

The Role of Attention on the Minimum Presentation Duration Required for Scene Recognition



Berhan Şenyazar¹, Albert Ali Salah^{1,2}, İnci Ayhan^{1,3}

¹Cognitive Science MA Program, Boğaziçi University, Istanbul / Turkey
²Department of Computer Engineering, Boğaziçi University, Istanbul / Turkey
³Department of Psychology, Boğaziçi University, Istanbul / Turkey

{berhan.senyazar, inci.ayhan}@boun.edu.tr

Introduction

It has been suggested that little if any attention is required for scene recognition (Li, VanRullen, Koch, & Perona, 2002). The absence of the role of attention in scene recognition, however, has been challenged by Cohen, Alvarez, and Nakayama (2011) showing that basic-level scene categorization and object identification performance degrade while simultaneously performing an attention-demanding task. Here, we used a similar dual-task paradigm except that we used psychophysical methodology in order to find duration thresholds in a broader range of scene recognition tasks (Table 1).

Experiment

Participants

- Two authors and four psychophysically trained naive students all with normal or corrected-to-normal vision

Stimuli

- Scene images selected from SUN Database (Xiao, Hays, Ehinger, Oliva, & Torralba, 2010), scene attributes obtained from SUN Attribute Database (Patterson & Hays, 2012)
- 210 target and 210 distractor images (256x256 pixels) in each block
- Algorithm of Portilla and Simoncelli (2000) used to generate masking images using the spectral information of the image set of each block.

Table 1: Images For Each Task

Task	Target	Distractor
Global Property (Scene Function)	Highest Sports	Lowest Sports
Global Property (Spatial Structure)	Highest Openness	Lowest Openness
Superordinate-level Categorization	Highest Naturalness	Highest Man-made
Basic-level Categorization	Beach	Mountain
Scene Detection	Unused Random Images	Masking Images

Procedure

- The experiment conducted in a dark room using a Philips 109B40/20 CRT monitor (1024 × 768 screen resolution at 85 Hz refresh rate)
- A chinrest used to keep viewing distance at 45 cm
- Method of constant stimuli with seven levels of presentation duration (30 trials for each data point)

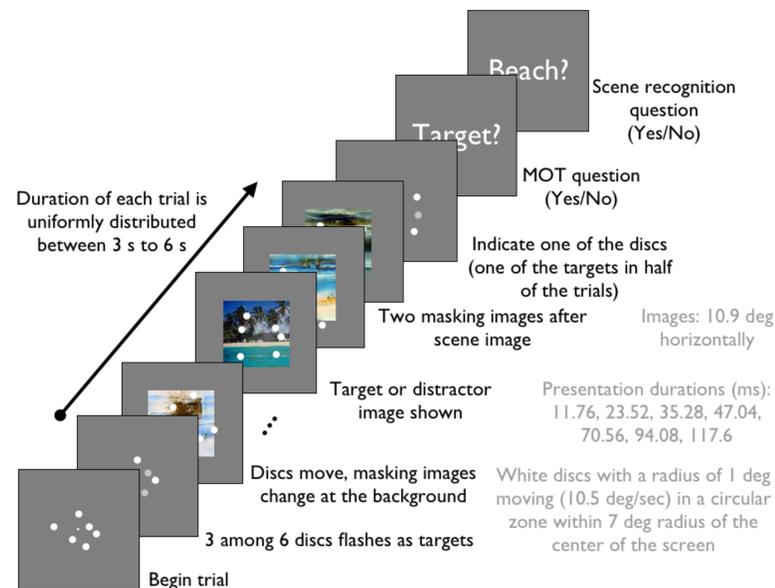


Figure 1: Procedure of dual-task condition. Single-task procedure is the same but participants are instructed to ignore discs and no MOT question asked at the end.

Results

- Data points fit into a Weibull function to determine the minimum duration at which the percentage of correct answers reached 75% point
- Data analyzed using a 5 × 2 within-subjects ANOVA

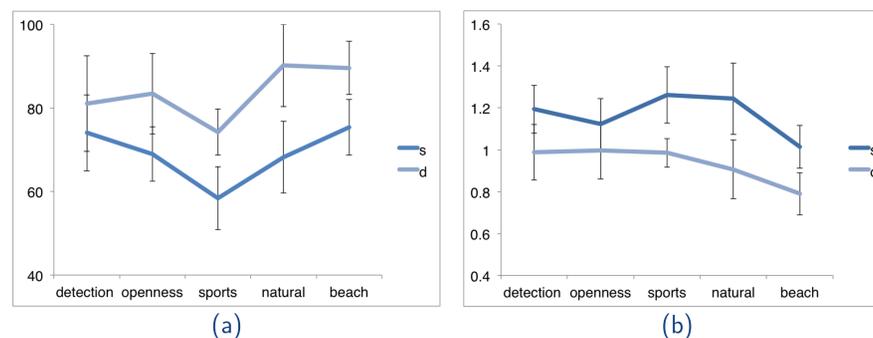


Figure 2: Means of 75% correct thresholds (ms) (2a) and d' values (2b) for single- and dual-task conditions of each task. Error bars show standard error.

