Japanese Psychological Research THE JAPANESE PSYCHOLOGICAL ASSOCIATION

Japanese Psychological Research
2018, Volume 60, No. 3, 179–187

doi: 10.1111/jpr.12187

Brief Report

Failure to Replicate the Increased Temporal Resolution Induced by Images That Give Impression of Danger

KEISHI NOMURA and YUKO YOTSUMOTO* *University of Tokyo*

Abstract: Kobayashi and Ichikawa (2016) recently reported that briefly presented images with dangerous impressions were detected with higher accuracy than images with safe impressions and concluded that the emotion evoked by such images improves temporal resolution of visual perception. In this study, we assessed confounding effects of the color saturations of the images used in their study. While attempting to replicate their results, we found the opposite results—that is, images with safe impressions were detected with higher accuracy than those with dangerous impressions. This likely reflected an observed correlation between color saturations and detection thresholds. To confirm the effects of color saturations, in subsequent experiments, we independently examined the effects of emotion and of color characteristics. We concluded that the previously reported increased temporal resolution was due to the confounding effect of color saturation, and not by the evoked emotion.

Key words: visual perception, temporal resolution, emotion, affective picture, color.

People sometimes anecdotally report the experience of time slowing down when suddenly exposed to a life-threatening event. Although little empirical research to date has focused on this phenomenon, a study by Stetson, Fiesta, and Eagleman (2007) is the most pertinent. They predicted that, if time does indeed slow down for such people, this would entail increased temporal resolution. They introduced an analogy of the motion picture: Slow-motion replay allows finer temporal discrimination compared to replay at a normal speed, because of increase in the number of mental snapshots. To test this prediction, they measured participants' temporal thresholds during free-fall by presenting rapidly alternating digits on a hand-held device. They found no evidence of increased temporal resolution despite the fact that participants retrospectively overestimated the duration of their own falls. Their conclusion was that the experience of time slowing down is due to retrospective retrieval of an emotionally salient memory rather than increased temporal resolution.

Kobayashi and Ichikawa (2016), on the other hand, challenged this view. They selected negative and positive pictures with high arousal from the International Affective Picture System (IAPS; Lang, Bradley, & Cuthbert, 2008), an internationally distributed and widely used stimulus database for affective color pictures. They reported that observers detected briefly presented images that gave dangerous impressions with high accuracy and further concluded that the emotion evoked by such images improves temporal resolution of visual perception.

*Correspondence concerning this article should be sent to: Yuko Yotsumoto, Department of Life Sciences, University of Tokyo, Komaba, Meguro-ku, Tokyo 153-8902, Japan. (E-mail: cyuko@mail.ecc.u-tokyo.ac.jp)

However, the Kobayashi and Ichikawa (2016) study suffers from serious methodological limitations. The images used in their study were likely to be biased in terms of their image features. In fact, a previous study indicates that subsets of pictures selected from the IAPS tend to be biased in terms of their spatial frequency content (Delplanque, N'diaye, Scherer, & Grandjean, 2007). In a pilot study, we found that there were considerable biases in several image features, such as mean pixel values, between the two image groups. Therefore, it remains unclear whether the reported increased temporal resolution was due to the emotion evoked by the images or to the confounding effects of their image features.

Our initial aim was to assess possible confounding effects of basic image features on the increased temporal resolution reported in Kobayashi and Ichikawa (2016). In an attempt to replicate their results, we found a correlation between color saturation and detection thresholds (Experiment 1). Then we assessed the confounding effects of color saturations on the increased temporal resolution reported by Kobayashi and Ichikawa. In Experiment 2, we examined the effect of dangerous impressions of the images on temporal resolution of visual perception, independently of their image features, by means of texture synthesis. In Experiment 3, we investigated the effect of image color characteristics on temporal resolution measurements, separately from their dangerous impressions, using color transfer.

General Methods

Participants

Seven university students (one female and six males, mean age 23.3 ± 4.1 years) participated in all three experiments (Experiments 1, 2, and 3). All participants reported normal or corrected-to-normal vision. Four participated in Experiment 1 first while the other three first participated in Experiment 2. All participated in Experiment 3 after Experiments 1 and 2.

