Gaze behavior during 3-D face identification is depth cue invariant

Hassan Akhavein

Reza Farivar

Gaze behavior during scene and object recognition can highlight the relevant information for a task. For example, salience maps—highlighting regions that have heightened luminance, contrast, color, etc. in a scene—can be used to predict gaze targets. Certain tasks, such as face recognition, result in a typical pattern of fixations on high salience features. While local salience of a 2-D feature may contribute to gaze behavior and object recognition, we are perfectly capable of recognizing objects from 3-D depth cues devoid of meaningful 2-D features. Faces can be recognized from pure texture, binocular disparity, or structure-from-motion displays (Dehmoobadsharifabadi & Farivar, 2016; Farivar, Blanke, & Chaudhuri, 2009; Liu, Collin, Farivar, & Chaudhuri, 2005), and yet these displays are devoid of local salient 2-D features. We therefore sought to determine whether gaze behavior is driven by an underlying 3-D representation that is depth-cue invariant or depth-cue specific. By using a face identification task comprising morphs of 3-D facial surfaces, we were able to measure identification thresholds and thereby equate for task difficulty across different depth cues. We found that gaze behavior for faces defined by shading and texture cues was highly comparable, but we observed some deviations for faces defined by binocular disparity. Interestingly, we found no effect of task difficulty on gaze behavior. The results are discussed in the context of depth-cue invariant representations for facial surfaces, with gaze behavior being constrained by low-level limits of depth extraction from specific cues such as binocular disparity.

Introduction

The inspection of a visual object is accompanied by complex gaze behavior that has proven difficult to predict (Jovancevic, Sullivan, & Hayhoe, 2006; Turano, Geruschat, & Baker, 2003). Aspects of this gaze behavior are driven by stimulus attributes, but task demands can also drastically influence gaze behavior (Navalpakkam & Itti, 2005). Stimulus attributes include low-level features such as brightness, contrast, color, or spatial frequency, but may also include high-level information such as depth cues. The perception of depth can have the effect of over-riding some of the salient 2-D cues. For example, the addition of perspective information during the viewing of a simple geometric capsule has the effect of biasing gaze towards the center of gravity for the object in 3-D, instead of the center of the 2-D projected form (Vishwanath & Kowler, 2004). In more complex tasks such as face perception, there is evidence that the addition of stereopsis can influence gaze behavior (Chelnokova & Laeng, 2011), but it is difficult to determine whether depth per se changed eye movement behavior or depth simplified the task, resulting in a change in inspection strategy.

The addition of depth information can influence complex object recognition. For example, face recognition improves with the addition of binocular disparity information (Liu & Ward, 2006). We have shown previously that a number of depths cues, such as structure-from-motion (SFM) and texture (Farivar et al., 2009; Liu et al., 2005), can alone serve to empower complex object recognition. Interestingly, not all depth cues are processed by similar cortical mechanisms. For
example, stereopsis and SFM are processed largely by the dorsal visual pathway (Backus, Fleet, Parker, & Heeger, 2001; Heeger, Boynton, Demb, Seidemann, & Newsome, 1999; Kamitani & Tong, 2006; Tootell et al., 1995; Tsao et al., 2003; Zeki et al., 1991), a pathway that begins in V1 and courses dorsally through area MT before terminating in the posterior parietal lobe, while texture and shading largely depend on the ventral visual pathway (Merigan, 2000; Merigan & Pham, 1998), a pathway that also begins in V1, but courses ventrally through V4 and terminates in the anterior temporal lobe.

In displays where a single depth cue is used to render an object, low-level cues such as luminance and contrast are unavailable, weak, or unrelated to the surface of the object. This means that saliency models such as described by (Itti & Koch, 2000) would poorly predict gaze behavior during 3-D object recognition, independent of task demands, simply because low-level features in such displays are unrelated to depth. Therefore, gaze behavior in such displays must be driven primarily by the perception of depth and not the manner in which depth is rendered—the depth cue ought to be unimportant. We call this the depth-cue invariance hypothesis—that gaze behavior is driven by the depth profile of an object, independent of low-level features or task demands. The key prediction of this view is that the gaze behavior for a given object recognition task would remain constant across different depth cues. An alternative hypothesis is that gaze performance is not invariant to depth cue, but tasks using each depth cue create different demands for gaze, resulting in a distinct pattern of gaze behavior that depends on the depth cue of the stimulus.

