Vision Research 135 (2017) 54–64

Contents lists available at ScienceDirect

Vision Research

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

Attentive pointing in natural scenes correlates with other measures of attention

Daniel M. Jeck[a,b,*], Michael Qin[c], Howard Egeth[d], Ernst Niebur[a,d,e]

[a] Zanvyl Krieger Mind/Brain Institute, Johns Hopkins University, Baltimore, MD, United States
[b] Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD, United States
[c] Department of Biomedical Engineering, University of Connecticut at Storrs, United States
[d] Department of Psychological and Brain Sciences, Johns Hopkins University, Baltimore, MD, United States
[e] Solomon Snyder Department of Neuroscience, Johns Hopkins University, Baltimore, MD, United States

ARTICLE INFO

Article history:
Received 20 September 2016
Received in revised form 13 March 2017
Accepted 6 April 2017

Keywords:
Pointing
Tapping
Salience
Attention
Probability density estimation
Saliency map
Natural scenes
Fixations
Interest points
Probability distributions

ABSTRACT

Finger pointing is a natural human behavior frequently used to draw attention to specific parts of sensory input. Since this pointing behavior is likely preceded and/or accompanied by the deployment of attention by the pointing person, we hypothesize that pointing can be used as a natural means of providing self-reports of attention and, in the case of visual input, visual salience. We here introduce a new method for assessing attentional choice by asking subjects to point to and tap the first place they look at on an image appearing on an electronic tablet screen. Our findings show that the tap data are well-correlated with other measures of attention, including eye fixations and selections of interesting image points, as well as with predictions of a saliency map model. We also develop an analysis method for comparing attentional maps (including fixations, reported points of interest, finger pointing, and computed salience) that takes into account the error in estimating those maps from a finite number of data points. This analysis strengthens our original findings by showing that the measured correlation between attentional maps drawn from identical underlying processes is systematically underestimated. The underestimation is strongest when the number of samples is small but it is always present. Our analysis method is not limited to data from attentional paradigms but, instead, it is broadly applicable to measures of similarity made between counts of multinomial data or probability distributions.

© 2017 Elsevier Ltd. All rights reserved.

1. Introduction

Factors influencing selective attention can notionally be separated into top-down and bottom-up influences. Top-down influences depend on the internal state of the observer, including his or her goals (*e.g.* Yarbus, 1967; DeAngelus & Pelz, 2009). Bottom-up influences are factors that draw attention independently of any task and past experience with particular stimuli (*e.g.* Anderson, Laurent, & Yantis, 2011). For example, a bright flash in an otherwise still scene will usually attract attention (Yantis & Jonides, 1984). The ability of parts of a visual scene to attract attention in a bottom-up fashion has been called the *salience* of this region (Koch & Ullman, 1985), a definition we adopt here.

While the definitions of top-down and bottom-up attention are clear, it is in practice difficult to dis-entangle their effects. For instance, observers who repeatedly perform tasks designed to measure bottom-up attentional effects may form expectations of what the next trial may be. These expectations will change their internal state and therefore add a top-down component to their responses. One of the goals of this study is to reduce such effects. Specifically, our goals are to:

- Introduce open ended self reports as a new experimental assay for selective attention and show that it can be measured efficiently using a pointing/tapping paradigm.
- Develop a new experimental design in which each participant views only a small numbers of scenes. This reduces the contamination of bottom-up attentional effects by top-down expectations due to participants viewing similar stimuli many times.
- Compare the results of this experiment with three other measures of attention and salience: fixations, interest points, and computed saliency.

* Corresponding author at: Zanvyl Krieger Mind/Brain Institute, Johns Hopkins University, Baltimore, MD, United States.

E-mail address: danny.jeck@gmail.com (D.M. Jeck).

http://dx.doi.org/10.1016/j.visres.2017.04.001
0042-6989/© 2017 Elsevier Ltd. All rights reserved.

- Analyze the effects of sample size on estimating correlation between maps. The small number of samples from the pointing/tapping paradigm results in a statistical effect that causes the correlation between different maps to be systematically underestimated. We will clarify the influence of finite numbers of samples on the correlation between maps.

1.1. Determining bottom-up saliency from human behavior

There are several methods that allow researchers to characterize items or regions that observers direct their attention to. One very influential approach has been visual search. Search for targets that differ from distractors by one of several low-level features (*e.g.* luminance, color, orientation contrast) takes a (generally short) time that is nearly independent of the number of distractors in the display (Egeth, Jonides, & Wall, 1972; Treisman & Gelade, 1980). In contrast, targets that could be distinguished from distractors only by combinations of such features require search times that increased roughly linearly with the number of distractors (Egeth, Virzi, & Garbart, 1984; Treisman & Gelade, 1980). These and related results were fundamental in the construction of computational models for visual search (Wolfe, 1994, 2007; Wolfe, Cave, & Franzel, 1989) and for saliency determination and attentional selection (Niebur & Koch, 1996; Itti, Koch, & Niebur, 1998; Itti & Koch, 2001).