All participants gave written informed consent for their participation in the experimental protocol, which was approved by the institutional review boards of the University of Tokyo. In particular, participants were advised that some of the images could make them feel uncomfortable and that they could withdraw their participation at any time. All the participants completed the whole experiment without reporting any problems.

Stimuli and Apparatus

The stimuli were generated by MATLAB (MathWorks Inc., Natick, MA, USA) with the Psychtoolbox extension (Brainard, 1997; Pelli, 1997). All the stimuli were the same as those used in Kobayashi and Ichikawa (2016). The 12 negative images (image numbers in the IAPS: 1,050, 2,683, 2,811, 3,500, 6,520, 6,550, 6,560, 8,480, 9,414, 9,902, 9,908, and 9,940) and 12 positive images (image numbers in the IAPS: 1,440, 1,441, 1,630, 1,710, 1,920, 2,347, 5,199, 5,760, 5,780, 5,825, 7,325, and 7,492) were selected from the IAPS (Lang et al., 2008).

Using MATLAB, 24 grayscale images (1,024 pixels in width, 768 pixels in height) were generated from their color versions by eliminating hue and saturation but retaining luminance with the rgb2gray function (scale ranging from 0 to 255). The rgb2gray function calculates luminance by forming a weighted sum of the R, G, and B components: $0.2989 \times R + 0.5870 \times G + 0.1140 \times B$.

The stimuli were presented on a gamma-corrected 22-in. CRT monitor (Mitsubishi DiamondtronM2 RDF223H, luminance range of 0.033–83.186 cd m^{-2}) controlled by OptiPlex 9,020 (DELL, 800 × 600 pixels, 100 Hz refresh rate). The experiment was conducted in a dark room. The viewing distance was 57.3 cm, and each participant was asked to stabilize his or her head on a chin rest.

Rating

Following each experiment, all participants rated impressions of each color image using a 7-point bipolar scale (–3.0 to +3.0 with 0 as *neutral*). Impressions of each image were evaluated for dangerousness (*dangerous–safe*), pleasantness (*pleasant–unpleasant*), and impact (*shocking–ordinary*) in the same way as in Kobayashi and Ichikawa (2016). Each color image was presented once at the center of the black display for 2,000 ms in a random order for each participant.

To account for truncation near the lower or upper limit—that is, −3.0 or 3.0—we applied a simple statistical model. For each image, we set uniform prior distribution between −3.0 and 3.0 on the true average of impression ratings and modeled observed ratings as drawn from the normal distribution centered on this true average of impression ratings. Thus, the 95% credible intervals of the posterior distributions successfully fell within the reasonable range.

Experiment 1

We first investigated whether the results from the Kobayashi and Ichikawa (2016) study could be replicated. In each trial, an original color image was presented for 1,000 ms followed by a mask. In 75% of the trials, the image was briefly switched to the same image in monochrome for a variable duration. We measured a 50% threshold for detecting the monochrome image as an index of temporal resolution of visual perception. We then compared the average of the 50% thresholds between the images with dangerous impressions and those with safe impressions.

Methods

Stimuli. We used MATLAB to create the monochrome images (1,024 × 768 pixels) with the rgb2gray function (scale ranging from 0 to 255). The stimulus size was 12.0° wide and 9.0° high. The background of the display was black.

Procedure. Each trial began with a central fixation cross (0.7° × 0.7°), which was presented for 500 ms, immediately followed by one of the original color images for 1,000 ms. In 75% of the trials, the color image was then briefly switched to the same image in monochrome. There were six duration conditions for the monochrome image: 10, 20, 30, 40, 50, or 60 ms. In the other 25% of the trials, the color image remained for another 30 or 60 ms instead of the same image in monochrome. Then, the image turned into a mask, and the participants reported whether they saw the monochrome image by pressing the "j" or "k" button. The visualized trial format can be found in the original paper (Kobayashi & Ichikawa, 2016, p. 276, Figure 1).