Distinguishing between these two hypotheses is challenging, because it requires comparison of gaze performance across different depth cues in a manner that separates the influence of task demands. While objects can be defined by different depth cues, recognition performance across depth cues is not consistent. For example, face recognition performance with SFM on a 1:8 matching task is approximately 50% (four times above chance), while the same task with shaded faces results in performance close to 85% (Farivar et al., 2009). This means that we cannot directly compare gaze performance across depth for such a task—the gaze performance difference could be due to the difference in the task difficulty, independent of the depth cue itself (Dehmoobasadsharifabadi & Farivar, 2016).

In the studies described below, we have sought to overcome this challenge and measure gaze behavior during face recognition, where faces are defined by different depth cues. We were able to equate task difficulty across the different depth cues by measuring face identification thresholds—the amount of morph between a given facial surface and the “average” facial surface needed to result in correct performance. This allowed us to equate task difficulty across depth cues and therefore directly compare how gaze performance was influenced by depth cues for similar levels of performance on each depth cue.

**Methods**

**Participants**

Nine subjects (including one of the authors, HA) participated in the study. All participants (aged 23 to 38; two women, seven men) had normal or corrected-to-normal vision. All procedures were approved by the Research Ethics Board of the Montreal Neurological Institute. Two subjects completed all but the stereo task.

**Stimuli and display**

The stimuli consisted of frontal view of faces with different 3-D surface information (Shading, Texture, and Stereo). Facial surfaces with four different identities (two male and two female) were generated using FaceGen Modeller 3.5. Each identity was morphed from 100% towards the average face in steps of 10% using the morph feature in the software. Facial surfaces were edited using a professional 3-D modelling package (3D Studio Max 2013, Autodesk) to remove the identity information from the size and contour of the morphs by resizing them to the dimension of the average face and then applying a random surface noise to the external contour (Figure 1). The sizes of the final objects were approximately 180 × 280 × 220 in 3Ds Max units. The stimuli were presented on a Samsung (S23A700D) 3-D screen operating at 1920 × 1080 and with a refresh rate of 120 Hz, driven by a dual HDMI cable connected to the stimulus computer. Subjects viewed the display from a distance of 60 cm. Each stimulus subtended 16° of visual angle vertically and 10.5° horizontally. As described further below, each stimulus was presented at a random location in order to measure the location of the first fixation landing. The face stimuli were presented against a grey background.

**Shading**

All the textures were removed from the surface and a directional light source with 45° angle from horizon was introduced. The frontal view of the face was rendered in orthographic projection to avoid perspec-
The texture stimuli were generated using the method described by (Liu et al., 2005). A random regular noise texture with 100% self-luminance was applied to the surface. The maximum and minimum thresholds for the noise were set to be 75% to 25% of the maximum image luminance range. The size of the noise was 12.7 cm overlaid on the face object about 4 m in width. The images were rescaled for presentation on the screen. The self-luminance was required to remove shading information from the stimuli. The size of the noise minimized nonuniformity of the procedural random texture on the surface. The images were rendered in orthographic projection. In this condition the depth of the object would be well defined by changes in texture gradient across the surface (Figure 2b).

