Proceedings Paper:

Fotios, S., Castleton, H., Lin, Y. et al. (1 more author) (2016) Varying facial expressions in studies of interpersonal judgements and pedestrian lighting. In: Proceedings of CIE 2016: Lighting Quality & Energy Efficiency. CIE 2016 Lighting Quality and Energy Efficiency Conference, 03-05 Mar 2016, Melbourne, Australia. Commission Internationale de L'Eclairage (CIE) , pp. 275-282. ISBN 9783902842657

VARYING FACIAL EXPRESSIONS IN STUDIES OF INTERPERSONAL JUDGEMENTS AND PEDESTRIAN LIGHTING

Steve Fotios1, Holly Castleton2, Yandan Lin 3,4, Biao Yang 5

1 School of Architecture, University of Sheffield

2 Department of Engineering and Mathematics, Sheffield Hallam University

3 Institute for Electric Light Sources, Fudan University, Shanghai, China

4 Engineering Research Center of Advanced Lighting Technology, Ministry of Education, Shanghai, China

5 Department of Civil, Environmental and Geomatic Engineering, University College London

Abstract

This paper reports two analyses carried out to investigate the influence of facial expression choice in work exploring interpersonal judgements and lighting for pedestrians. An experiment was conducted to compare performance on a facial recognition task using target and reference images of either the same or different expressions: these data demonstrated that matching with different expressions led to a significant reduction in recognition. Regarding the results from previous studies of facial emotion recognition, a post-hoc analysis was carried out to draw conclusions from analysis of individual expressions rather than the collation of all six expressions: it was concluded that these data were consistent for all individual expressions.

1 Introduction

An assumption within the tenets of lighting for pedestrians is that they "*need to be able to take a 'good look' at the other users of streets – identification of persons or of intentions*.." [Caminada and van Bommel, 1984]. This assumption has been supported through eye tracking studies involving pedestrians walking outdoors which found that a large number of visual fixations at critical moments are directed at other people [Fotios et al 2015a, 2015b]. A reason to examine other people is to understand their intentions, i.e. whether they are friendly, aggressive or indifferent [BSI, 1992] and thus whether it is safe to approach them or if avoidance action is required. After dark, road lighting should be designed to enhance the performance of such interpersonal judgements.

In this article the focus is on visual cues. There are also acoustic cues and the interaction between these requires further study. Possible visual cues include body posture, facial expression, gaze direction (direct or averted), personal characteristics of the other persons (age, gender, body shape), gait (e.g. walking or running, apparent confidence, swagger), the number of people and their proximity. Further research is required to consider the relative importance attached to these different cues and which are the most appropriate tasks for investigating effects of lighting for pedestrians. The focus of the current paper is judgements concerning the face.

Past research has used two approaches to measure the effect of changes in lighting, facial identify recognition [Caminada and van Bommel, 1984, Knight 2010, Lin & Fotios, 2015] and facial emotional recognition [Fotios et al 2015c, 2015d, Johansson and Rahm, 2015, Yang & Fotios 2015]. One procedure used to study facial identity recognition is matching, requiring a target face to be matched against one of a set of reference faces [Boyce and Rea, 1990, Bruce et al, 1991, Dong et al 2015, Konar et al 2013, Rea, Bullough and Akashi, 2009]. The matching reference image may be identical to that of the target image, in which case the task is simply matching a luminance pattern, or they may be different (e.g. different expressions or photographs taken from different viewpoints) which would demand a greater cognitive load. This choice affects task difficulty which may affect conclusions drawn about the effect of changes in lighting [Lin and Fotios, 2015]. To investigate this the experiment reported here was carried out in which the target and reference faces had the same and different expressions.

Facial emotion recognition seeks a forced choice response as to which of the six universally recognised emotions (happy, sad, fear, anger, disgust and neutral) is being portrayed through facial expression. One limitation of this approach that while target images contained all six expressions, it may be that expression choice matters, for example, that only some expressions are critical for pedestrians' interpersonal evaluations or that some expressions are easier than others to distinguish which would affect task difficulty. To investigate this, a post-hoc analysis was carried out using the results of two previous studies [Fotios et al, 2015d, Yang & Fotios, 2015].

