Top-Down Models Explain Key Aspects Of A Speed-Of-Sight Character Recognition Task

Garrett T. Kenyon[1,2], Shawn Barr[2], Michael Ham[1], Vadas Gintautas[1], Cristina Rinaudo[1,2], Ilya Nemenman[3], Marian Anghel[1], Steven Brumby[1], John S. George[1], Luis M. A. Bettencourt[1]

[1] Los Alamos National Laboratory, Los Alamos, NM   [2] New Mexico Consortium, Los Alamos, NM   [3] Emory University, Atlanta, GA

garkenyon@gmail.com

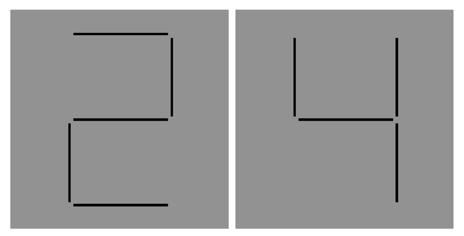
Abstract

Object recognition can be very rapid, typically reaching completion within 150 msec following image onset, consistent with inter-saccade intervals in humans. In a Speed of Sight task, the recognition process can be interrupted by presentation of a mask after a short delay—termed the Stimulus Onset Asynchrony (SOA). Uniform gray or white noise images are minimally effective as masks, even at very short SOAs (e.g. 20 ms). Optimal masks can compromise object identification at SOAs of 80 ms or more. We conducted a 2AFC experiment in which subjects reported the location (left/right) of a target presented next to a distractor, with both images quickly replaced by identical masks. To limit image parameters while allowing task difficulty to be varied, images were depicted on a 7-segment LED-like display. Targets were always a specific digit (e.g. 2 or 4). Masks and distractors consisted of digits, letters or non-semantic symbols composed from the same 7 segments. To account for the observed variability in mask efficacy for different target–mask combinations, we constructed a model that combined dynamical variables representing feedforward feature detectors—corresponding to the 7 image segments—with high-level pattern detectors for targets, masks and distractors. In human subjects: the numeral 8 was an effective universal mask, requiring SOAs of nearly 80 msec to reach criterion, whereas the numeral 1 was a poor mask, allowing many targets to be reliably detected after only a 20 msec SOA. To account for such a wide range of effective SOAs, top-down influences from model pattern detectors were needed to maintain the activity of low-level segment detectors following target offset. Our results suggest that masking occurs at the level of component features and is strongly modulated by top-down processes, inconsistent with pure feedforward models often proposed to account for Speed of Sight results.

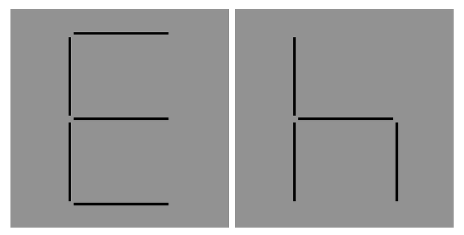
Methods

LED Task

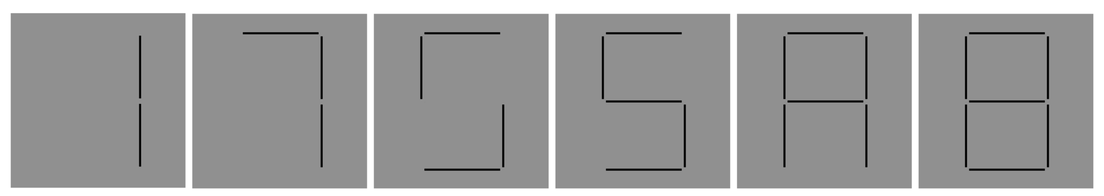
In order to allow more complete, parametric control of the relationship between target and mask images, we explored a task in which targets and masks were composed of a subset of 7 oriented, "LED" segments. Unlike other tasks, this allowed the subject to predict the precise visual characteristics of the target.



(a) Targets



(b) Distractors



(c) Masks

Figure 1: *Stimuli and masks were composed of familiar LED patterns. Each mask contained a unique number of segments, ranging from 2-7.*

