Twin mechanisms: Rapid scene recognition involves both feedforward and feedback processing

Zhijun Zhang, Xukun Zhang, Yusheng Sun, Zhejiang University, P. R. China.

Human visual system interacts with the environment to obtain the information we needed in real time. Previous studies have confirmed our ability to rapidly recognize natural images. Human could efficiently saccade to the side containing an animal in 120ms under a force-choice saccade task (Kircher & Thorpe, 2006), or completed the animal detection task well within 150ms after target onset (Thorpe, Fize, & Marlt, 1996). Therefore, some researchers supposed that the rapid scene recognition was a feedforward process based on the ventral stream from V1 to inferotemporal cortex (Serre, Oliva, & Poggio, 2007).

Scene images are composed of components with different space frequencies. Low spatial frequencies (LSF) conveyed coarse information about global shape rapidly, while high spatial frequencies (HSF) provided fine information about edges and details slowly (Delplanque, N'Diaye, Scherer, & Grandjean, 2007). LSF and HSF were mainly processed by magnocellular pathway and parvocellular pathway respectively (Nowak & Bullier, 1997; Vuilleumier, Armony, Driver, & Dolan, 2003).

So far, little was known about the integration of LSF and HSF. Some researchers supposed that rapid scene recognition followed a feedforward coarse-to-fine sequence (Musel, Giavarini, Chauvin, Guyader, & Peyrin, 2011).

However, recently, researchers proposed a mix model which suggested that rapid scene recognition could also involve feedback processing (Wyatte, Jilk, & O'Reilly, 2013; Sun, Zhang, & Wu., 2017).

The present study aimed to examine the interaction of different spatial frequency during rapid scene recognition. That is, whether different spatial frequencies were delivered in a strictly feedforward coarse-to-fine sequence.

Here, we adopted an adjusted response prime paradigm. Experiment 1 aimed to explore whether a rapid feedforward sweep of original natural images could reveal a response priming effect. Further, in the Experiment 2a, the LSF and HSF components of the same picture were set to be prime and target, respectively. According to the coarse-to-fine theory, a LSF-to-HSF sequence may reproduce the response priming effect. As a comparison, Experiment 2b adopted a HSF-to-LSF sequence.

Analysis of accuracy yielded a significant main effect of consistency ($F(1,14)=204.30$, $p<0.01$, $\eta^2=0.94$) in Experiment 1. The interaction of consistency and SOA was also significant ($F(4,56)=3.18$, $p<0.05$, $\eta^2=0.19$). The further simple effect analysis showed that the prime effect was larger when the prime-target SOA was 144ms, rather than 36ms, as shown in Fig.1

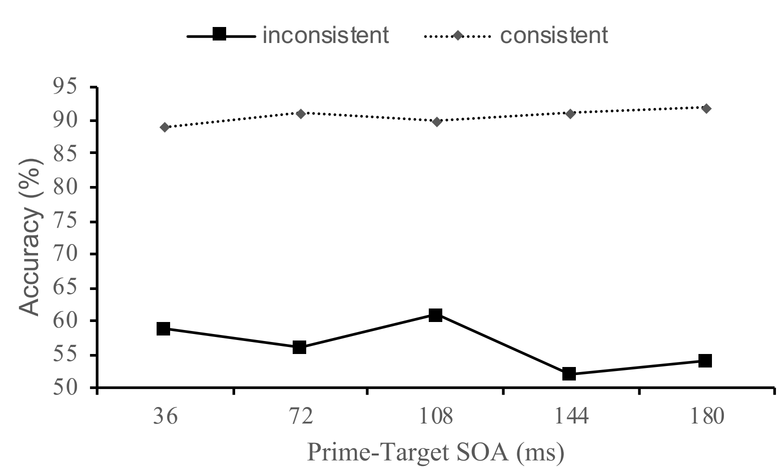

Fig. 1 The impact of consistency and Prime-Target SOA on accuracy in Experiment 1

The analysis of RT also yielded a significant main effect of consistency ($F(1,14)=60.22$, $p<0.01$, $\eta^2=0.80$). The interaction of consistency and SOA was also significant ($F(4,56)=3.90$, $p<0.05$, $\eta^2=0.22$). The further simple effect analysis showed that the prime effect was larger when the prime-target SOA was 144ms (42.28ms), rather than 36ms (22.19ms), as shown in Fig.2.