Each image appeared twice in each of the six duration conditions for the monochrome images (288 trials) and two duration conditions for the color images (96 trials). A short break was given every 128 trials, and each participant completed 384 trials in total. The stimuli were presented in a random order for each participant.

Image analysis. Using MATLAB, the RGB values of all original color images were converted to CIE 1976 L*a*b* values with the rgb2lab function. The rgb2lab function converts RGB values to CIE 1931 XYZ values, and then converts the XYZ values to CIE 1976 L*a*b* values. For more details on the conversion formula, refer to the documentation by MathWorks Inc. (https://jp.mathworks.com/help/vision/ref/colorspaceconversion.html).

Saturation of a color image was calculated as the Euclidean distance from the origin on the a*b* color plane in L*a*b* color space (white point = D65).

Results and Discussion

Figure 1a shows the average impression ratings across participants for each original color image. The participants' impressions of the images were consistent with those reported by Kobayashi and Ichikawa (2016; dangerous for the negative images; safe for the positive images).

The average false-alarm rate was below 5% in both conditions: 1.2% (images with dangerous impressions) and 3.6% (images with safe impressions). A paired t test revealed no significant difference in false-alarm rates between the two image groups, $t(6) = -1.430$, $d_{Diff} = 0.584$ (Cohen's $d = 0.637$), $p = .202$. The subsequent analyses were restricted to the trials in which the monochrome image was presented.

A cumulative normal function was fitted to the raw data for each participant in each duration condition, from which we obtained the 50% thresholds for the detection of the monochrome images. Figure 2a shows the average of the 50% thresholds of all the participants for both conditions. A paired t test revealed that the 50% thresholds for the images with dangerous impressions were significantly higher than for the images with safe impressions, $t(6) = 5.324$, $d_{Diff} = 2.173$ (Cohen's $d = 0.468$), $p = .002$.

We tested whether there was any systematic relationship between the mean saturation of each original color image and the 50% thresholds calculated by each image. Our reasoning was that the saliency of a monochrome image is determined by the magnitude of the color–monochrome saturation change, because any other image features, including luminance, remain unchanged.

First, we compared averages of the mean saturations between the two image groups. Wilcoxon's rank sum test revealed no significant difference, $z = -1.241$, $r = -.253$, $p = .215$, although group average of mean saturation of the images with dangerous impressions ($M = 21.63$, $SD = 13.17$) was considerably