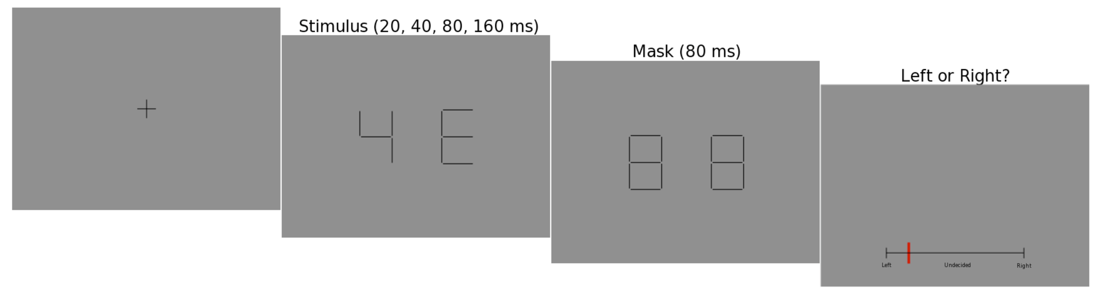
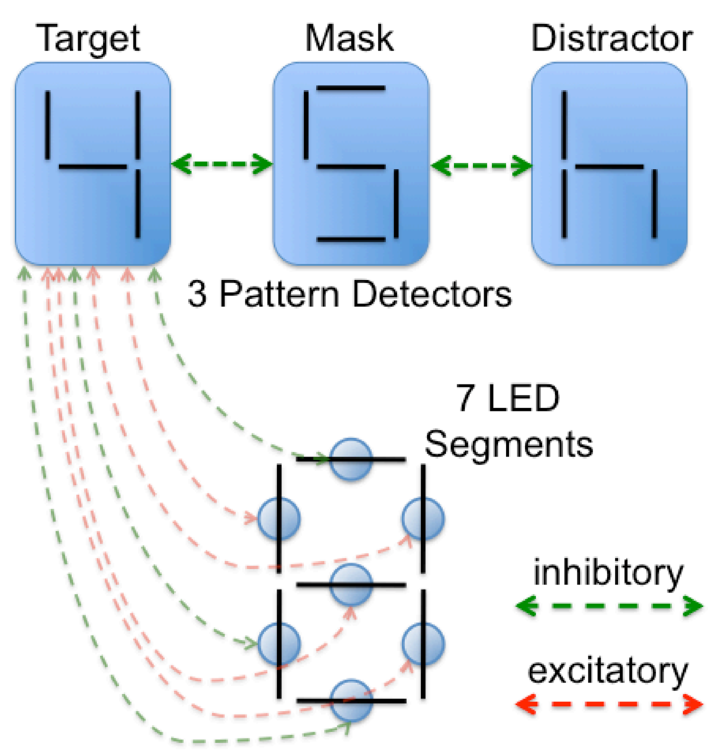


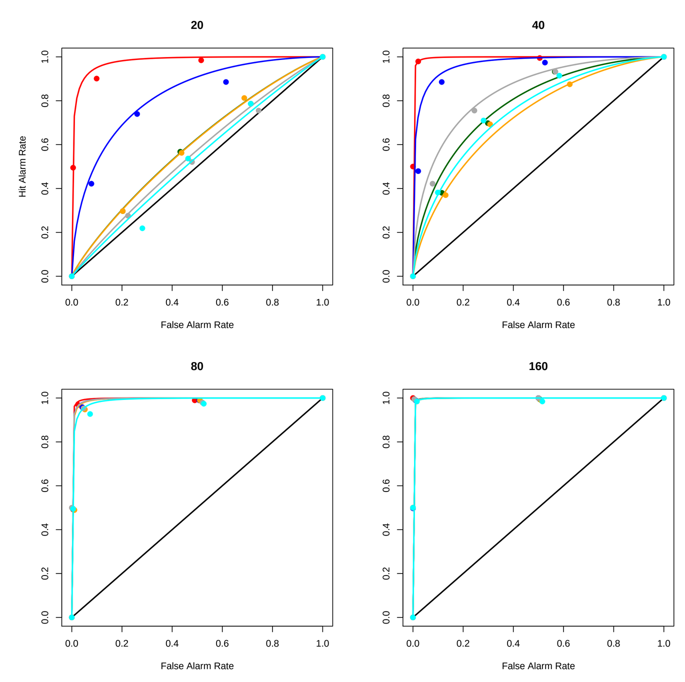
Figure 2: *Schematic of the 2AFC experimental trial. Each trial began with a fixation cross followed by the stimulus. The stimulus consisted of a target and a distractor presented side-by-side for a duration of 20, 40, 80 or 160 ms. A mask immediately followed the stimulus. Participants were instructed to respond by marking their confidence along the response bar.*

