounter hundreds of objects at one time, but do not have difficulties in seeing them all. How can these hundreds of objects survive the severe limits of the processing bottleneck? One possible answer is that the visual system represents multiple objects in the compressed form of ensemble summary statistics (Alvarez, 2011; Cohen, Dennett, & Kanwisher, 2016). From lossy individual representations, a rather precise summary of many objects is computed (Alvarez & Oliva, 2008; Alvarez, 2011; Ariely, 2001; Parkes, Lund, Angelucci, Solomon, & Morgan, 2001). There is evidence that ensemble summaries are encoded directly as perceptual properties (Burr & Ross, 2008; Corbett, Wurnitsch, Schwartz, & Whitney, 2012; Norman, Heywood, & Kentridge, 2015); they require as much attention as representing individual properties of a single object but without focusing on each object (Alvarez & Oliva, 2008; Chong & Treisman, 2005a; Huang, 2015; Robitaille & Harris, 2011; Utochkin & Tiurina, 2014; but see Myczek & Simons, 2008; Allik, Toom, Raidvee, Averin, & Kreegipuu, 2013; Maule & Franklin, 2015) and not much resource demanding (Alvarez & Oliva, 2008; Bauer, 2017; Epstein & Emmanouil, 2017; but see Jackson-Nielsen, Cohen, & Pitts, 2017). Ensemble summaries can be computed for basic sensory domains: size (Ariely, 2001; Chong & Treisman, 2003), orientation (Alvarez & Oliva, 2008; Dakin & Watt, 1997), color (De Gardelle & Summerfield, 2011; Huang, 2015; Maule & Franklin, 2015), brightness (Bauer, 2009), direction and speed of motion (Emmanouil & Treisman, 2008; Watamaniuk & Duchon, 1992) and for complex perceptual features (Haberman & Whitney, 2007; Yamanashi-Leib, Kosovicheva & Whitney, 2016). The distributional properties of ensembles also influence how elements are grouped or segregated in perception. For example, whether two groups of elements are discriminated from each other depends on their mean difference normalized by variance (Corbett et al., 2012; Fouriezos, Rubenfeld, & Capstick, 2008; Im & Halberda, 2013; Rosenholtz, 2000; Utochkin & Tiurina, 2014). Mean-variance ratios also play an important role in determining which elements become salient and gain priority in

* Corresponding author at: Psychology Dept, Higher School of Economics, 101000, Armyansky per., 4, room 419, Moscow, Russian Federation.
E-mail address: isutochkin@inbox.ru (I.S. Utochkin).

https://doi.org/10.1016/j.cognition.2018.06.016
Received 29 December 2017; Received in revised form 18 June 2018; Accepted 23 June 2018

the control of attention and gaze (Avraham, Yeshurun, & Lindenbaum, 2008; Haberman & Whitney, 2012; Nothdurft, 1992, 1993a, 1993b; Palmer, Verghese, & Pavel, 2000; Rosenholtz, 2001; Rosenholtz, Huang, & Ehinger, 2012; Rosenholtz, Huang, Raj, Balas, & Ilie, 2012). Recent studies have shown that ensemble processing can go beyond mean and variance and that complex distributional properties can be learned explicitly (Oriet & Hozempa, 2016) or implicitly (Chetverikov, Campana, & Kristjánsson, 2016, 2017a, 2017b).

1.1. Segmentability: distribution-based segmentation and categorization

Our recent work (Utochkin & Yurevich, 2016; Utochkin, 2015) focused on testing the idea that a set of multiple items can be grouped or segregated based on the shape of the feature distribution (namely, the peaks and gaps in it). This theory explains how a distribution of continuous visual features can be rapidly transformed into discrete classes of objects in perception. If one peak is present, then the set is more likely to be perceived as consisting of same-type objects. By contrast, if several peaks are presented and interleaved with gaps then the items represented by each peak are more likely to form different-type objects, or categorical classes (Utochkin, 2015). Correspondingly, same-type objects are better grouped, while different-type objects are readily segmented, even if the types are intermixed in the space. For example, leaves on a tree can widely vary in the fall from green to red as extremes, but individual shades can be intermediate. The presence of these intermediate shades makes the transition between green and red smooth, so this produces a single-peak distribution recognized as a set of one-type objects. By contrast, in summer, leaves and ripe berries also vary between green and red. But the transition between the extremes is much more abrupt, so one would more easily see this set as two overlapping sets of different-type objects. Utochkin and Yurevich (2016) tested this theory in visual search experiments with basic visual features, including size and orientation. Their critical manipulation concerned the transition between feature values of distractors. In some trials, the transition was sharp, when features values were distributed with a big (e.g., 0°, 22.5°, and 45° of orientation) or even extreme (e.g., 0° and 45°) step. In other trials, the transition was smooth (e.g., 0°, 5°, 10°, …, 45°). Utochkin and Yurevich (2016) found that search efficiency was related to transition step non-monotonically: The search among smoothly distributed distractors was the fastest one; the search among "extremely" distributed distractors (two-peak distribution) was slower; and three-peak sharp distribution yielded the slowest search. A concept of "segmentability" was introduced to explain the non-monotonic effect of transition. Sharp transitions between the feature values provide the internal distribution with peaks corresponding to each presented value and large gaps between these peaks, which should lead to segmentation of the set into categorically different subsets (Utochkin, 2015). Smooth transition provides a single-peak broadband internal distribution without large gaps, which leads to the representation of all the items as one group. Each subset is analyzed as a separate chunk and rejected serially, making search among sharply distributed distractors slower (Duncan & Humphreys, 1989; Humphreys & Müller, 1993; Müller, Humphreys, & Donnelly, 1994).