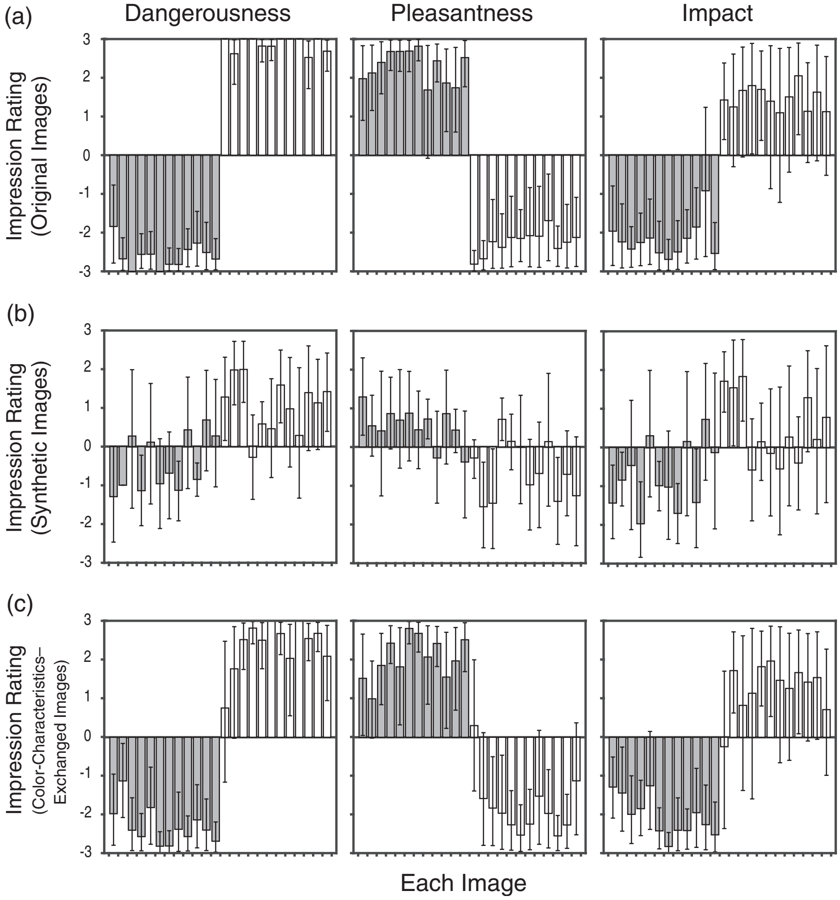

Figure 1 Average of the impression ratings across participants for each image. Error bars represent 95% credible intervals. Each bar in each panel corresponds to each image. Shaded and white bars represent negative and positive images, respectively. The original image numbers in the IAPS are, from left to right: 1,050, 2,683, 2,811, 3,500, 6,520, 6,550, 6,560, 8,480, 9,414, 9,902, 9,908, 9,940, 1,440, 1,441, 1,630, 1,710, 1,920, 2,347, 5,199, 5,760, 5,780, 5,825, 7,325, and 7,492.

lower than that of the images with safe impression (M = 26.26, SD = 9.06).

Although the difference in mean saturation did not reach significance, this might be due to the small sample size (N = 24) or to the inappropriate statistical testing. We considered it more appropriate to use a statistical procedure that can take full advantage of the continuous variables. Figure 2b shows the relationship between the mean saturation of each original color image and the 50% thresholds calculated by each image. We calculated Spearman's rank-correlation coefficient, as there was some non-linearity in the scatter plot. We found a moderate negative correlation between the mean saturations of the original color images and the 50% thresholds, rho = −.430, n = 22, p = .047. Note that we excluded image numbers 2,683 and 9,908 in the IAPS, which did not fit to a psychometric function. Their monochrome versions were detected in less than 50% of trials even for the longest duration (60 ms).

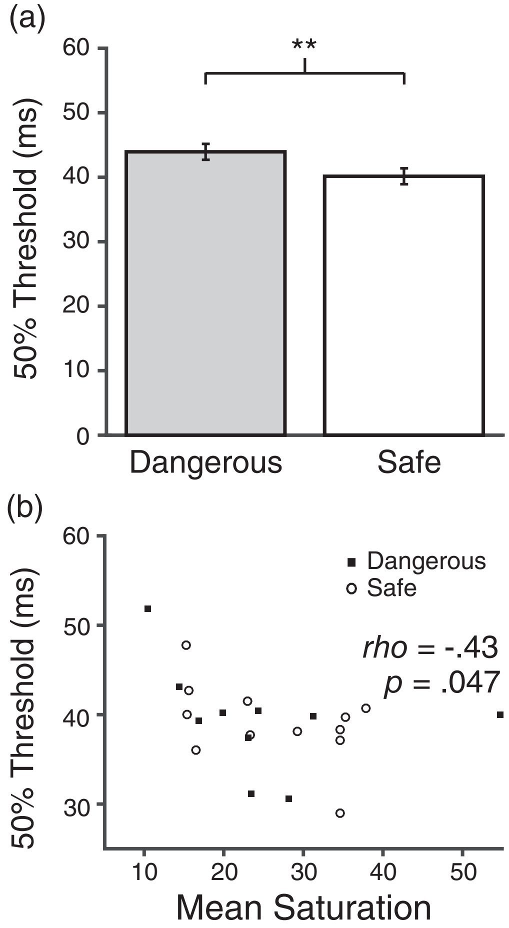

Figure 2 Results of Experiment 1. (a) Average of the 50% thresholds of all the participants for both conditions. Error bars represent 95% within-participant confidence intervals (Morey, 2008). (b) Correlation between the mean saturations of the original color images and the 50% thresholds. **p < .01.