Given past success in utilizing features that promote efficient search, it is tempting to continue using visual search as a way to test models of visual salience. However, search tasks are limited in their applicability to measuring salience because participants are typically informed about the types of images they are about to see (*e.g.* "an image in which there is a single target and many distractors"), and the target and distractors are often described before the task begins. This information generates top-down influences that are likely to interact with bottom-up selection mechanisms. Even when participants are only told to look for a unique target, without being informed how it will differ from other objects ("odd-man out" tasks), they are still being informed about the structure of the image. It is then difficult to decide whether the participants find the target due to its bottom-up saliency features, or because of its uniqueness (Bacon & Egeth, 1994). Results therefore may reflect a mixture of bottom-up (saliency) and top-down components of unknown composition.

This concern applies also to measurements of salience where participants give their subjective assessment of which of two stimuli is more salient (*e.g.* Nothdurft, 2000). These experiments require that participants know that a stimulus will appear made up of oriented bars where two of them (one to the left and one to the right of fixation) will differ from the rest. As with search tasks, this information potentially biases the response of the participant. Indeed Nothdurft refers to needing additional concentration (clearly a top down process) to make difficult salience assessments. Furthermore, even if participants are not informed explicitly about the nature of the visual scene they are observing, the process of performing a task many times will likely give them information about what to expect.

While top-down influences can probably never be excluded entirely, our goal in this project is to reduce them. One possible way to mitigate top-down influences is to use "overt attention" in a free viewing task as an indicator for covert attention. In this approach, introduced by Parkhurst, Law, and Niebur (2002) and used in many subsequent studies (for a review see Borji & Itti, 2013), observers look at images (or videos) which can be natural or abstract scenes while their eye movements are tracked. Areas of the scene that are fixated are taken to be attended, a conclusion supported by findings from Deubel and Schneider (1996) that visual discrimination performance is enhanced at saccade targets. In the absence of a specific task ("free viewing"), it seems reasonable to assume that at least for the first few images, and for the first few fixations in these images, observers let themselves be guided by the visual input, rather than by some more complex strategy. This assumption becomes less plausible, however, the longer the sequence of images becomes and the longer the duration becomes that observers look at any given image. Indeed, Parkhurst et al. (2002) found that the agreement between eye fixation data and predictions of a purely bottom-up computational model of saliency decreased with viewing time/fixation number for a given image. It is not known whether the level of agreement depended on how many images had been viewed previously.

In principle it is possible to use the eye tracking method, with naïve participants viewing only a small number of scenes. In practice, the overhead of setting up an eye tracker system for each participant would make gathering fixation data for a small number of images per participant a very cumbersome task. We recruited 252 participants in this study, an order of magnitude more than participated in the latest saliency benchmark by Borji and Itti (2015), making eye-tracking each subject prohibitive.

To counteract this difficulty, we developed a novel experimental paradigm with the goal of gathering data from many participants where each participant only performed a small number of trials. The new paradigm is centered on showing subjects a short sequence of images and recording the response of each subject to each image. Some of the images are simple displays (similar to typical visual search arrays like those used by Treisman & Gelade (1980)) that are designed to test a specific hypothesis about what features of an image affect salience. Future work will discuss the structure of these images and the results gathered. Alternating with these images are natural scenes, the focus of this report. The goal in presenting these scenes to participants is to determine the extent to which salience as measured in our new experimental paradigm comports with salience data from previous studies. The natural scenes were therefore a subset of those used in a previous study (Masciocchi, Mihalas, Parkhurst, & Niebur, 2009), and we will compare results obtained in our new paradigm with those from that study.

The data being compared here are attentional maps aggregated over a pool of participants. Such maps have been used in the study of salience extensively (Borji & Itti, 2013), and because they are population averages we can gather data to make attentional maps from a similar population without needing to gather new fixation data from the same subjects.

1.2. Reporting attended locations by pointing to them

Our new experimental paradigm for fast assessment of attentional selection was inspired by a study by Firestone and Scholl (2014) although those authors used a very different stimulus set and had a different motivation. The main idea is that, instead of recording eye movements, we ask participants to communicate their selections in a natural way by tapping on a screen with their (index) finger. Specifically, we ask the subjects to "tap the first place you look when the image appears." This instruction gives us a quick way to communicate in a non-technical manner that the participant should select the first attended location on the image, rather than an arbitrary point as requested by Firestone and Scholl (2014). Even though instructions refer to where the participants look first, we do not attempt to determine whether any single individual is able to report their eye movements successfully. Instead, we are concerned with whether the population-level attentional maps we derive from the responses reflect previous measures of attention. We will validate our method by comparing these maps on when gathered for the same set of images.

Download English Version:

https://daneshyari.com/en/article/5705910

Download Persian Version:

https://daneshyari.com/article/5705910

Daneshyari.com