1.2. Segmentation of multiple conjunctions: A general framework

In real-world perception, multiple objects rarely show variation, grouping, or segmentation along a single dimension. Each object varies in many features, forming an individual feature conjunction; taken together, multiple objects can provide a vast variety of conjunctions (Tsotsos, 1987). The variety of conjunctions as a function of their constituent feature statistics can be described in terms of inter-feature *correlation*. The correlation (or any other concordance measure) is an effective way to estimate how likely certain features in one dimension go with certain features in another dimension. An analogy can be made between descriptive ensemble statistics (e.g. average size) in global feature discrimination, and correlational statistics in global conjunction discrimination. Comparing two sets of sticks with different average sizes, we can say where longer sticks prevail. Likewise, comparing two sets with different orientation-size correlations, we can say where horizontal and bigger items prevail. Therefore, our discrimination between conjunction-defined sets of multiple objects is the matter of "seeing" feature correlations.

The question of how the visual system treats multiple feature conjunctions has rich theoretical links. Perhaps, the most fundamental topic this question is related to is the "binding problem" (Cave & Wolfe, 1999; Treisman, 1999). Perceived objects and scenes are presumably represented as elementary features and parts in the early sensory analysis performed by isolated and independent sensory modules (Treisman, 2006; Yantis, 2014) and should be somehow integrated correctly. As complete and momentary binding seems to be very computationally demanding (Tsotsos, 1987), it is prone to some limitations. A theoretical debate concerns the locus of binding limits and strategies that the visual system uses to deal with them (e.g., Di Lollo, 2012; Duncan & Humphreys, 1989; Cave & Wolfe, 1999; Rosenholtz, Huang, et al., 2012; Treisman, 2006; Treisman & Gelade, 1980, etc.). The perception of correlated features in multiple items is a question of binding to some degree. How accurately can orientations and sizes be ascribed together to the sticks from our example, given that even separate features are represented quite approximately – as an ensemble summary rather than a set of precise values for each stick? Taking into account the fundamental problem related to binding of elementary features the questions of our work are following: Can the mechanism of ensemble-based segmentation, which was described for one-dimensional case, work over sets defined by a correlation of features? How efficient it might be?

Based on our segmentability theory (Utochkin, 2015), our general prediction is that the segmentability of conjunction-defining features should facilitate the discrimination between sets with different feature correlations, because their statistics make it easier to select local and more contrast classes for comparison. Specifically, we propose that an ability to perform segmentation allows to restrict processing to one of the categorical subsets formed along one dimension, thus making differences along another dimension more pronounced. We ran four experiments to test the effect of segmentability on the discrimination of sets (textures) defined by two correlated features and to probe a potential mechanism for this effect. We chose the conjunctions of orientation and length, as these two features are very common and well-studied in numerous previous work on ensemble and texture perception (orientation: Alvarez & Oliva, 2008; Attarha & Moore, 2015; Cha & Chong, 2018; Dakin & Watt, 1997; Nothdurft, 1992, 1993a, 1993b; Parkes et al., 2001; Rosenholtz, 2000, 2001, etc.; size - length or area: Ariely, 2001; Attarha, Moore,& Vecera, 2014; Chong & Treisman, 2003; Bauer, 2017; Im & Halberda, 2013; Myczek & Simons, 2008; Oriet & Brand, 2013; Robitaille & Harris, 2011; Utochkin & Tiurina, 2014, etc.).

2. Experiment 1

2.1. Methods

2.1.1. Participants

Five expert observers, including the authors of this article, participated in the experiment. Their age varied between 20 years and 43 years old, median age was 21 years old. All had normal or corrected-to-normal vision and no neurological problems.

2.1.2. Apparatus and stimuli

Stimulation was developed and presented through PsychoPy for Linux (Pierce, 2007). Stimuli were presented on a standard VGA monitor with a refresh frequency of 75 Hz and a 800 × 600-pixel spatial resolution. A 26° × 26° square at the center of the screen was used as the "working" field for presenting stimuli; the rest screen space