The observed non-linearity can explain the significant difference in the 50% thresholds with absence of significant difference of the mean saturations. The 50% thresholds showed a sharp rise in inverse proportion to the mean saturations at a mean saturation of around 15 or lower, and such abrupt increase was not observed with the other mean saturations. Images with dangerous impressions contained some extremely desaturated images (image numbers 2,683 and 9,908 in the IAPS had mean saturations of 7.19 and 5.88, respectively; the average of the mean saturations of all images was 23.94 ± 11.07). These images thus pulled up the average of the 50% thresholds for the images with dangerous impressions even though there was no significant difference in mean saturation at the image group level.

We have no clear answer as to why our results showed the trend opposite to that obtained in Kobayashi and Ichikawa (2016). They excluded the data from three participants because their data did not fit to a psychometric function. The frequencies at which these participants detected the monochrome pictures were higher than 50% even for the shortest duration (10 ms). Such accuracies are much higher than ordinal behavioral accuracies with visual backward masking (Bacon-Macé, Macé, Fabre-Thorpe, & Thorpe, 2005), and higher than accuracies obtained in our experiment (the average detection accuracy across participants in our experiment was 3.9 ± 3.7%). Such high performances are likely due to dropping frames. The generally low 50% thresholds reported by Kobayashi and Ichikawa raises questions about the reliability of their stimulus presentation.

Experiment 2

In Experiment 1, we revealed a correlation between color saturation and detection thresholds. Therefore, we hypothesized that the results from Kobayashi and Ichikawa (2016) were due to the confounding effects of color saturation, and not based on dangerous impressions of the images. This hypothesis led to the prediction that the overall trend of the results would remain unchanged even when the impressions were reduced. To test this prediction, we used a texture synthesis algorithm proposed by Portilla and Simoncelli (2000) to reduce the impressions of the images while retaining their image statistics.

Methods

Figure 3a,b shows an example of texture synthesis. Using a texture synthesis algorithm proposed by Portilla and Simoncelli (2000), 24 images were synthesized from the same images as in Experiment 1.

The pixel intensity distributions of the original images are imposed on the images synthesized by their algorithm and the new images thus retain their original mean saturations. They also retain higher-order statistics, such as auto-correlation of magnitude of each subband.

While images synthesized by their algorithm retain their original image statistics, their global shapes are destroyed (Banno & Saiki, 2015). A previous study indicates that degraded availability of global shape information makes object recognition difficult (Rokszin et al., 2015). We thus expected that the impressions of the resultant synthetic images would be reduced relative to those of the original images because of difficulty in object recognition.

The experimental procedure was the same as that in Experiment 1.

Results and Discussion

As expected, the participants' impressions of the images decreased in the measures of dangerousness, pleasantness, and impact (Figure 1b).

The average false alarm rate was below 5% in both conditions: 2.7% (the synthetic images that previously gave dangerous impressions) and 4.5% (the synthetic images that previously gave safe impressions). A paired *t* test revealed no significant difference in false-alarm rates between the two image groups, $t(6) = -1.035$, $d_{Diff} = 0.423$ (Cohen's $d = 0.312$), $p = .341$. The subsequent analyses were restricted to the trials in which the monochrome image was presented.

A paired *t* test revealed that the 50% thresholds for the synthetic images that previously gave dangerous impressions were still significantly higher than those for the synthetic images that previously gave safe impression (Figure 4a), $t(6) = 6.389$, $d_{Diff} = 2.608$ (Cohen's $d = 0.959$), $p < .001$. That is, the overall trend of the results remained unchanged from that of Experiment 1 despite reducing the dangerous impressions of the images while retaining their image statistics. This fact suggests that the effect of dangerous impressions on temporal resolution of visual perception is not critical, and thus supports our hypothesis.

There was a marginally significant negative correlation between the mean saturations of the synthetic color images and the 50% thresholds (Figure 4b), $rho = -.405$, $n = 23$, $p = .057$, in line with that observed in Experiment 1. Note that we excluded the image synthesized from the original image number 9,908 in the IAPS, which did not fit to a psychometric function. Their monochrome versions were detected in less than 50% of trials even for the longest duration (60 ms).