**Stereo**

Using the 3D Studio Max software, two images were captured from the 3-D surfaces from two virtual cameras shifted in the horizontal direction to simulate the position of the two eyes at 60 cm distance from the full screen monitor (±2.86°). The surfaces were covered with a high-density regular noise texture. The maximum and minimum thresholds for the noise were set to be 75% to 25% of the maximum image luminance range. The size of the noise was 7.62 cm overlaid on the face object about 4 m in width. The surface was also set to 100% self-luminance to eliminate shading. The two images were presented alternatively on the screen, and the subject could detect the surface of the face through binocular disparity information provided by use of 3-D shutter eye glasses. In this condition all the shading was removed from the surface with self-luminance and the size of the texture was small enough that it did not allow for monocular detection of changes in the texture gradient, meaning that depth could be resolved only...
when the stimuli were viewed binocularly—monocular viewing resulted in a flat image of random noise (Figure 1c).

**Eye movement recording**

Two-dimensional movement of the left eye was recorded using a custom designed eye-tracker sampling at 400 fps with precision of 0.1° of visual angle at 60 cm distance from the monitor running on the GazeParser software (Sogo, 2013). The position of the eye was calculated based on the relative distance of the pupil and the first Purkinje image. A nine-point linear transform was used to calibrate the camera coordinates to the screen coordinates. To further increase the accuracy, a chin rest and an IR filter on the camera was used. Participants who required prescription glasses wore their contact lenses for the studies. 3-D glasses provided by the display manufacturer were used for the stereo condition. A data file generated by the eye-tracker containing the eye-position and stimulus timings was saved for later analysis.

**Procedure**

The experiment started with the training session where the subjects became familiar with the identity of the faces and were asked to associate each identity to a specific key on the keyboard. Participants were trained for all types of stimuli (Texture, Shading, and Stereo) in different sessions. During the training, one of the four faces was randomly chosen and was presented for 2 s while the subject had unlimited time to respond. No fixation points were presented, and the stimuli were always at the center of the screen. The training session was repeated until the subject performance reached more than 90% accuracy on the 60% morph level for all the three stimuli types. Each training session lasted about five minutes and depending on the subject, four to five training sessions were required to reach the sufficient accuracy.

During the testing session, a face with one of the four identities and 10 morph levels was randomly selected and presented for 2 s. The position of the stimulus was random on the screen with no fixation point so as to be able to record the preparatory fixation of the subject. After two seconds, the stimulus was removed and the screen was filled with a gray background. Participants then made their response without a time limit (Figure 3). Each face and morph level was tested 20 times, for a total of 2,640 trials per subject (3 depth conditions × 4 identities × 11 morph levels × 20 trials per condition). Different depth cues were presented in separate sessions and because of the long duration of each session, subjects had three breaks and the eye tracker was recalibrated after each break. Subjects were tested for each session on separate days, and the overall duration...
of the testing was slightly less than three hours combined.

**Analysis of psychophysical data**

Weibull Functions were fit to the 4AFC data for each subject and depth cue using the Palamedes Toolbox (Prins & Kingdom, 2009). The identification threshold was defined as the amount of face morph needed to achieve 72.2% accuracy. The guessing and the lapse rates were fixed to 0.25 and 0.001, respectively. The recognition thresholds were calculated for each individual subject for each of the three depth-cues.

**Gaze data: Preprocessing and inclusion criteria**

We used Matlab for the preprocessing of eye-data to correct the random positioning of the stimuli on the screen. The eye-data was then analyzed using Ogama 4.3 (Vosskuhler, Nordmeier, Kuchinke, & Jacobs, 2008) to calculate the position and the duration of fixations from raw data. The maximum distance a data point may vary from the average fixation point to be considered as part of fixation was set to be 0.5° of visual angle, and the minimum fixation duration was set to 70 ms. Trials containing less than two and more than eight fixations (±two standard deviations away from average number of fixations per trial) where excluded from further statistical analysis.

