Evaluation of Humanoid Robot Design Based on Global Eye-Tracking Metrics

Fan LI [a,1], Danni CHANG [b], Yisi LIU [a], Jian CUI [a], Shanshan FENG [c], Ning HUANG[d], Chun-Hsien CHEN [e] and Olga SOURINA [a]

[a] *Fraunhofer Singapore*
[b] *School of Design, Shanghai Jiao Tong University, Shanghai, China*
[c] *Inception Institute of Artificial Intelligence, Abu Dhabi, United Arab Emirates*
[d] *Huazhong University of Science and Technology*
[e] *School of Mechanical and Aerospace Engineering, Nanyang Technological University, Singapore*

Abstract. The first impression of robot appearance normally affects the interaction with physical robots. Hence, it is critically important to evaluate the humanoid robot appearance design. This study towards evaluating humanoid robot design based on global eye-tracking metrics. Two methods are selected to extract global eye-tracking metrics, including bin-analysis-based entropy and approximate entropy. The data are collected from an eye-tracking experiment, where 20 participants evaluate 12 humanoid robot appearance designs with their eye movements recorded. The humanoid robots are evaluated from five aspects, namely smartness, friendliness, pleasure, arousal, and dominance. The results show that the entropy of fixation duration and velocity, approximate entropy of saccades amplitude are positively associated with the subjective feelings induced by robot appearance. These findings can aid in better understanding the first impression of human-robot interaction and enable the eye-tracking-based evaluation of humanoid robot design. By combining the theory of design and bio-signals analysis, the study contributes to the field of Transdisciplinary Engineering.

Keywords. Humanoid robot design, evaluation, eye-tracking, global metrics

Introduction

Comparing with industrial robots, the design of humanoid robots requires more attention to their appearance design. Industrial robots are created for the sole purpose of performing repetitive tasks. On the contrary, the humanoid robots are normally designed for service and consumers would have to be comfortable being around and interacting with them. Hence, the appearance of humanoid robots, which plays a critical role in influencing human-robot interactions [1, 2], should be studied.

The appearance of humanoid robots provides six meanings, communication of aesthetic, symbolic, functional and ergonomic information, attention-drawing and categorization [1, 3]. In general, questionnaires and scales are widely used for evaluating product appearance [4]. Nevertheless, evaluation data extracted from these methods suffer from the problems of false feelings of inner states and subjective bias [5]. Hence,

---

[1] Corresponding author, Mail: FLI011@e.ntu.edu.sg.

increasing studies have been done to objectively evaluate product appearance from physiological aspects [4], including eye movements, heart rate, blood pressure, facial muscle activity, voice pitch analysis, and brain imaging [6, 7]. Among them, vision is the first channel and receives the most information about a product. Hence, the eye-tracking received most of attention in evaluating product appearance [8]. It is expected that the eye-tracking method is agreeable to evaluating humanoid robot appearance design.

The previous studies found that eye movement parameters, such as fixations and saccades indicated user experience. Nevertheless, these parameters are normally affected by visual attention mechanisms and visual cues. Hence, instead of using these statistical parameters, this study investigates the possibility of evaluating humanoid robot design based on global eye-tracking metrics. Global pattern measures refer to the parameters derived from an overall perspective, and these parameters can be measured without reference to the scene to be viewed [9].

1. Literature review

Several studies evaluated the product design by tracking eye movements and explored the indexes of eye movements which can reflect the user experience of the product [4, 10, 11]. Recent studies showed that it is possible to find correlations between gaze and different aspects of design evaluation. The six most commonly used metrics are overall fixation count, percentage of the total time spent on each area of interest, average fixation duration, fixation count on each area of interest, average dwell time on each area of interest, and overall fixation rate [10, 12]. The average fixation counts normally have a positive correlation with the appearance evaluation score. It was found that users took shorter time to the first fixation and spent longer fixation time to the one they preferred [4].

Besides fixation parameters, saccade parameters have been used for usability studies, too [13]. The saccade duration represents eye movements well and corresponds to all other saccade and fixation parameters. Hence, it has been used to evaluate image quality [13]. Moreover, pupil size is the one widely used to evaluate product design, as it indicates emotions and cognitive states. However, there are some contradictory results in the correlations between pupil size and product design. It was found that users had a smaller pupil size when browsing the product they liked. While other studies found that users had a larger pupil size when evaluating the product they liked. The results may be explained by the effects of visual attention mechanisms [4]. Some parameters are generated based on fixations and saccades, such as the length of scanpath. It was found that a product with high user experience normally induces efficient scan patterns [12]. In other words, users took short scanpath on the product with a high user experience.

These findings may provide a foundation and reference for developing eye-tracking based evaluation methods. However, limited studies investigated the possibility of using global eye-tracking metrics to evaluate product design. Hence, it is really challenging to generalize the results of a specific product to other products. In this study, several global eye-tracking parameters are proposed.