Experiment 3

We hypothesized that the results from Kobayashi and Ichikawa (2016) were due to the confounding effects of color saturation, and not based on dangerous impressions of the images. This hypothesis led to another prediction that the overall trend of the results would be reversed if the color characteristics were exchanged between the two image groups. To test this prediction, we used a color transfer algorithm proposed by Xiao and Ma (2006). This algorithm enabled us to exchange the color characteristics between the two image groups while keeping their impressions nearly unchanged.

Methods

Figure 3c shows an example of color transfer. Using a color transfer algorithm proposed by Xiao and Ma (2006), we exchanged the color characteristics between each of the original images with dangerous impressions and each of those with safe impressions. The color characteristics of the original image numbers 1,050, 2,683, 2,811, 3,500, 6,520, 6,550, 6,560, 8,480, 9,414, 9,902, 9,908, and 9,940 were exchanged for those of the original image numbers 1,440, 1,441, 1,630, 1,710, 1,920, 2,347, 5,199, 5,760, 5,780, 5,825, 7,325, and 7,492, respectively. This generated 24 color-characteristics-exchanged images.

The experimental procedure was identical with that of Experiment 1.

Results and Discussion

The participants' impressions of the images remained nearly unchanged from those of the original images in the measures of dangerousness, pleasantness, and impact (Figure 1c).

The average false alarm rate was below 5% in both conditions: 2.1% (the images with dangerous impressions) and 3.0% (the images with safe impressions). A paired *t* test revealed no significant difference in false-alarm rates between the two image groups, $t(6) = -1.000$, $d_{Diff} = 0.409$ (Cohen's $d = 0.362$), $p = .355$. The subsequent analyses were restricted to the trials in which the monochrome image was presented.

A paired *t* test revealed that the 50% thresholds for the images with dangerous impressions were significantly lower than those for the images with safe impressions (Figure 5a), $t(6) = -7.202$, $d_{Diff} = 2.940$ (Cohen's $d = 1.315$), $p < .001$. That is, the overall trend of the results was reversed by exchanging the color characteristics between the two image groups while keeping the impressions of danger nearly unchanged. This suggests that temporal resolution measurements are strongly influenced by the color characteristics of the images in this experimental paradigm, and thus supports our hypothesis.

There was a moderate negative correlation between the mean saturations of the color-characteristics-exchanged images and the 50% thresholds (Figure 5b), $rho = -.340$, $n = 22$, $p = .122$, in line with that observed in Experiments 1 and 2. Note that we excluded the image generated from the original image numbers 1,441 and 7,325 in the IAPS whose color characteristics were replaced by that of the original image numbers 2,683 and 9,908, respectively, because the performances did not fit to a

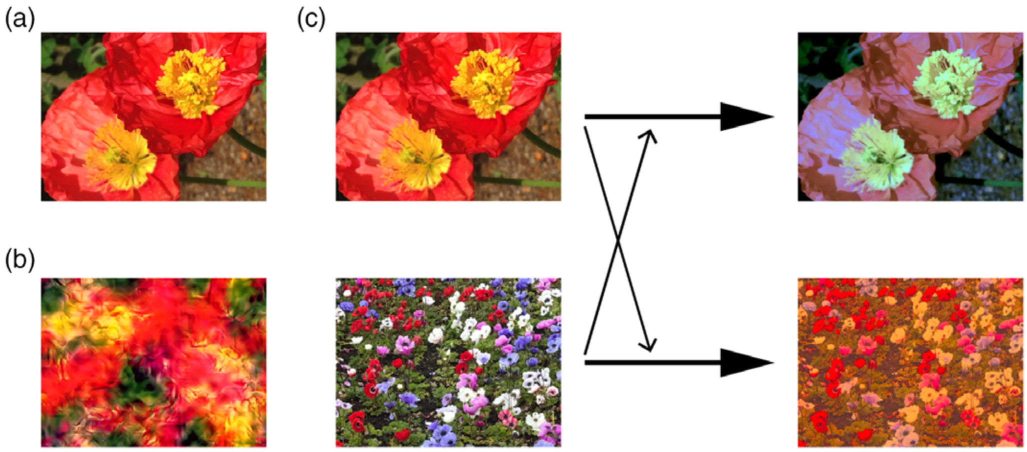

Figure 3 Example stimuli. (a) Original image. (b) Image synthesized from the original image (Experiment 2). (c) Color-characteristics-exchanged images (Experiment 3). Note that these images were not taken from the IAPS for copyright reasons.

psychometric function. Their monochrome versions were detected in less than 50% of trials even for the longest duration (60 ms).