Table 2: ANOVA Results of Duration Thresholds

Task	$F(4,20) = 2.14$	$MSe = 1.50$	$p = .113$	$\eta_p^2 = .30$
Attention	$F(1,5) = 13.46$	$MSe = 1.74$	$p = .014$	$\eta_p^2 = .73$
Task * Attention	$F(4,20) = .68$	$MSe = .91$	$p = .613$	$\eta_p^2 = .12$

Table 3: ANOVA Results of d'

Task	$F(4,20) = 2.38$	$MSe = .04$	$p = .087$	$\eta_p^2 = .32$
Attention	$F(1,5) = 21.19$	$MSe = .04$	$p = .006$	$\eta_p^2 = .81$
Task * Attention	$F(4,20) = .65$	$MSe = .03$	$p = .637$	$\eta_p^2 = .11$

Effect of Image Variability in SUN Dataset

- Intracategory variety in SUN Dataset controlled by analyzing the performance of participants for each image
- Hard items answered correctly by < 3 participants
- Easy items answered correctly by > 4 participants
- Statistical analysis repeated by excluding those items:

Table 4: Effect of Attention on Hard, Easy and Full Sets

Easy items excluded	$F(1,5) = 16.04$	$MSe = 1.85$	$p = .010$	$\eta_p^2 = .76$
All items included	$F(1,5) = 13.46$	$MSe = 1.74$	$p = .014$	$\eta_p^2 = .73$
Hard items excluded	$F(1,5) = 8.94$	$MSe = 1.68$	$p = .030$	$\eta_p^2 = .64$

Discussion

- That the minimum duration required to recognize a particular scene category or property does not significantly differ provides evidence against the hierarchical models that claims precedence of superordinate-level over basic-level categorization or the other way around.
- The relatively higher presentation duration required for the detection task indicates the need for further studies on expectation and top-down effects on scene recognition (i.e. semantics).
- There is a trend such that the effect of attention on the performance of detection task is negligible (Fig. 2a), which further supports studies that provide evidence for the decoupling of detection from other recognition tasks (Mack & Palmeri, 2010), as opposed to models of scene recognition where categorization is achieved at the time of detection (Grill-Spector & Kanwisher, 2005).
- More controlled datasets for scene recognition are needed in order to study the underlying mechanisms more reliably.

References

Cohen, M. A., Alvarez, G. A., & Nakayama, K. (2011, September). Natural-Scene Perception Requires Attention. *Psychological Science*, 22(9), 1165-1172.
 Grill-Spector, K., & Kanwisher, N. (2005, February). Visual Recognition As Soon as You Know It Is There, You Know What It Is. *Psychological Science*, 16(2), 152-160.
 Li, F. F., VanRullen, R., Koch, C., & Perona, P. (2002, July). Rapid natural scene categorization in the near absence of attention. *Proceedings of the National Academy of Sciences*, 99(14), 9596-9601.
 Mack, M. L., & Palmeri, T. J. (2010). Decoupling object detection and categorization. *Journal of Experimental Psychology: Human Perception and Performance*, 36(5), 1067.
 Patterson, G., & Hays, J. (2012). Sun attribute database: Discovering, annotating, and recognizing scene attributes. In *Computer vision and pattern recognition (cvpr), 2012* (see conference on (pp. 2751-2758)).
 Portilla, J., & Simoncelli, E. P. (2000). A Parametric Texture Model Based on Joint Statistics of Complex Wavelet Coefficients. *International Journal of Computer Vision*, 40(1), 49-70.
 Xiao, J., Hays, J., Ehinger, K. A., Oliva, A., & Torralba, A. (2010). SUN database: Large-scale scene recognition from abbey to zoo. *2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 3485-3492.

The Role of Attention on the Minimum Presentation Duration Required for Scene Recognition



Berhan Şenyazar¹, Albert Ali Salah^{1,2}, İnci Ayhan^{1,3}

¹Cognitive Science MA Program, Boğaziçi University, Istanbul / Turkey

²Department of Computer Engineering, Boğaziçi University, Istanbul / Turkey

³Department of Psychology, Boğaziçi University, Istanbul / Turkey

{berhan.senyazar, inci.ayhan}@boun.edu.tr

*

References

Cohen, M. A., Alvarez, G. A., & Nakayama, K. (2011, September). Natural-Scene Perception Requires Attention. *Psychological Science*, 22(9), 1165–1172.

Grill-Spector, K., & Kanwisher, N. (2005, February). Visual Recognition As Soon as You Know It Is There, You Know What It Is. *Psychological Science*, 16(2), 152–160.

Li, F. F., VanRullen, R., Koch, C., & Perona, P. (2002, July). Rapid natural scene categorization in the near absence of attention. *Proceedings of the National Academy of Sciences*, 99(14), 9596–9601.

Mack, M. L., & Palmeri, T. J. (2010). Decoupling object detection and categorization. *Journal of Experimental Psychology: Human Perception and Performance*, 36(5), 1067.

Patterson, G., & Hays, J. (2012). Sun attribute database: Discovering, annotating, and recognizing scene attributes. In *Computer vision and pattern recognition (cvpr), 2012 IEEE conference on* (pp. 2751–2758).

Portilla, J., & Simoncelli, E. P. (2000). A Parametric Texture Model Based on Joint Statistics of Complex Wavelet Coefficients. *International journal of computer vision*, 40(1), 49–70.

Xiao, J., Hays, J., Ehinger, K. A., Oliva, A., & Torralba, A. (2010). SUN database: Large-scale scene recognition from abbey to zoo. *2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 3485–3492.