Infant Behavior & Development 36 (2013) 32–41

Contents lists available at SciVerse ScienceDirect

Infant Behavior and Development

Where and how infants look: The development of scan paths and fixations in face perception

Masaharu Kato[a,*], Yukuo Konishi[b]

[a] NTT Communication Science Laboratories, NTT Corporation, Japan
[b] Center for Baby Science, Doshisha University, Japan

ARTICLE INFO

Article history:
Received 11 July 2012
Received in revised form
21 September 2012
Accepted 22 October 2012

Keywords:
Face perception
Scan path sequence
Configural process

ABSTRACT

It is well known that infants prefer faces but relatively little is known about how infants look at face. The present work examined the development of face perception by using the scan paths. Infants (aged 6–13.5 months) and adults were presented with images of upright and inverted faces, and looking times and scan paths were compared. Similarity between participants' scan paths demonstrated that infants collect facial information more efficiently from upright faces than from inverted faces, and this ability gradually develops with age. Analyses of looking times also revealed that preferences for upright faces were gradually replaced by preferences for inverted faces by approximately 10 months of age. The results indicate that the processing of configural information gradually develops throughout infancy.

1. Introduction

Research suggests that infants can find information from facial images (see Pascalis & Slater, 2003, for a review); however, questions remain as to how infants find information from a face. Knowledge about where in the face infants' look, and how they look, should help to address these questions. An early study revealed that infants as young as 3–5 weeks mostly scan the edges of a face and begin to scan the eyes by 9–11 weeks (Haith, Bergman, & Moore, 1977). Two-month-old infants also prefer to orient toward the eyes (Maurer & Salapatek, 1976). However, as far as we know, detailed age-related studies examining the development of facial processing have yet to be conducted. Thus, this paper examined developmental changes of detailed looking times and scan paths during face processing in developing infants.

Where and how infants look is influenced by both top-down cognitive operations, such as perceptual sets toward images, and bottom-up sensory factors, such as local contrast, color, and orientation. Bottom-up factors are referred to as saliency effects. Koch and Ullman (1985) introduced the construct of a saliency map that is expressed by figures in 2D space. These maps suggest that highly salient areas automatically attract an observer's attention. Given that eyes tend to move serially from high to low salient areas, saliency maps have been used to predict scan paths. These maps have explained—to a certain degree—the paths by which scenes are viewed (Foulsham & Underwood, 2008; Itti & Koch, 2001).

To determine the cognitive meaning of detailed looking times and scan paths, the effect of saliency on the looking times and the scan paths should be excluded. One method for accomplishing this goal is to prepare both upright and inverted facial images and compare detailed processing of those images. This is an advantageous method given that both images are identical in terms of saliency, while they differ in terms of cognitive meaning.

* Corresponding author.
E-mail address: kato.masaharu@lab.ntt.co.jp (M. Kato).

0163-6383/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.infbeh.2012.10.005

Facial recognition is less accurate when a face is presented in an inverted position (Yin, 1969). A related phenomenon is referred to as the Thatcher effect (Thompson, 1980): here, a face with vertically flipped eyes and mouth are quite grotesque when the face is upright but not so grotesque when the face is inverted. In both cases, information based on spatial relations among facial regions (i.e., eyes, nose, mouth, ear, chin, etc.) is corrupted by inversion; this corruption results in the Thatcher effect (Farah, Tanaka, & Drain, 1995; Freire, Lee, & Symons, 2000; Rhodes, Brake, & Atkinson, 1993). Developmentally, humans can make use of this orientation-specific configuration information (referred to as configural cues) at an early age. Four-month-old infants can recognize an upright face but cannot recognize an inverted one (Turati, Sangrigoli, Ruely, & Schonen, 2004). They also look longer toward internal facial features when the pictures are upright (Gallay, Baudouin, Durand, Lemoine, & Lecuyer, 2006). However, children are still fooled by manipulating facial expressions and paraphernalia (e.g., hair-styles and eyeglasses) when identifying faces until they are 10 years old. This suggests that their ability to make use of configural cues is still developing (Carey & Diamond, 1977). A similar conclusion can be made by using a set of faces, where each face differs only in the spacing of the eyes and mouth (Mondloch, Le Grand, & Maurer, 2002). They presented those images in upright and inverted positions to find that the inversion effect did not decrease even in adults, while the effect diminished if a face set differed each other in terms of facial features. Thus, comparing infants' scan paths when viewing upright and inverted faces should be helpful in identifying the development of scan path sequences essential for successfully perceiving configural cues.