Importantly, we consistently observed the correlations between the mean saturations of the color images and the 50% thresholds throughout the experiments. This fact adds further evidence that the saturation cues of the color images led to superficially increased temporal resolution in this experimental paradigm.

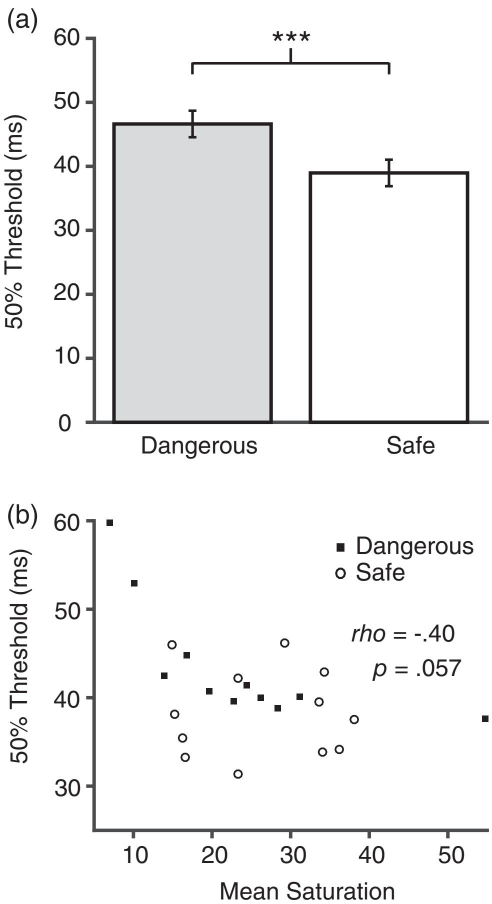

Figure 4 Results of Experiment 2. (a) Average of the 50% thresholds of all participants for both conditions. Error bars represent 95% within-participant confidence intervals (Morey, 2008). (b) Correlation between the mean saturations of the synthetic color images and the 50% thresholds. ***$p < .001$.

General Discussion

We clarified the confounding effects of color saturation on increased temporal resolution reported by Kobayashi and Ichikawa (2016). There were two main findings. First, the more saturated an image was, the higher the temporal resolution obtained. Second, temporal resolution of visual perception was affected by the color characteristics of the images rather than their dangerous impressions. These findings suggest

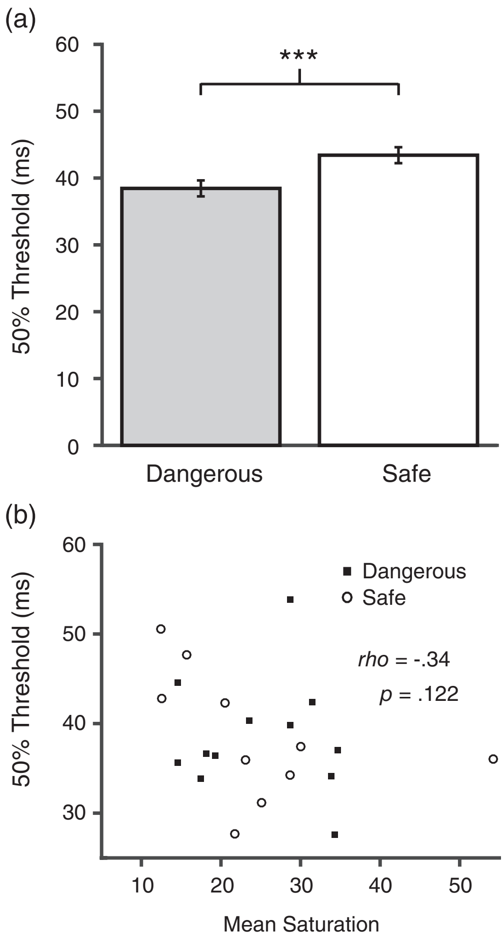

Figure 5 Results of Experiment 3. (a) Average of the 50% thresholds of all the participants for both conditions. Error bars represent 95% within-participant confidence intervals (Morey, 2008). (b) Correlation between the mean saturations of the color-characteristics-exchanged images and the 50% thresholds. ***$p < .001$.