Model

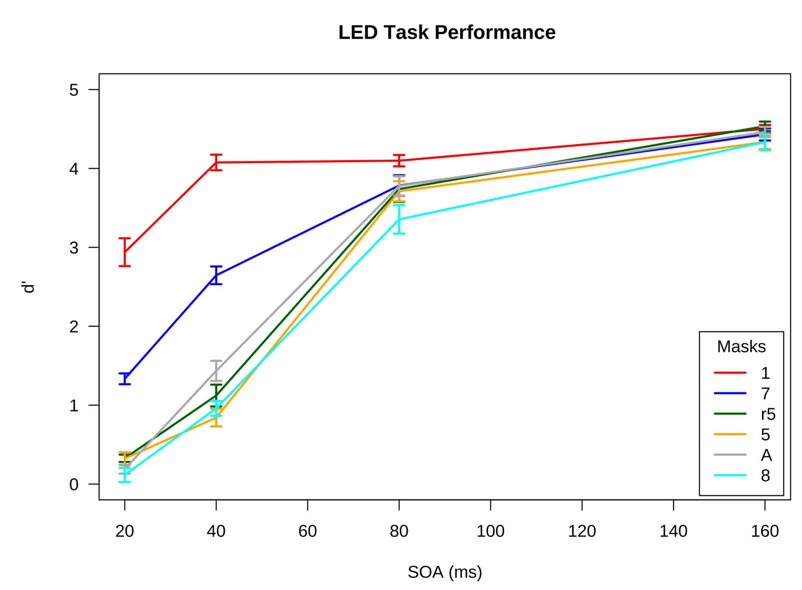


- Pattern detectors for each of the seven segments.
- Reciprocal excitatory (green arrows) or inhibitory (red arrows) connections between patterns and detectors.
- Feedback enhanced of suppressed appropriate low-level features.
- Performance was measured as the duration of the response exceeding a threshold.

Results

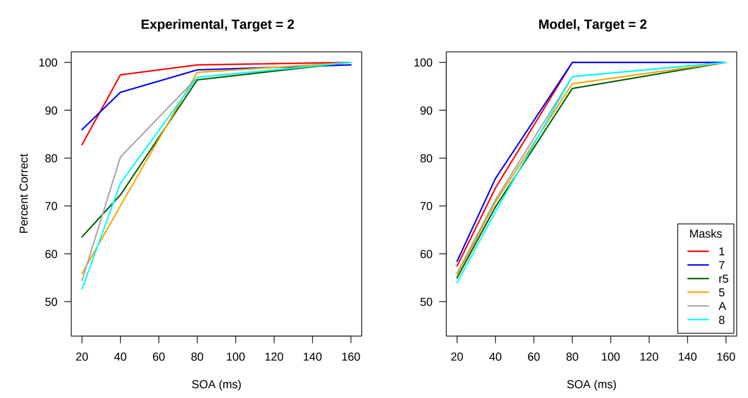


(a) ROC Curves

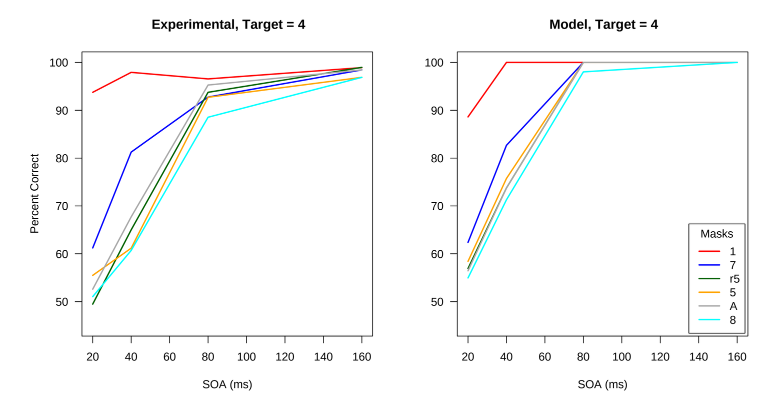


(b) Average d'

Figure 4: *For short SOAs the "1" and "7" mask are clearly less effective. (Data from both plots were averaged over targets and distractors.)* **(a)** *ROC curves for each mask were constructed by partitioning confidences into four equally-spaced bins. Hit/false alarm rate was plotted as a point for each bin; equal-variance, binormal curves were fitted to the average d'.* **(b)** *d' averaged over subjects (N=8) was plotted as a function of SOA. Error bars show standard error of the mean.*



(a)



(b)

Figure 5: **(a)** *In both panels, the numeral "1" is the least effective mask, allowing performance to reach approximately 90% for a 20 msec SOA. Due to the top-down feedback implemented in the model, the "4-detector" remained active for a prolonged period following the onset of the mask, thereby delaying activation of the "1-detector". Because the "1" LED pattern contained no segments that were not also present in the "4" LED pattern, there was no direct inhibition to counter the top-down feedback. Similar considerations explain the intermediate effectiveness of the "7" mask, which contained only 1 extra LED segment not present in the "4" LED pattern.* **(b)** *Same organization as previous figure, except the target object was the LED pattern "2". The model accounted for the slightly reduced effectiveness of the "1" and "7" LED masks, although each contained one additional segment and thus was predicted to be a relatively effective mask. We anticipate that incorporating expansive non-linearities into the model may explain why masks containing two additional LED segments are more effective at impeding recognition than are masks containing only 1 extra segment.*

This work was supported by the the Los Alamos National Laboratory LDRD program under project 20090006DR.