2 Face identity recognition

To investigate the effect of expression variation in facial recognition we repeated the matching task of Dong et al [2015]. A target face was presented on display screen 1 and ten reference images were presented on a second, display screen 2. Photographs on screen 1 were sized so that at the experiment observation distance of 4.5 m they simulated a real distance of 10 m. In trials, test participants were required to state which, if any, of the reference images on screen 2 matched the target on screen 1. A response of 'no-match' was permitted.

Target images were drawn from two databases. European faces were taken from the FACES database [Ebner et al, 2010], these being the same four actors (old male, old female, young male, young female) as used in previous work [Fotios et al, 2015c, 2015d, Yang & Fotios, 2015]. A new database of Chinese faces was created for this work, comprising five younger (students) and five older (teachers) people from Fudan University, including males and females. In both databases, there are six photographs of each actor, these portraying the six distinct expressions.

For a given trial ten reference images were presented on screen 2. The choice of reference images was made according to the three objectives being tested as shown in Table 1. Variations in observation duration (0.5 and 5.0 s) and image luminance (0.1 and 1.0 cd/m^2) were set to give three levels of task difficulty (four experimental conditions) according to the results of Dong et al [2015] as shown in Table 2.

Table 1. Selection of reference images for facial identity recognition test

Objective	**Screen 1 (target image)**	**Screen 2 (10 reference images)**
Identity match with different expression	Actor x with expression y	• **Actor x with an expression that is not y.** (i.e. the correct match) • 7 different actors from the database with any expression: chosen at random (i.e. a possible match on a different trial) • 2 people from the distraction database. (i.e. never a correct match within this experiment)
Identity match with same expression (control)	Actor x with expression y	• **Actor x with expression y.** • 7 different actors from the database with any expression: chosen at random. • 2 people from the distraction database.
No match face (control)	Actor x with expression y	• Any 10 images from the set provided that none of them is a match for the target.

Table 2. Task difficulty as defined by the product of target luminance and duration of presentation

Difficulty	**Luminance (cd/m^2)**	**Duration (s)**	**L*D**
Easy	1	5	5
Middle	1	0.5	0.5
Middle	0.1	5	0.5
Difficult	0.1	0.5	0.05

Table 3. Distribution of task difficulty, experimental and control conditions during trials.

Test condition		**Number of trials (target images) for each test objective**			
Luminance (cd/m^2)	Duration (s)	Match with different expression	Match with same expression (control)	No match face (control)	Total number of trials
1.0	0.5	20	5	5	30
0.1	5.0	20	5	5	30
1.0	5.0	20	5	5	30
0.1	0.5	20	5	5	30
Total		*80*	*20*	*20*	*120*

In an experiment session, each test participant carried out 120 matching trials within 4 blocks (Table 3). The 120 target faces were allocated randomly to these trials. The presentation order of target images was randomized. There were 18 test participants, recruited from the students and staff of Fudan University. All test participants were required to have normal or corrected to normal visual acuity, and all had normal colour vision, which were tested before the trials. Each test participant carried out 120 trials giving 2160 trials in total.

The results are shown in Figure 1, this being the proportion of correct responses plotted against task difficulty as defined by the product of luminance and duration. For the different-face and same-face trials the chance level for making a correct response is 0.1: the results for both conditions exceed chance. It can be seen that the different-face trials have a lower proportion of correct responses than did same-face trials: comparing these two sets of results within the test conditions using the Wilcoxon test suggests that the difference is significant ($p<0.01$, except for condition M2 where $p<0.05$).

These data were also used to examine the proposed definition of task difficulty, i.e. the product of luminance x duration. This was done in two ways, first to determine whether higher levels of difficulty lead to lower performance (lower proportion of correct responses) and second to determine whether the two middle levels of task difficulty lead to the same performance (similar proportions of correct responses).