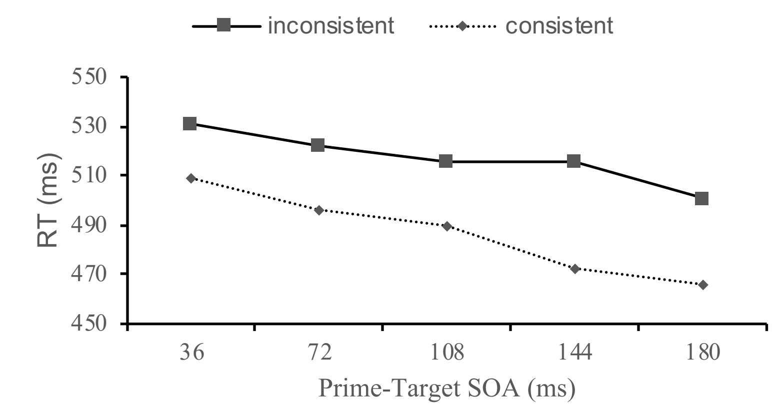

Fig. 2 The impact of consistency and Prime-Target SOA on RT in Experiment 1

So, using original natural images, we found a typical response prime effect of both accuracy and RT. The findings of Experiment 1 demonstrated that the processing of prime and target was behaviorally equivalent to a feedforward system even in a rapid sequence. A rapid sweep of prime image was sufficient enough to evoke target-related signal processing.

Analysis of accuracy yielded a significant main effect of consistency ($F(1,14)=50.33$, $p<0.01$, $\eta^2=0.78$) in Experiment 2a, as shown in Fig.3 . Analysis of RT yielded a significant main effect of consistency ($F(1,14)=14.91$, $p<.01$, $\eta^2=0.52$) and SOA ($F(4,56)=6.58$, $p<.01$, $\eta^2=0.33$), as shown in Fig.4. Similarly, analysis of accuracy yielded a significant main effect of consistency ($F(1,14)=130.80$, $p<0.01$, $\eta^2=0.91$) in Experiment 2b. Analysis of RT also yielded significant main effects of consistency ($F(1,14)=24.36$, $p<0.01$, $\eta^2=0.63$) and SOA ($F(4,56)=2.84$, $p<0.05$, $\eta^2=0.17$) .

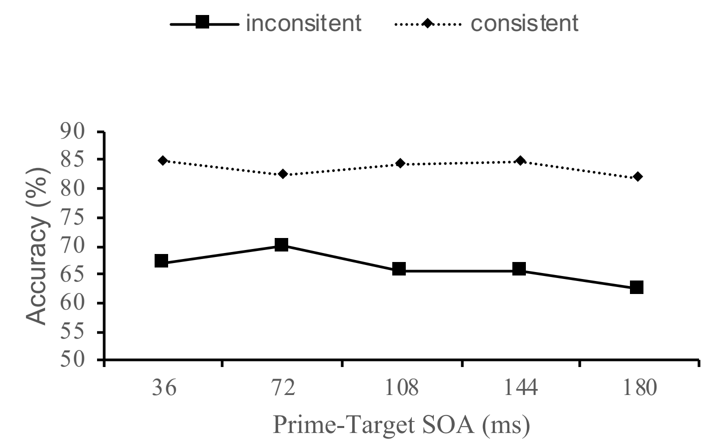

Fig. 3 The impact of consistency and Prime-Target SOA on accuracy in Experiment 2a

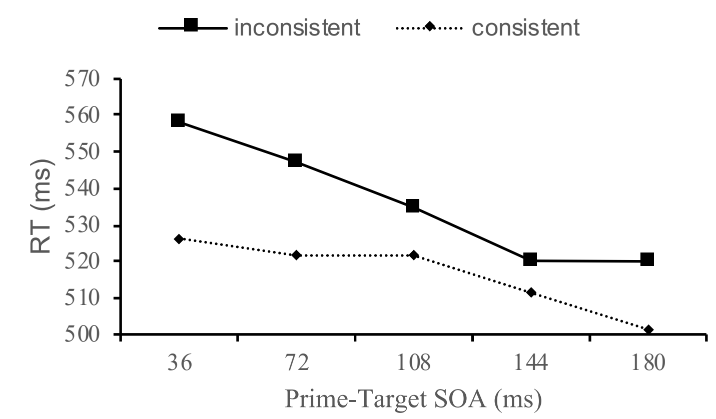

Fig. 4 The impact of consistency and Prime-Target SOA on RT in Experiment 2a

Therefore, LSF and HSF components both significantly influence the rapid scene recognition. However, neither the LSF-to-HSF sequence nor the HSF-to-LSF sequence reproduced the response priming effect as the original images. That is, the rapid scene recognition needed far more than a feedforward coarse-to-fine sweep, and the integration of different spatial frequency needed some early feedback loops

Corresponding author: zjzhang@zju.edu.cn (Z. Zhang)