that the previously reported increase in temporal resolution while viewing highly emotional images was not actually experienced. Rather, this superficial improvement was likely due to the confounding effect of color saturation, and not the emotion evoked by such images.

We caution against using natural images to study the effect of emotion on temporal resolution of visual perception for two reasons. First, a previous study suggested that subsets of pictures selected from a certain stimulus database tend to be biased in terms of their physical properties (Delplanque et al., 2007), and it is difficult to completely remove the artifacts from image statistics. Instead, we suggest using other visual stimuli that are defined only by luminance, such as simple flashes, and obtaining flash fusion thresholds as a measure of temporal resolution of visual perception. Second, affective pictures do not seem sufficient to offer "a life-threatening condition." Life-threatening situations are unlikely to be satisfied by visual stimuli alone, and may require more inputs from other senses, such as somatic senses. From the viewpoint of ecological validity, we propose that it would be useful to design an experimental paradigm that offers a powerful multisensory experience, as in the free-fall experiment conducted by Stetson et al. (2007).

References

Bacon-Macé, N., Macé, M. J. M., Fabre-Thorpe, M., & Thorpe, S. J. (2005). The time course of visual processing: Backward masking and natural scene categorisation. *Vision Research*, *45*, 1459–1469. https://doi.org/10.1016/j.visres.2005.01.004

Banno, H., & Saiki, J. (2015). The use of higher-order statistics in rapid object categorization in natural scenes. *Journal of Vision*, *15*, 1–20. https://doi.org/10.1167/15.2.4

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 433–436. https://doi.org/10.1163/156856897X00357

Delplanque, S., N'diaye, K., Scherer, K., & Grandjean, D. (2007). Spatial frequencies or emotional effects? A systematic measure of spatial frequencies for IAPS pictures by a discrete wavelet analysis. *Journal of Neuroscience Methods*, *165*, 144–150. https://doi.org/10.1016/j.jneumeth.2007.05.030

Kobayashi, M., & Ichikawa, M. (2016). Emotions evoked by viewing pictures may affect temporal aspects of visual processing. *Japanese Psychological Research*, *58*, 273–283. https://doi.org/10.1111/jpr.12125

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2008). *International affective picture system (IAPS): Affective ratings of pictures and instruction manual.* (Technical Report A-8). University of Florida, Gainesville, FL.

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorials in Quantitative Methods for Psychology*, *4*, 61–64. https://doi.org/10.20982/tqmp.04.2.p061

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442. https://doi.org/10.1163/156856897X00366

Portilla, J., & Simoncelli, E. P. (2000). Parametric texture model based on joint statistics of complex wavelet coefficients. *International Journal of Computer Vision*, *40*, 49–71. https://doi.org/10.1023/A:1026553619983

Rokszin, A. A., Gyori-Dani, D., Linnert, S., Krajcsi, A., Tompa, T., & Csifcsák, G. (2015). The interplay of holistic shape, local feature and color information in object categorization. *Biological Psychology*, *109*, 120–131. https://doi.org/10.1016/j.biopsycho.2015.05.002

Stetson, C., Fiesta, M. P., & Eagleman, D. M. (2007). Does time really slow down during a frightening event? *PLoS One*, *2*(12), e1295. https://doi.org/10.1371/journal.pone.0001295

Xiao, X., & Ma, L. (2006). Color transfer in correlated color space. *Proceedings of the 2006 ACM International Conference on Virtual Reality Continuum and Its Applications*, 13, 305–309. https://doi.org/10.1145/1128923.1128974

(Received July 20, 2017; accepted December 7, 2017)