**Gaze data: Region-of-interest (ROI) analysis**

Eight regions of interest were selected corresponding to the facial features (left eye, right eye, left cheek, right cheek, nose, forehead, mouse, and chin). The schematic view of the defined facial regions in this study is shown on (Figure 4). The amount of time spent on individual facial areas has been reported as a...
Results

Face identification performance

A repeated-measures ANOVA was performed on the identification thresholds for each depth cue (Shading, Texture, and Stereo). As expected, the identification threshold was significantly different across depth-cues, $F(2, 14) = 17.5, p = 0.001$; Figure 5. Identification thresholds were lowest for shading and highest for texture. In general, subjects found the tasks difficult—although all subjects were trained before the experiment, not all of them were able to perform at 100% accuracy even with 100% identity strength with texture and stereo. By measuring the identification performance for each subject and depth cue and analyzing the eye movement data in relation to the threshold level of performance, we were then able to remove the effect of task difficulty on the eye movements during face identification.

Preferred fixation locations for faces defined by 3-D depth cues

After the onset of the stimuli, the first saccades were towards the center of the face and the consecutive fixations were accumulated in facial regions. The main facial region of interest for shading and texture was the nose with a slight shift towards the right side. The overall amount of time spent on each individual facial feature is shown in Figure 6. The results are the percentages of accumulated durations spent on each area calculated only from fixations, and the times during saccades have been excluded (Findlay & Walker, 1999; Thiele, Henning, Kubischik, & Hoffmann, 2002; Volkman, 1986). Gaze performance for each subject is reported in relation to their psychophysical performance, and we report gaze performance below, at, and above face identification thresholds.

Panels a, b, and c of Figure 6 represent the results for the average face (0% identity), at face-identification threshold, and at 100% identity for each depth-cue, respectively. A repeated-measures ANOVA was performed on duration of fixations with depth-cue (Texture, Shading, and Stereo), difficulty (average face, threshold, and full identity) and facial features (eight facial areas) as within-subject factors. There was no main effect of depth-cue, $F(2, 14) = 0.298, p = 0.657$; while there were main effects of difficulty, $F(2, 14) = 5.291, p = 0.022$, and ROI, $F(7, 49) = 45.998, p < 0.0001$. There were no interaction between difficulty and ROI, $F(14, 98) = 0.278, p = 0.123$, which suggested there was no significant shift in the gaze strategy of subjects into different ROIs as a function of difficulty. We observed a strong interaction between depth-cues and ROIs, $F(14, 98) = 0.343, p < 0.0001$, which suggest a significant shift of gaze strategy for different depth cues.

To compare gaze performance across depth cues and to remove the effect of task difficulty, we only considered gaze behavior during threshold-level face identification performance—for each depth cue, we considered gaze performance during viewing of face morphs that represented the minimum amount of identity information required for correct identification. A posthoc comparison was performed on duration of fixation at threshold with depth cues (shading and texture) and ROI as factors. There was no significant interaction between ROI and depth cue, $F(7, 56) = 3.147, p < 0.08$, suggesting that the interaction effect we detected previously was mainly driven by performance on the disparity-defined faces. Figure 6d represents the duration of fixations for each depth cue at threshold level of face identification performance to facilitate visualization of between-depth-cue effects.

Effect of task difficulty on gaze performance during the recognition of faces from pure depth cues

Previous studies have suggested longer fixation durations are correlated to the difficulty of the task (Rayner, 1998; Underwood, Jebbett, & Roberts, 2004).
We tested the effect of the difficulty of the task on the duration of fixations by using morphed faces. Figure 7 shows the duration of fixation for the first three fixations. The preparatory fixation is the amount of time taken for the gaze to move towards the face after the onset of the stimulus. This duration mainly reflects the detection of the stimulus position after the onset of the stimulus and the difficulty of selecting the first

Figure 6. The overall duration spent on each individual facial area for (a) shading, (b) stereo, and (c) texture at three levels of difficulty (Average face, 100% identity, and at threshold). The results were calculated based on subjects’ individual identification performance. The y axis represents the percentage of the time spent on an area over the entire duration of the trial (2 s). The amount of time during the saccades has been excluded, and only the time spent during the fixations has been included. (d) Represents the same data at threshold for all three depth-cues for better comparison of the effect of depth on fixation positions. There was no significant difference in duration of fixation between shading and texture. Error bars are one standard error across subjects.

