NeuroImage 106 (2015) 21–33

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Affective processing in natural scene viewing: Valence and arousal interactions in eye-fixation-related potentials

Jaana Simola [a,b,c,*], Kevin Le Fevre [a,b], Jari Torniainen [c], Thierry Baccino [a,b]

[a] Laboratoire Vision Action Cognition (VAC), Université Paris Descartes – Institut de Psychologie, France
[b] Chart-Lutin, Université Paris8, Saint-Denis, France
[c] Cognitive Science & Cognitive Brain Research Unit (CBRU), University of Helsinki, Finland

ARTICLE INFO

Article history:
Accepted 13 November 2014
Available online 21 November 2014

Keywords:
emotion
attention
emotional salience
eye movements
EEG
eye-fixation-related potentials

ABSTRACT

Attention is drawn to emotionally salient stimuli. The present study investigates processing of emotionally salient regions during free viewing of emotional scenes that were categorized according to the two-dimensional model comprising of valence (unpleasant, pleasant) and arousal (high, low). Recent studies have reported interactions between these dimensions, indicative of stimulus-evoked approach or withdrawal tendencies. We addressed the valence and arousal effects when emotional items were embedded in complex real-world scenes by analyzing both eye movement behavior and eye-fixation-related potentials (EFRPs) time-locked to the critical event of fixating the emotionally salient items for the first time. Both data sets showed an interaction between the valence and arousal dimensions. First, the fixation rates and gaze durations on emotionally salient regions were enhanced for unpleasant versus pleasant images in the high arousal condition. In the low arousal condition, both measures were enhanced for pleasant versus unpleasant images. Second, the EFRP results at 140–170 ms [P2] over the central site showed stronger responses for high versus low arousing images in the unpleasant condition. In addition, the parietal LPP responses at 400–500 ms post-fixation were enhanced for stimuli reflecting congruent stimulus dimensions, that is, stronger responses for high versus low arousing images in the unpleasant condition and stronger responses for low versus high arousing images in the pleasant condition. The present findings support the interactive two-dimensional approach, according to which the integration of valence and arousal recruits brain regions associated with action tendencies of approach or withdrawal.

Introduction

Real-world scenes include several salient features. Visual attention allocates foveal resources in order to select the most relevant information by prioritizing the processing of certain parts of the incoming information. Stimulus-driven ("bottom-up") models of visual attention define saliency in terms of low-level properties, such as color, orientation and luminance, which are then integrated to produce a saliency map (e.g., Itti and Koch, 2000). The maximum of a saliency map corresponds to the attended location. Previous research also shows that higher level ("top-down") factors, such as the semantic content of the images, co-occurrence of objects, and task constraints play a key role in attention allocation (De Graef, 2005; Henderson et al., 1999; Tatler et al., 2011; Torralba et al., 2006; Yarbus, 1967). Affective salience creates yet another source of bias on perceptual processing. The tendency of an item to stand out relative to its neighbors due to an association between its emotional meaning provides a means to select sensory information rapidly and efficiently (Humphrey et al., 2012; Niu et al., 2012; Pourtois et al., 2013).

Detecting and reacting to emotional information, for example, avoiding harmful events and approaching pleasant events, is essential for survival. Therefore, affective stimuli have a processing advantage over neutral information. This finding has been demonstrated in several studies. Behavioral studies have used visual search paradigms to show that emotional stimuli, in particular, fear-relevant information is detected faster than neutral information (e.g., Blanchette, 2006; Flykt, 2005; Fox et al., 2007; Öhman et al., 2001). Previous eye movement research further indicates that when an emotional image is presented simultaneously with neutral images, the probability of initially fixating the emotional stimulus is higher (Calvo and Lang, 2004; Nummenmaa et al., 2006). Moreover, in a spatial cuing task, saccade reaction times were shorter when an emotional picture rather than neutral picture was cued, while more saccadic errors occurred when the cue pointed toward neutral than toward emotional picture (Nummenmaa et al., 2009).

These studies have successfully revealed preferential processing of emotional versus neutral information, but they all use relatively

* Corresponding author at: University of Helsinki, Cognitive Science & Cognitive Brain Research Unit (CBRU), Siltavuorenpenger 1B, FI-00014 University of Helsinki, Finland. Tel.: +33 7 82 30 51 82 (mobile).
E-mail addresses: jaana.simola@helsinki.fi (J. Simola), kevin.le-fevre02@univ-paris8.fr (K. Le Fevre), jari.torniainen@aalto.fi (J. Torniainen), thierry.baccino@univ-paris8.fr (T. Baccino).

http://dx.doi.org/10.1016/j.neuroimage.2014.11.030
1053-8119/

unnaturalistic viewing conditions. The visual search paradigms, for example, contain a small number of independent images (e.g., four images presented in a 2 x 2 matrix). Because the images are presented independently, their contents are unrelated to their location, unlike in real-world scenes, where the content of an item mostly determines its location (e.g., objects do not float in the air). Moreover, presenting the emotional stimuli in predetermined locations may increase the expectation of emotional content at these locations. Further, the task conditions, such as visual search does not directly measure the affect-biased attention as a form of natural viewing behavior involving pre-tuning to specific stimulus categories, since the participants are often explicitly instructed to attend or ignore the emotional stimuli. Based on previous studies, thus, it is unclear whether the emotional effects remain under more natural conditions where the emotional items are embedded within crowded scenes and in locations determined by the scene context and when the capacity limited attentional system selects which objects are given the processing priority (however see Acunzo and Henderson, 2011).