To examine scan paths during face perception, a working hypothesis was developed; if facial processing is efficient, facial information should be detected with limited fixation points and scan paths. This hypothesis is derived from studies investigating expertise in collecting visual information and eye movements. For example, expert chess players make fewer fixations than less-skilled players during a check detection task (Reingold, Charness, Pomplun, & Stampe, 2001). Expert pilots have better defined and compact eye scan patterns; however, novice pilots exhibit more complicated eye scan patterns under a visual flight rules flight condition (i.e., a condition where it is clear enough for a pilot to guide an aircraft with reference to the ground; Kasarskis, Stehwien, Hickox, Aretz, & Wickens, 2001). Moreover, video game players given training on efficient eye scan practices (in order to minimize eye movements and optimize scan paths) performed better on a video game as compared to both participants who were given inefficient eye scan training and control participants (Shapiro & Raymond, 1989). Although it is not know beforehand what scan path is the most efficient for collecting facial information, it is reasonable to assume that participants' scan paths simplify to the most efficient path. This suggests that infants' scan paths come to resemble each other as a function of age and orientation of the image.

To quantify similarity of scan paths, a similarity index was introduced. The string edit method is a widely used method that estimates the similarity between two sequences based on the editing cost, transforming one sequence into the other with three operations (insertion, deletion, and substitution; Levenshtein, 1966). The string edit and derivative methods (e.g., Okuda, Tanaka, & Kasai, 1976; Wagner & Lowrance, 1975) have been used in several studies assessing scan paths (e.g., Brandt & Stark, 1997). However, the string edit method has a fundamental problem: one sequence is transformed into the others, which means that the final absolute position of every fixation in both sequences is the same. Since it is not known when cognitive processes begin during face presentation, and begin to appear on the scan paths, it is reasonable to transform one sequence in reference to the other sequence in order to find which part of the sequence matches, irrespective of the absolute position in the sequence. Thus, for the current study, we derived a method that compares sequences relative to each other (the details of the algorithm are documented in Section 2).

2. Materials and methods

2.1. Participants

Fifty-one healthy, full-term infants participated in the experiment. Eleven were excluded from the analyses due to fussiness, or because they did not meet our criteria of looking times (viewing stimuli for at least 5 s). The remaining 40 infants were divided into four groups according to age: 6 months (188 ± 31 [mean $\pm$ SD] days), 8.5 months (262 ± 20 days), 11 months (336 ± 31 days), and 13.5 months (413 ± 17 days). Each group consisted of five males and five females. Ten adult participants (four females and six males, 20.2 ± 1.5 years) also participated in the experiment as a control group.

The study was conducted in accordance with ethical standards specified in the 1964 Declaration of Helsinki and approved by the ethics committee to the authors' respective organizations. Adult participants and the families of participating infants were informed about the purpose of the study, and the adult participants and parents of infants signed consent forms. Infants and their parents received a token worth 1000 Japanese yen, and adult participants received 1000 Japanese yen for their participation in our experiment.

2.2. Apparatus and stimuli

The experiment was conducted in a dark, quiet room. Visual stimuli were presented on a 17-in. LCD display. Just below the display, a cornea-reflection eye-tracking system (120×, Tobii Corporation, Stockholm, Sweden), which measures the direction of each eye, separately, at 60 Hz was attached to the monitor by a hand-made aluminum basket. The distance between the participants and the display was approximately 60 cm.