Figure 7. Duration of fixation as a function of morph level in (a) Texture, (b) Shading, and (c) Stereo. The data are shown only for the first three fixations and the preparatory fixation of the eye. No significant change was detected for the duration of fixation across morph levels within each depth cue.
saccade target (Bichot & Schall, 1999; Pomplun, Garaas, & Carrasco, 2013; Pomplun, Reingold, & Shen, 2001), and the subsequent fixation durations reflect the visual processing required for selecting the following fixations (Hooge & Erkelens, 1998). The first two fixations are reported to be the most informative—the recognition performance is not affected if the subject is restricted to only two fixations (Hsiao & Cottrell, 2008). Therefore, we did not consider fixations that occurred after the fourth saccade in our analysis.

Repeated-measures ANOVAs were performed on the duration of fixations with fixation count (preparatory, first, second, and third fixation) and morph levels (0% to 100% in 10 steps) as factors. There were no significant changes detected for duration of fixations across different morph levels in any of the depth cues—for Shading, $F(10, 80) = 1.911, p = 0.13$; for Texture, $F(10, 80) = 1.204, p = 0.326$; or for Stereo, $F(10, 60) = 3.29, p = 0.119$. Thus, the amount of facial identity information provided did not alter the duration of fixations: Whether subjects were performing the identification task with high accuracy or were completely at chance, fixation durations remained unchanged.

While fixation durations did not vary as a function of task difficulty, fixation durations did vary as a function of the fixation count in a manner that was consistent across all depth cues. The duration of fixations significantly varied between consecutive fixations within a trial (Figure 8).

Because the durations of fixations were constant across all the morphed levels, the pooled data from all morph levels were used for the analysis of fixation durations by fixation count. A repeated-measures ANOVA was performed on the duration of fixations as a function of fixation count (preparatory, first, second, and third fixation) and depth-cue. There was a main effect of fixation count, $F(3, 18) = 92.77, p < 0.001$; and depth-cue, $F(2, 12) = 14.92, p = 0.004$; but with no interaction between the depth-cue and fixation count, $F(6, 36) = 2.57, p = 0.114$. This suggests significant difference in the duration of fixations for each consecutive saccade.

The second fixation was longer than the third for textured, $t(8) = 2.85, p = 0.02$; and for stereo-defined stimuli, $t(8) = 2.6$ and $p = 0.04$; though for shaded stimuli the second and the third were comparable, $t(8) = 1.115, p = 0.3$. The second fixation was significantly longer than the first for all depth cues, all $t(8) > 4.8, p < 0.001$.

### Discussion

In this study, we examined the effect of depth cues on gaze behavior during face identification. By equating task difficulty across depth cues and using precisely controlled 3-D stimuli, we were able to assess how gaze performance during face identification varies across depth cues. First, we found that the amount of time spent on a given facial region was highly comparable between faces defined by shading and texture, but gaze performance for disparity-defined faces differed—more fixations were spent on the cheeks than with the other depth cues. Second, we observed a comparable pattern of fixation durations for all depth cues across the most informative saccades (first, second, and third), suggesting that similar amounts of surface information are extracted in each fixation regardless of depth cue. Third, we did not observe a role of task difficulty on fixation durations on different facial regions nor on the fixation count—fixation durations across facial regions were highly conserved regardless of whether the face being identified was the average face, close to the threshold, or well-above identification threshold.

These findings extend previous psychophysical studies on face recognition from 3-D depth cues (Dehmoobadsharifabadi & Farivar, 2016; Farivar et al., 2009; Liu et al., 2005) and suggest that depth cue invariant face representations may be formed by comparable (e.g., for shading and texture) as well as divergent (e.g., for disparity) gaze strategies.

