Vision Research 118 (2016) 25–30

Contents lists available at ScienceDirect

Vision Research

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

On the possible roles of microsaccades and drifts in visual perception

Ehud Ahissar [a,*], Amos Arieli [a], Moshe Fried [b], Yoram Bonneh [c]

[a] Department of Neurobiology, Weizmann Institute of Science, Rehovot, Israel
[b] Goldschleger Eye Research Institute, Tel Aviv University, Tel Hashomer, Israel
[c] Department of Human Biology, University of Haifa, Haifa, Israel

ARTICLE INFO

Article history:
Received 22 July 2014
Received in revised form 7 November 2014
Available online 19 December 2014

Keywords:
Active sensing
Active vision
Retinal coding
Closed loops
Visual sampling

ABSTRACT

During natural viewing large saccades shift the visual gaze from one target to another every few hundreds of milliseconds. The role of microsaccades (MSs), small saccades that show up during long fixations, is still debated. A major debate is whether MSs are used to redirect the visual gaze to a new location or to encode visual information through their movement. We argue that these two functions cannot be optimized simultaneously and present several pieces of evidence suggesting that MSs redirect the visual gaze and that the visual details are sampled and encoded by ocular drifts. We show that drift movements are indeed suitable for visual encoding. Yet, it is not clear to what extent drift movements are controlled by the visual system, and to what extent they interact with saccadic movements. We analyze several possible control schemes for saccadic and drift movements and propose experiments that can discriminate between them. We present the results of preliminary analyses of existing data as a sanity check to the testability of our predictions.

1. Introduction

Saccades (Ss) are rapid eye movements that shift the line of sight ("visual gaze") of both eyes simultaneously. Large saccades (LSs) have amplitudes >2 deg, peak velocities of up to >500 deg/s, and durations of 20–100 ms. Within their primary natural working range (<10 deg) (Dorr et al., 2010) LSs obey the "main sequence" relationship: the ratio between their amplitude and peak velocity is constant (Bahill, Clark, & Stark, 1975; Findlay & Walker, 2012). Therefore, saccade durations are confined to smaller modulations; for perceptual relevant saccades, when amplitudes change by 20 folds (from 0.5 to 10 deg), durations change only by 2-fold (from 20 to 40 ms) (Bahill, Clark, & Stark, 1975).

Microsaccades (MSs) are small saccades; they exhibit all characteristics of LSs described above, but their amplitudes are smaller than the diameter of the foveal region, i.e., typically less than 2 deg (Engbert, 2006). During free viewing LSs occur several times per second, typically 2–4 times. Each LS shifts the gaze rapidly to a new target, where the eye dwells for several hundreds of millisecond before jumping to the next target (Fig. 1A). Each dwelling period (e.g., Fig. 1B) is termed a "fixational pause" (Barlow, 1952) and its duration is affected by global planning as well as by the visual information acquired during the pause (Sereno & Rayner, 1992). The sequence of saccadic targets, which forms the scanning path of the entire scene, is also affected by global planning as well as visual information acquired during the scan of the scene (Noton & Stark, 1971; Walker-Smith, Gale, & Findlay, 1977). MSs' occurrence rate depends on the task (e.g., Bonneh et al., 2014; Bonneh et al., 2010; Fried et al., 2014; Rolfs, Kliegl, & Engbert, 2008; Siegenthaler et al., 2014; Stampe & Reingold, 2002). In fixation tasks, where subjects are instructed to fixate at one point, MSs occur at rates similar to, though a bit lower than, LSs (Otero-Millan et al., 2008). During free viewing they occur at much lower rates as described below.