Prior electrophysiological research also shows preferential processing of emotional content (reviewed in Olofsson et al., 2008). These studies present the images typically one at a time, indicating how the stimuli are encoded when they are presented singly. Further, when the neural responses are time-locked to scene onset, it is hard to disentangle how the information extracted from the full range of individual fixations contributes to the response. The relatively novel technique of simultaneous recording of eye movements and EEG provides an efficient tool to investigate attention processes under naturalistic viewing (e.g., Baccino, 2011; Dimigen et al., 2011). This approach has several important advantages. First, eye movements are part of our attentional system, which largely determines what information is entering the visual system. The eye movement events, thus, constitute natural markers to segment the ongoing neural activity, and the analysis of eye-fixation-related potentials (EFRPs) allows for investigation of neural activity during self-paced perceptual and cognitive behavior. The co-registration technique provides also a high temporal resolution to investigate the time course of perceptual and cognitive processes, which eventually lead to observable behavior.

Previous research indicates that the co-registration technique is a valid technique to investigate attention to emotional stimuli. Simola et al. (2013) showed that even though emotional information is highly salient and therefore also processed in parafoveal/peripheral vision (Calvo and Lang, 2005; Coy and Hutton, 2012; Nummenmaa et al., 2009) such preview effects did not confound the analysis of brain responses time-locked to the point when the eyes had landed on the emotional stimulus. In order to validate the co-registration technique during an emotional scene perception task, Simola et al. compared the responses recorded during free viewing to ERP responses to serially presented emotional images. Their results showed similar responses across the presentation conditions, indicating that EFRPs can be reliably recorded in an emotional attention task. Also other researchers have successfully applied the EFRP method during a scene perception task, for example, when the task was to find a target face among a crowded scene containing multiple faces (Kaunitz et al., 2014). Intracranial EEG (iEEG) research has further established that category-specific visual areas (i.e., the fusiform face area, FFA for faces and the visual word form area, VWFA for words) maintain their selectivity during free-viewing of crowded natural scenes, such that the specialized visual areas take turns in processing distinct elements of the scene depending on the current location of the gaze (Hamamé et al., 2014). Intracranial EEG was also utilized by Nagasawa et al. (2011) to show saccade related augmentation of gamma activity in the medial occipital region during an active viewing task involving eye movements.

The present study used the EFRP-approach in order to understand how attention is allocated to process emotional information during natural scene viewing. The underlying theoretical assumption was the division of emotions into two dimensions: valence (pleasant–unpleasant) and arousal (calm–highly arousing) (Bradley et al., 2001). Both factors are assumed to elicit activation of two underlying systems: the approach and withdrawal systems. The approach system is activated in contexts promoting survival, including sustenance, procreation, and nurturance, with the basic behaviors of ingestion, copulation, and caregiving. Conversely, the withdrawal system is activated when the context involves threat, with the basic behavioral repertoire of withdrawal, escape, and attack. The judgment of valence indicates which motivational system is activated, while arousal indicates the intensity of the activation of the given system. High arousal stimuli activate the motivational systems more strongly than low arousing stimuli, because they convey information that is relevant for immediate survival.

A number of studies assume independence of valence and arousal (reviewed in Olofsson et al., 2008). However, Robinson et al. (2004) propose a model that predicts approach or withdrawal tendencies for low and high arousal stimuli independently of whether the stimulus is positively or negatively valenced. According to their model, arousal, no matter whether positive or negative, is negatively coded, because high arousing stimuli are associated with potential danger or threat. Therefore, high arousal elicits withdrawal and low arousal elicits approach tendencies, suggesting that the unpleasant high-arousing and pleasant low-arousing stimuli are congruent in their valence and arousal dimensions, and that the unpleasant low-arousing and pleasant high-arousing stimuli are incongruent. Robinson et al. found that participants evaluated congruent stimuli faster than incongruent stimuli. An interactive influence on affective processing has also been observed without an explicit emotion evaluation goal, suggesting that arousal and valence are coded and integrated indirectly and that the interaction effect is not based on specific response strategies (Eder and Rothermund, 2010). Previous research using the ERP technique also indicates enhanced amplitudes for congruent versus incongruent stimuli (Feng et al., 2012). These enhancements occurred within both early and late stages of emotional processing, as reflected in increased amplitudes for the P2a (100–200 ms) and P3 (300–400 ms) responses.

Based on previous research (Eder and Rothermund, 2010; Feng et al., 2012; Robinson et al., 2004), we expected that after the emotionally salient items were overtly attended, valence and arousal would interactively modulate the eye fixation-related-potentials (EFRPs) at different stages of processing. Consistent with previous studies, the congruent conditions were expected to influence the EFRP responses more strongly than the incongruent conditions. Results on these lines would indicate that valence and arousal affect emotional processing interactively also under naturalistic viewing conditions where the emotional items are embedded in natural scene context. To our knowledge, no earlier study has used co-registration of eye movements and EEG to examine the effects of valence and arousal on the processing of emotionally salient items when these items are part of a natural scene.

Material and methods

Participants

Eighteen (11 female, 19-49 years, mean age 26 years) volunteers participated in the experiment. The required sample size for F tests (repeated measures ANOVA, within-subjects factors) was estimated by a power analysis (GPower 3.1.7) (Faul et al., 2007). Using a full within-factors design (rmANOVAs, see below for the design) and 185 trials, the power analysis predicted 4 participants to be enough for reaching a significance level of $p = .05$ (power = .95; effect size = .25). Thus, the current sample was well above the sample size required. All participants were right handed and had normal or corrected to normal vision. Before the experiment, each participant provided a written informed consent according to the recommendations of the declaration of Helsinki. At the end, the participants were rewarded with a 20 € gift card for participating.