### Gaze behavior during face recognition

The mechanism of encoding information from different depth-cues is believed to be different in texture and shading (Grill-Spector, Kushnir, Edelman, Itzchak, & Malach, 1998; Sereno, Trinath, Augath, & Logothetis, 2002). Our result suggested that although...
different mechanisms are involved, the brain tends to gather the information with the same strategy in the case of position of fixations. While the only shared attribute of the stimuli used is the depth profile of the object, this can suggest that the information from different processes aggregated to generate a comprehensive depth-map of the object, and this depth-map is acting as a guiding force that dictates where to look.

We observed more fixations around the cheeks in the stereo task, which was consistent with a previous report. Chelnokova and Laeng (2011) interpreted the increased fixation time spent on cheeks during the 3-D viewing as a preference for more volumetric properties of the scene during 3-D viewing. But these volumetric properties were consistent in our set of stimuli for different depth cues and could not modify the gaze behavior. Instead, we suggest that the preference of greater fixations on the cheek may be a feature specific to the use of disparity rather than a general preference for volumetric information.

Whereas our monocular recording conditions preclude a deep discussion of the role of vergence eye movements during the disparity tasks, we believe vergence eye movements played very little role in the pattern described here. The stereo-paired images shown to the two eyes were almost perfectly colocalized on the screen, allowing us to relate the monocular recording to the screen position of the facial features. It will be interesting to investigate the specific role of vergence eye movements and their speed in the perception of objects defined by disparity or joint with other depth cues.

The preferred fixation locations around the nose and the cheeks were consistent with previous finding of eye-movements during face recognition (Bindemann, Scheepers, & Burton, 2009; Chelnokova & Laeng, 2011; Hsiao & Cottrell, 2008). For shading-defined stimuli, these facial areas are also more salient—local gradients of luminance contrasts and the presence of sharp edges are suggested to be more attractive for the gaze (Itti & Koch, 2000; Olshausen, Anderson, & Van Essen, 1993; Treisman, 1998; Wolfe, 1994). These low-level saliencies are absent in texture- and stereo-defined faces; therefore, our results strongly suggested that the depth profile of the object overcomes any effect of these low-level features. Saliency in the area around the nose could also be due to this area being the facial center of gravity which is suggested to be the first target of gaze (Armann & Bulthoff, 2009; Hsiao & Cottrell, 2008; Peterson & Eckstein, 2012).

We found that the duration of consecutive fixations within a trial is significantly different. The fixation pattern starts with a short fixation (first fixation) and continues with a long fixation (second fixation) and a fixation with intermediate duration (third fixation). This effect, which was reported earlier in face recognition tasks (Hsiao & Cottrell, 2008), was consistent for all the depth-cues used in this experiment (Figure 8). One of the major views in determining the duration of a fixation is believed to be the ongoing visual and cognitive processes that take place during that fixation (Henderson & Ferreira, 1990; Morrison, 1984; Rayner, 1998). In this perspective, the duration of fixation is the time required for the visual system to register the input and calculate the best next saccade target in order to perform the task with maximum efficiency. In this study, we controlled the effect of one of the main factors of the cognitive process by equating the task difficulty. Therefore, the difference between the duration of fixations is mainly due to the time required for the visual system to process the facial surface from the different depth cues and not task difficulty. The difference in processing time can express itself by biasing the duration of all the fixations with a constant value between depth cues. This effect can be seen clearly as a constant shift of the graphs between depth-cues (Figure 8), suggesting that the relative amount of time needed between subsequent fixations is comparable across all depth cues, but the average total fixation duration may vary across depth cues.

**Effect of task difficulty on gaze behavior**

We did not observe an effect of task difficulty (i.e., morph level) on fixation durations. This was in contrast with our prediction that the amount of identity information (i.e., task difficulty) would affect fixation durations (Armann & Bulthoff, 2009; Rayner, 1998; Underwood et al., 2004). Longer fixation durations are often interpreted as demanding of more extensive processing, such as in reading of low-frequency words (Underwood et al., 2004). Our results corroborate that of Armann and Bulthoff (2009) as they also did not observe an effect of task difficulty on fixation duration during a 2AFC face identification task with face morphs. Therefore, the duration of fixation is more likely related to the amount of time the brain requires to extract surface information from the depth-defined stimuli, rather than identification of those surfaces.