The different-face data and the no-match data both suggested a significant change in task performance with difficulty and Figure 1 shows that this was the expected increase in correct responses for the easier task. The same-face data did not suggest a change in task performance with difficulty. One reason for this is that performance of this task was at a plateau for all conditions, and this plateau being near maximum performance suggests that the test conditions did not provide a sufficient challenge. The two middle levels of task difficulty were suggested to be significantly different for the different-face data ($p<0.01$) but were not suggested to be different for the same-face data ($p=0.22$) nor the no-match data ($p=0.25$).

Two conclusions were drawn from these data. First, these data confirm that matching faces of different expression leads to a lower level of correct responses than when matching faces having the same expression. We interpret this as being a more difficult task. Second, the results do not appear to consistently support the proposed characterisation of task difficulty as the product of luminance and observation duration.

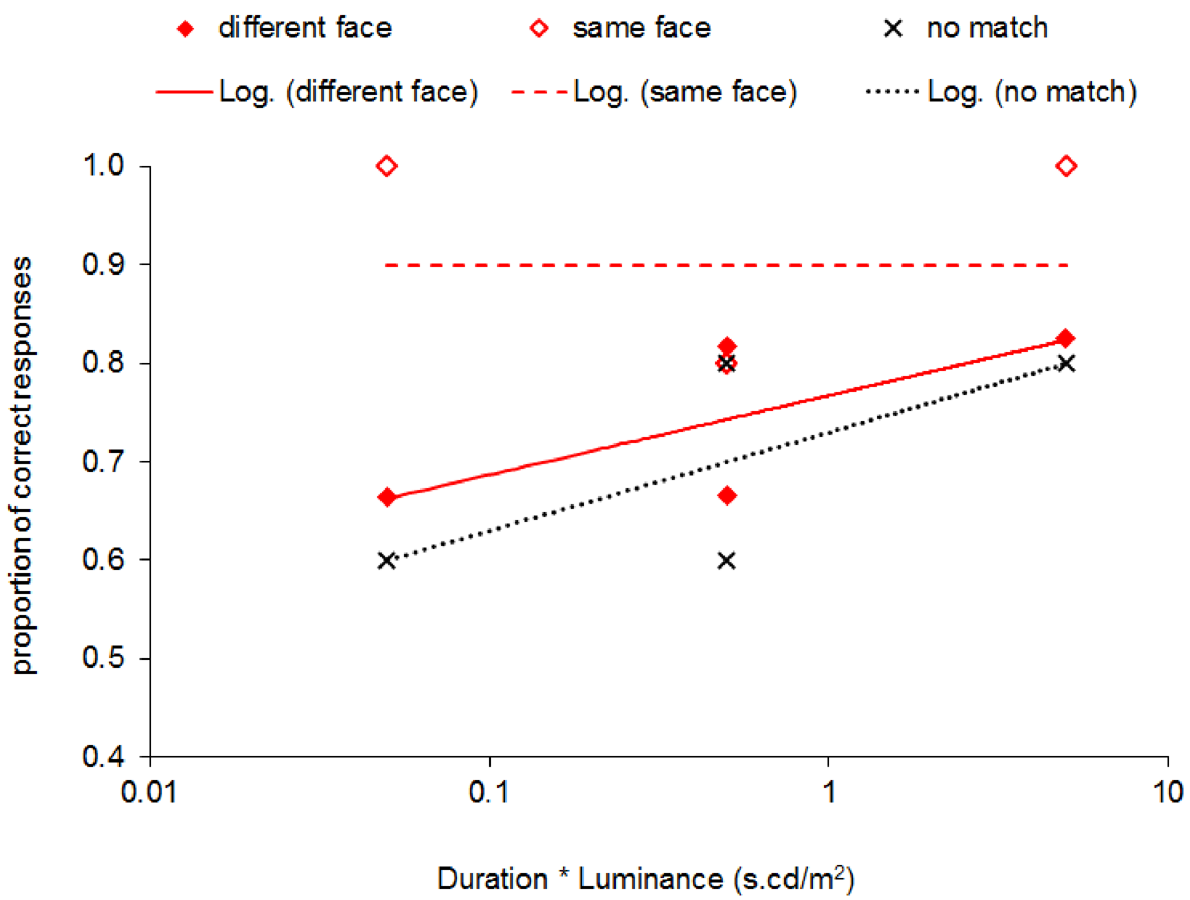

Figure 1. Results of facial recognition test carried out using reference faces having either the same or different face as the target face, and with the reference set not including the target face (no match).

3 Facial emotion recognition

In past studies of facial emotion recognition a test participant's performance was determined across all six expressions. To investigate the effect of expression choice on facial emotion recognition we repeated analysis of data from two studies [Fotios et al 2015d, Yang & Fotios 2015] but for each expression separately.

These studies used photographs from the FACES database, specifically four actors (a young male, a young female, an old male and an old female) each portraying the six universally recognised facial expressions [Ebner et al, 2010]. A photograph was presented for a brief (0.5 or 1.0 s) observation and a six-alternative forced choice of expression response was sought. The targets were observed under lighting of different spectral power distribution (SPD) and luminances, and image size was manipulated to present viewing distances of 4 m, 10 m and 15 m. In one study [Yang & Fotios 2015] the photographs were presented on a non-self-luminous screen with the surrounding environment being lit by the test lighting. In a further study [Fotios et al 2015d] the photographs were projected onto a screen, with the projector light and surrounding field providing identical SPD and luminance.