2. Microsaccades function

Several functions had been suggested along the years for MSs. Two such functions assign to MSs a central role in the perception of external objects. In one, it is suggested that MSs are not different in function from LSs, and as such they redirect the visual gaze to new locations within the foveal region (Cunitz & Steinman, 1969; Ko, Poletti, & Rucci, 2010; Zuber, Stark, & Cook, 1965). In the other it is implied that MSs are used to directly encode visual details while gaze shifts are determined by LSs (e.g., McCamy et al., 2014). Encoding by MSs can be done in principle in two ways: by scanning during movement or by "flashing" (i.e., resetting photoreceptors and reactivating them according to the new spatial configuration) upon landing (Rolfs, 2009a). "Flash" encoding by

* Corresponding author.
E-mail address: ehud.ahissar@weizmann.ac.il (E. Ahissar).

http://dx.doi.org/10.1016/j.visres.2014.12.004
0042-6989/

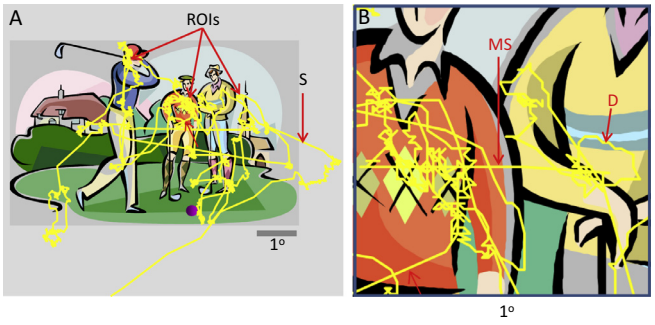

Fig. 1. (A) Example of ocular scanning of an image. The trajectory of a human subject's gaze (of one eye) during free viewing of an image presented on a computer screen is depicted. ROIs, regions of interest; S, saccade. Courtesy of Shira Ozana. (B) Zoom-in into a 1° × 1° area in the middle ROI of A. MS, microsaccade; D, drift.

MSs was examined and found to be a plausible mechanism for improving near-threshold visual detection (Elsner & Deubel, 1986). However, such a mechanism cannot underlie encoding of fine visual details because there is simply not enough information in the spikes generated at the retina upon MS landing and before smearing due to drift begins (Appendix 2 in Ahissar & Arieli, 2012).

We examined the plausibility of MS encoding by scanning. The suggestion is that the retinal slip (i.e., the shift of the visual image across the retina) induced by a MS activates photoreceptors and as a result visual details are encoded at the output of retinal ganglion cells. Critically, however, MS velocities are too fast for fine retinal encoding. Regular visual acuity can reach the limit of receptor granularity: spatial frequencies of about half of the receptor frequency can be perceived (Hirsch & Curcio, 1989). As the spacing between neighboring image peaks equals to the width of only two retinal receptors, a pre-requisite for the perception of this image is that the activities of two neighboring receptors are distinguishable. Visual encoding by MSs thus requires that activities of neighboring receptors should be distinguishable when the eye traverses a stationary grid during a MS (Fig. 2).

MS velocities obey the main sequence ratio and vary between ~20 and 90 deg/s for MSs of 0.1–1 deg (McCamy et al., 2014). The diameter of the smallest foveal receptors in typical human subjects is about 1/3 of an arcminute (or 1/180 of a deg). A retinal slip induced by such MS induces inter-receptor-delay (IRD) of about 0.06–0.3 ms. Evidently, these temporal delays are too brief for any significant retinal coding: First, single receptors will be hardly activated during such brief exposures even at very high contrasts (Vuong, Chabre, & Stryer, 1984). Second, if antagonistic (e.g., center-surround) RFs exist in the fovea, then their inhibitory and excitatory zones will be activated virtually simultaneously, likely preventing any meaningful retinal response (Amthor & Grzywacz, 1993). Third, even if retinal responses would be generated at these temporal intervals it is not conceivable that downstream stations could phase lock to them or decode them because synaptic and conduction temporal variances are an order of magnitude larger (Faisal, Selen, & Wolpert, 2008). Thus, if MSs are used for visual encoding this could work only for coding of much lower spatial frequencies.