In contrast to Armann and Bulthoff (2009), we did not observe an interaction between the amount of time spent on each facial feature and task difficulty, suggesting task difficulty did not alter fixation patterns...
in our task. Whereas Armann and Bulthoff (2009) used simultaneous presentation of two faces, our task required identification of a single face image. This difference in the task may have influenced how fixations were planned—the simultaneous presentation of two face images in Armann and Bulthoff (2009) may have allowed subjects to make an initial guess at the task difficulty and to adjust their gaze strategy accordingly. On our task, subjects had no way of knowing the difficulty of a face before initiating their inspection for identification and may have therefore adopted a single inspection strategy for all morph levels. We speculate that knowledge of task difficulty may alter fixation patterns, but task difficulty itself does not.

Generalizability to other depth cues and object categories

In this study, we were unable to include structure-from-motion mainly because the motion of the stimulus poses difficulties in both reliably analyzing the fixation of the eye and also comparing the fixations with stationary stimuli like shading and texture. Another factor that restricts us to generalize our findings to any other object category is the choice of our stimuli. In order to exclude the task difficulty and have a comparable fixation pattern, we used faces as the stimuli because facial surfaces and their morphs have comparable depth profile. While restricting our analysis of depth-driven eye movements to faces did facilitate equating for task difficulty across depth cues, it does reduce the generalizability of the findings to other object categories.

Depth cue invariance and gaze behavior

Our study sought to determine whether gaze behavior is invariant with regards to depth cues or whether it is depth cue specific. We found moderate support for a depth cue invariant mechanism in gaze planning but with an important caveat. The fact that shaded faces and textured faces resulted in highly comparable pattern of fixations and fixation durations despite large differences in low-level image features suggests low-level features did not influence gaze behavior on our task. Alone, the highly comparable gaze patterns between the shaded and textured faces would support a depth cue invariant model of gaze planning, were it not for the fact that preferred fixation positions for faces defined by binocular disparity differed significantly from that of both textured and shaded faces. While disparity-defined faces on average caused longer fixations, the pattern of fixation durations across the most informative fixations was highly comparable across depth cues, suggesting some aspects of gaze behavior are invariant with regards to depth cue.

We propose that the divergence of the results with disparity-defined faces may have more to do with the limit of stereopsis than depth cue invariance in gaze planning and execution. The perception of the depth profile of an object in stereo-defined stimuli is based on the accumulation of relative disparity information across the surface—the gradient of disparity change that defines the curvature of the cheeks, for example, is effectively the acceleration of disparity over space. Relative disparity sensitivity decreases drastically as a function of eccentricity or distant from fixation (Prince & Rogers, 1998; Rady & Ishak, 1955). Thus detecting the curvature of the cheek would be very difficult when one is fixating on the nose. In order to accurately perceive the changes in cheek curvature from stereopsis, one would need to fixate on the cheek. The increased number of fixations to the cheek area for disparity-defined faces may therefore reflect the limit of human disparity perception rather than a depth cue invariant mechanism for directing gaze at objects.

Conclusion

By equating task difficulty across depth cues and using precisely controlled 3-D stimuli, we were able to assess how gaze performance during face identification varies across depth cues. We found that the amount of time spent on a given facial region was highly comparable between faces defined by shading and texture, but gaze performance for disparity-defined faces did differ. We also observed a comparable pattern of fixation durations for all depth cues across the most informative saccades, suggesting that similar amounts of surface information is extracted in each fixation, regardless of depth cue. Task difficulty did not affect fixation durations or fixation count. The results lend support to a depth cue invariant mechanism for object inspection and gaze planning.

Keywords: face recognition, depth-cue, eye movement, shading, texture, disparity, task difficulty, depth perception

Acknowledgments

The research was funded by NSERC Discovery grant and an internal start-up award to RF by the Research Institute of the McGill University Health Centre.
References