For each combination of target luminance, light source SPD, presentation duration and image colourfulness there were 24 targets (4 actors x 6 expressions). The frequency of correct response was used to compare performance in each condition. Sample sizes for the two studies were 20 and 28 for Yang and Fotios [2015] and Fotios at al [2015d] respectively. The data were not drawn from a normally distributed population and hence analyses of differences were drawn using non-parametric statistical tests for repeated measures,. While repeated application of such tests raises the risk of falsely indicating a significant effect (a type I error), adjusting the threshold p-value to compensate raises the risk of failing to identify a significant effect (a type II error): Rothman [Rothman, 1990] recommends that we do not make adjustment for multiple comparisons. Following the example of previous work [Boyce & Cuttle, 1990] we retained the standard threshold of $p<0.05$ and drew conclusions by consideration of the overall pattern of results rather than by placing emphasis on any one result.

Four factors were analysed: the effect of luminance, lamp SPD, observation distance, and target colour (Fotios et al [2015d] used grey scale and colour photographs). Here we discuss only the effect of lamp SPD, that being the issue of most controversy, for which the original data did not suggest the effect of SPD to be significant. Tables 4 and 5 show the outcome of statistical testing, the p value determined using the Wilcoxon test (comparing lamp pairs) and the Friedman test (comparing three lamps).

Yang and Fotios used three lamps: high pressure sodium (HPS:2000K, S/P= 0.57, Ra = 25) and two types of metal halide (MH: 4200K, S/P = 1.77, Ra = 92, and CPO: 2868K, S/P= 1.22, Ra = 70). Tests were carried out at six luminances (0.01 to 3.33 cd/m^2), with images sized to simulate distances of 4 m and 15 m and presented for either 500 ms or 1000 ms. Table 4 shows the results of using the Friedman test to compare all three lamps. In only 5/144 cases is the difference suggested to be a significant effect (p<0.05) which does not suggest lamp SPD has a significant effect.

Table 4. P-values determined using the Friedman test to compare facial expression discrimination performance under three lamps: this is a post-hoc analysis of the data from Yang and Fotios [2015]. Values in **bold** are those where p<0.05.

