**V1 Saliency Hypothesis (V1SH, Li 1999, 2002): V1 creates a bottom-up saliency map to guide attentional/gaze shifts**

V1SH explains why color, orientation, and motion feature singletons attract attention

- **Orientation singleton**
  - Neurons responding to the vertical bar escape iso-orientation suppression
  - Neurons responding to the red bar escape iso-color suppression

- **Color singleton**
  - Neurons responding to the leftward moving iso color singleton escape suppression
  - Neurons responding to the red vertical bar escape iso-orientation and/or iso-color suppression

- **Motion singleton**
  - Neurons responding to the red bar escape iso-motion suppression

- **Motion-color orientation singleton**
  - Neurons responding to the vertical bar escape iso-motion and/or iso-color suppression

V1 neurons are tuned to orientation, color, or motion direction, and some of them respond simultaneously to color and orientation, or to motion direction and orientation

**V1SH prediction: Attention capture by an eye-of-origin singleton — because V1 neurons are tuned to eye-of-origin, and V1 has iso-eye-of-origin suppression confirmed experimentally (Zhaoping, 2008, 2012)**

An oculomotor singleton, though irrelevant in a task to search for an uniquely oriented bar, pops out to attract attention and gaze shift, even though observers do not name it visually distinctive and are often unaware of its presence. This attraction prolongs reaction time (RT) to find the target. This observation (Zhaoping, 2008, 2012), predicted by V1SH, is a fingerprint of V1, which is the only cortical area with substantial number of neurons tuned to eye-of-origin of visual inputs.

In the dichoptic input above, the uniquely oriented bar is salient, because the V1 neuron tuned to and responding to it escapes iso-orientation suppression. The ocular (eye-of-origin) singleton is also salient, because the responding neuron escapes the iso-eye-of-origin suppression. In this example, the ocular singleton is task irrelevant but is more salient than the orientation singleton, attracting the first saccade in the search task.

References:
Li Zhaoping, University of Tübingen, Max Planck Institute for Biological Cybernetics, Germany, li.zhaoping@tuebingen.mpg.de

**Abstract:** I proposed in the late 1990s and early 2000s that the primary visual cortex V1 creates a bottom-up saliency map from visual inputs (Li 1999, 2002). This is the V1 Salience Hypothesis (V1SH). According to V1SH, the highest V1 response to any visual location represents this location’s saliency, defined as the strength of exogenous attentional attraction. Hence, the visual location evoking the largest V1 response is most likely to attract visual attention or evoke a gaze shift driven by bottom-up visual inputs. For example, a vertical bar in a background of horizontal bars is salient; so is a red dot among green dots, because the feature singletons evoke higher V1 responses than the homogeneous background items. Contextual influences via the intracortical horizontal connections make V1 neurons’ responses dependent on contextual inputs outside the classical receptive field (e.g., Knierim & van Essen 1992). These contextual influences are such that neurons tuned to the same or similar features suppress each other, so that the neurons preferring and responding to the homogeneous background inputs suppress each other’s responses. The neurons responding to the feature singletons escape such suppression and the singletons are thus salient. Here I review some of the supporting evidence in human and monkey data from past works by my collaborators and myself. More info see http://www.lizhaoping.org/zhaoping/V1Saliency.html

**V1 SH prediction:** given the same visual input, higher V1 responses (due to neural response fluctuations or optogenetic stimulation) should lead to faster saccades to the feature singleton --- confirmed experimentally in monkeys (Yan et al. 2018)

**Monkey’s task:** start a trial by fixating at the central fixation point; after the bars appear, saccade to the orientation singleton (whose position is unpredictable) as quickly as possible.

**V1 neurons responses to the orientation singleton (spikes/sec) before the saccade onset (usually >200 ms)**

V1 neurons’ responses to the search target were measured while the monkey was doing the task. Usually it took more than 200 ms after stimulus onset before the saccadic onset. At around 40-50 ms after the stimulus onset, given the same visual input, with higher initial V1 neural responses to the target were more likely to evoke faster and accurate subsequent saccades to the target. This latency, at 40-50 ms, is so short that higher cortical areas or superior colliculus cannot provide this saliency signal to V1. This latency agrees with the time course of contextual influences observed in Knierim and van Essen 1992. While et al. 2017 did not see this saliency signal at this short latency, but they saw a signal at a latency longer than that in superior colliculus, in their monkeys doing a similar task. Yan et al. 2018 showed that a signal at such a longer latency arises from top-down factors.

**fMRI/ERP evidence: salient but imperceptible inputs (the cue below) activate early visual cortex but not the parietal cortex (Zhang, Zhaoping, Zhou, Fang, 2012)**

**V1 SH fully predicts reaction times using zero parameters, suggesting that V2 and other visual cortical areas do not contributing to saliency computation (Koen & Zhaoping 2007, Zhaoping & Zhe 2015)**

Mathematically, V1SH predicts the RT, of detecting singleton pop out by triple-feature (CMO) uniqueness in color (C), motion direction (M) and orientation (O) from RTs for single- and double-feature uniqueness (RTM, RTC, RTO, RTCO) by equation

\[ \min(\text{RTC}, \text{RTM}, \text{RTO}, \text{RTCO}) = \min(\text{RTC} + \text{RTM} + \text{RTO} + \text{RTCO}) \]

This prediction uses no parameters, and is derived from the fact that V2 has no cells simultaneously tuned to color, orientation, and motion direction. V2 and other extrastriate cortical areas do have such cells, so if they were involved in the saliency computation, the equality above would become an inequality. RT data is consistent with the V1 prediction, suggesting that V2 and above do not contribute to the saliency computation.

**References:**
- Li Z. (1999) Contextual influences in V1 as a basis for pop out and asynaptic in visual search. PNAS. USA 96:10530-10535