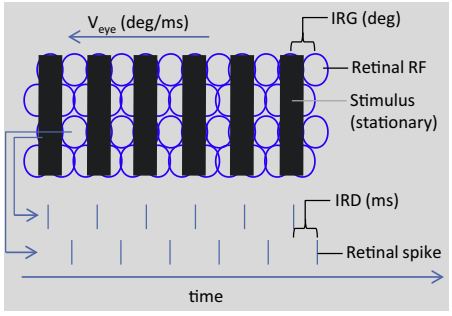

Fig. 2. A schematic illustration of scanning of a stationary spatial grid. The moving retinal receptor mosaic is illustrated by an array of blue circles, each representing a single receptive field (RF), which moves to the left at a velocity V_{eye}. The spikes generated via two single RFs are illustrated by two rows of vertical blue lines at the bottom, each row representing spikes of one RF (arrows). Inter-receptor-delay (IRD) is determined by inter-receptor-gap (IRG) divided by the velocity of the eye (V_{eye}).

However, MSs do not seem to be good candidates for encoding coarse vision either. Their occurrence rate is too low for any meaningful coding during free viewing. Free viewing is dominated by LSs. MSs occur only once per 3 to10 LSs, and when they occur they occur late in the fixational pause – typically later than 250 ms into the pause (McCamy et al., 2014). With such sparse occurrence it is not clear how any meaningful coding can be based on MSs. On the other hand, these data are consistent with a gaze-redirection role of MSs – more frequent redirections are expected to occur within more informative regions of the scene.

The conclusion from these considerations is that the role MSs play in vision is similar to that of LSs, only for within-fovea targets (Ko, Poletti, & Rucci, 2010; Poletti, Listorti, & Rucci, 2013), as originally suggested by Zuber, Stark, and Cook (1965) based on their similar kinematics. Thus, LSs and MSs initiate a new acquisition period, outside or within the current ROI, which lasts, typically, a few hundreds of milliseconds. This is consistent with the control mechanisms of LSs and MSs being shared (Hafed & Krauzlis, 2012), as originally suggested by Rolfs, Laubrock, and Kliegl (2006) and Rolfs, Kliegl, and Engbert (2008) based on theoretical considerations and supported by kinematic analysis (Otero-Millan et al., 2008). One interesting prediction of this suggestion is that during free viewing of natural scenes the clear preference for horizontal and vertical MS directions that is shown in artificial conditions (Rolfs, 2009a) should significantly decrease, or even disappear, as gaze redirection is expected in all possible directions.

3. Drift function

Acquisition of the visual details of stationary external objects can only be done via eye movements (eyeM) (Ahissar & Arieli, 2001). As MSs cannot be used for such acquisition, the only remaining candidate is the ocular drift (D). During every fixational pause the eye drifts around in a random-walk manner (Ahissar & Arieli, 2001; Barlow, 1952; Bengi & Thomas, 1972; Eizenman, Hallett, & Frecker, 1985; Ratliff & Riggs, 1950) (Fig. 1B). This drift motion is associated with high-frequency (>30 Hz) low-amplitude (<1′) fluctuations of the eye called 'tremor'. Tremor is often assumed to reflect the operation of an independent motor source, superimposed on the drift motion. However, a more parsimonious interpretation would assume that the tremor fluctuations reflect the fundamental 'steps' of the random-walk-like process underlying the drift motion. D mean velocities are in the order of 10–100 arcmin/s, and they typically cover areas of about ten to a few tens of arcminutes in diameter during a fixational pause (Cherici et al., 2012; Engbert, 2006). Importantly, D mean velocities translate to IRD values of ~3–30 ms at the smallest receptors of the fovea. Such temporal delays can be reliably decoded by neuronal circuits (Ahissar, 1998; Ahissar & Arieli, 2012). D occurs at all directions and thus can sample all visual details in a given image. Moreover, temporal encoding via D preserve all details required for perceiving shapes, textures, locations and motions of external objects (Ahissar & Arieli, 2012).

We thus propose the following scheme of visual encoding at the retina. Saccades (Ss, including LSs and MSs) shift the fovea from