Luminance	**Distance**	**Duration**	**Expression**					
			Angry	**Disgust**	**Fear**	**Happy**	**Neutral**	**Sad**
3.33	4	1000	0.412	0.284	0.307	0.779	0.846	0.323
	4	500	0.973	0.458	0.294	0.097	0.041	0.751
	15	1000	0.736	0.696	0.360	0.494	0.864	0.439
	15	500	0.247	0.362	0.078	0.751	0.135	0.138
1.00	4	1000	0.476	0.754	0.684	0.108	0.382	0.862
	4	500	0.819	0.657	0.735	0.651	0.110	0.103
	15	1000	0.155	0.336	0.087	0.565	0.497	0.068
	15	500	0.393	0.717	0.465	0.771	0.206	0.581
0.33	4	1000	0.321	**0.034**	0.282	0.811	0.140	0.581
	4	500	1.000	**0.008**	0.180	0.468	0.882	0.936
	15	1000	0.983	0.270	0.544	0.565	0.832	0.979
	15	500	0.607	0.708	0.744	0.526	0.687	0.569
0.10	4	1000	0.839	0.880	**0.012**	0.721	0.100	0.385
	4	500	0.942	0.563	0.106	0.464	0.743	0.478
	15	1000	0.938	0.172	0.288	0.236	0.437	0.884
	15	500	0.467	0.427	0.587	0.814	0.598	0.498
0.03	4	1000	0.839	0.502	0.599	0.799	0.683	0.125
	4	500	0.313	0.926	0.449	0.627	0.528	0.798
	15	1000	0.350	0.811	0.950	0.729	0.049	0.488
	15	500	0.238	0.735	0.951	0.538	0.288	0.148
0.01	4	1000	0.301	0.670	0.310	**0.010**	0.982	0.898
	4	500	0.307	0.879	0.146	0.859	0.849	0.397
	15	1000	0.289	0.656	0.767	0.074	0.819	0.937
	15	500	0.157	0.305	0.439	0.627	0.863	**0.015**

Fotios et al [2015d] used two lamps, the HPS and MH as described above. Tests were carried out at three luminances (0.1, 0.33 and 1.0 cd/m^2), with images sized to simulate distances of 4 m and 15 m and presented for 500 ms. The face images were shown in their original colour format and also in grey scale. Table 5 shows the results of using the Wilcoxon test to compare the two lamps. In only 5/72 cases is the difference suggested to be a significant (or close to significant) effect ($p<0.05$) which again does not suggest lamp SPD has a significant effect.

Further analyses with data from both studies confirmed that emotion recognition is significantly affected by target distance (i.e. target size) and luminance when the expressions were analysed individually.

Table 5. P-values determined using the Wilcoxon test to compare facial expression discrimination performance under two lamps: this is a post-hoc analysis of the data from Fotios et al [2015d]. Values in **bold** are those where $p<0.05$.

Luminance	Distance	Expression					
		Angry	Disgust	Fear	Happy	Neutral	Sad
		grey scale targets					
1.0	4	0.317	0.458	1	0.414	0.705	0.317
0.33	4	0.088	0.485	1	0.655	0.48	0.365
0.1	4	0.205	**0.034**	0.642	**0.059**	0.782	0.499
1.0	15	0.302	0.706	0.785	0.09	0.361	0.496
0.33	15	0.194	0.861	0.414	0.195	**0.018**	0.831
0.1	15	0.941	**0.008**	0.185	0.405	0.705	0.605
		coloured targets					
1.0	4	0.394	0.302	0.705	0.157	0.705	0.251
0.33	4	0.804	0.728	0.527	0.739	0.527	0.403
0.1	4	0.449	0.811	0.266	0.157	0.417	0.225
1.0	15	0.549	0.231	0.128	1	0.32	0.209
0.33	15	0.456	**0.024**	0.806	0.416	0.398	0.415
0.1	15	0.617	0.202	0.559	0.177	0.756	0.672

4 Conclusion

This paper explored the importance of facial expression choice in experiments of lighting and pedestrians interpersonal judgements. Two conclusions were drawn regarding the influence of facial expression:

1. Performance of a facial identify recognition matching task decreases when the target and reference faces present different expressions compared with performance when the target and reference images are identical. Matching non-identical images is the more-representative task and these results indicate that the additional cognitive load is important.
2. Conclusions drawn about the effect of changes in lighting on facial emotion recognition are the same for consideration of individual expressions as they were for analysis of all six expressions combined.

Acknowledgement

This work was carried out through funding received from the Engineering and Physical Sciences Research Council (EPSRC) grant number EP/H050817.

References

BOYCE PR, CUTTLE C. 1990. Effect of correlated colour temperature on the perception of interiors and colour discrimination. Lighting Research & Technology, 22(1); 19-36.

BOYCE, P.R. AND REA, M.S. 1990. Security lighting: effects of illuminance and light source on the capabilities of guards and intruders. Lighting Research and Technology, 22, 57–79.

BRITISH STANDARDS INSTITUTION. 1992. BS 5489-3: 1992. Road Lighting - Part 3: Code of Practice for Lighting for Subsidiary Roads and Associated Pedestrian Areas. London: BSI.

BRUCE, V., HEALEY, P., BURTON, M., DOYLE, T., COOMBES, A. AND LINNEY, A. 1991. Recognising facial surfaces. Perception, 20, 755-769.

CAMINADA JF, VAN BOMMEL WJM. 1984. New lighting criteria for residential areas. Journal of the Illuminating Engineering Society, 13(4); 350-358.

DONG M, FOTIOS S, LIN Y. 2015. The Influence of Observation Duration and Procedure on Luminance Required for Recognition of Pedestrian' Faces. Lighting Research & Technology, 47(6); 693-704.

EBNER N, RIEDIGER M, LINDENBERGER U. 2010. FACES—A database of facial expressions in young, middle-aged, and older women and men: Development and validation. Behavior Research Methods; 42: 351-362.

FOTIOS S, UTTLEY J, CHEAL C, HARA N. 2015a. Using Eye-Tracking To Identify Pedestrians' Critical Visual Tasks. Part 1. Dual Task Approach. Lighting Research & Technology, 47(2); 133-148.

FOTIOS S, UTTLEY J, YANG B. 2015b. Using Eye-Tracking To Identify Pedestrians' Critical Visual Tasks. Part 2. Fixation on pedestrians. Lighting Research & Technology, 47(2); 149-160.

FOTIOS S, YANG B, CHEAL C. 2015c. Effects of Outdoor Lighting on Judgements of Emotion and Gaze Direction. Lighting Research & Technology, 47(3); 301-315.

FOTIOS S, CASTLETON H, CHEAL C, YANG B. 2015d. Investigating The Chromatic Contribution To Recognition Of Facial Expression. Lighting Research & Technology. First published online November 24, 2015 as doi:10.1177/1477153515616166.

JOHANSSON M, RAHM J. 2015. Perceived lighting qualities and pedestrian performance. Proceedings of 28th session of the CIE. Manchester. 2015. 324-333.

KNIGHT C. 2010. Field surveys investigating the effect of lamp spectrum on the perception of safety and comfort at night. Lighting Research and Technology, 42(3); 313-330

KONAR Y, BENNETT PJ, SEKULER AB. 2013. Effects of aging on face identification and holistic face processing. Vision Research, 88; 38-46.

LIN Y, FOTIOS S. 2015. Investigating methods for measuring facial recognition under different road lighting conditions. Lighting Research & Technology, 47(2); 221-235. DOI: 10.1177/1477153513505306

REA MS, BULLOUGH JD, AKASHI Y. 2009. Several views of metal halide and high pressure sodium lighting for outdoor applications. Lighting Research and Technology, 41(4); 297-314.

ROTHMAN KJ. 1990. No Adjustments Are Needed for Multiple Comparisons. Epidemiology, 1(1); 43-46.

YANG B, FOTIOS S. 2015. Lighting and Recognition of Emotion Conveyed by Facial Expressions. Lighting Research and Technology, 47(8); 964-975.