2. Methods

2.1. Data collection

The eye-tracking data and evaluation of humanoid robot design are collected in the study of [14]. Eye movements are recorded by Tobii-X3-120 with a sampling rate of 120 Hz. The raw eye-tracking data is preprocessed by the I-VT fixation filter [15]. The eye-tracking data is time-series data, which can be presented as $E = \{g_x, g_y, p_r, p_l, F\}$, where g_x and g_y are the gaze position of x and y coordinates, p_r and p_l are the right and left pupil size, respectively, and F is the fixation index. The gaze points that have the same fixation index belong to a fixation.

2.2. Evaluation of humanoid robot design

As discussed in the introduction section, the evaluation of product appearance includes the subjective feeling of aesthetics and functionality. In this study, the subjective feeling of aesthetics is measured with an emotion scale, which includes pleasure, arousal, and dominance aspects [16]. It is expected that emotional characteristics to robot led to users' acceptance of humanoid robots [17]. Hence, several subdimensions of emotions are questioned. The functionality is measured with smartness and friendliness scales [18], which is modified from the PHIT-40 questionnaire. Questions are adapted to suit the picture interaction with a robot. The PHIT-40 is proposed to evaluate humanoid robots, and has been modified and widely used to evaluate user preference of robot [17, 19]. The nine-point Likert scales are adopted to measure these aspects.

2.3. Eye-tracking data

2.3.1. Fixation counts, fixation duration of AOIs

A velocity-based method named I-VT is applied to identify fixations from raw eye-tracking data [15]. For each photo of robot design, six areas of interest (AOIs) are defined, namely, head, chest, right leg, left leg, right arm, and left arm. The fixation counts and duration are calculated for each AOI. AOIs, are user-defined subregions of a displayed stimulus [20].

2.3.2. ApEn of gaze velocity and pupil diameter

The gaze velocity is obtained by the differentiating method. Specifically, the Euclidean distance of the consecutive gaze points is obtained and multiplied by the sampling rate. The gaze velocity is a time-series data, $v_1, \ldots, v_T$. T is the number of gaze points in each pack of eye-tracking data. Similar to the pupil diameter, the ApEn of pupil diameters is calculated based on the normalized data.

   The ApEn is calculated based on a series of data and two predefined parameters m and r. m is the window length, referring to the number of points in each window. r is the threshold of distance. To ensure the statistical validity of the ApEn, m can be set as 1 or 2, and r should be set as the value of the standard deviation of the entire time-series data multiplied by 0.1 to 0.25 [21].

2.3.3. Entropy of fixation distribution [22]

The entropy of fixation distribution can reflect users' choices in visiting more or fewer of AOIs, on more or fewer occasions. The entropy of fixation distribution is investigated from two aspects: fixation duration and fixation count. Fixation duration refers to the dwell time on the specific AOI, represented as Fd_k, k is the kth AOI. Fixation count refers to the number of fixations of specific AOI, represented as Fc_k. K is the number of AOIs.

$$FD = \sum_{k=1}^{K} Fd_k \tag{1}$$

$$Ep_{fd} = -\sum_{k=1}^{K} (Fd_k/FD) * \log (Fd_k/FD) \tag{2}$$

$$FC = \sum_{k=1}^{K} Fc_k \tag{3}$$

$$Ep_{fc} = -\sum_{k=1}^{K} (Fc_k/FC) * \log (Fc_k/FC) \tag{4}$$

2.4. Data analysis

Gaze points out of the photo of robot design are excluded for analysis. All the statistical analysis is performed with MATLAB 2019a. Two analyses are performed. First, repeated measures ANOVA is used to assess the statistical significance of differences in the fixation counts and fixation durations among the six AOIs. Post hoc analysis is conducted using the Tukey's honestly significant difference (HSD) criterion if any significance is found. The significance level is set at 0.05.

Second, the correlation analysis is applied to study the relationships between global eye-tracking metrics and subjective feelings. Pearson's coefficients are obtained and presented afterward. Following the study of [23], the eye-tracking data (E_{ip}) of each subject is classified into two groups based on the evaluation level of robot design. Evaluation levels are ranked as follows: "high" for the robot design that rated "6-9"; and "low" for the robot design that rated "1-5." The pupil diameters are subjected to great individual differences. Hence, the series of pupil diameters are normalized based on the maximum and minimum data and analyzed following the cloud model. The expected value (Ex), entropy (En), and hyper entropy (Hn) across the two levels are analyzed.

3. Results

3.1. Eye-tracking parameters and AOIs

Fixation counts of defined AOIs are shown in Figure 1(a). The ANOVA analysis reports significant differences in fixation counts between six AOIs ($F_{(5,102)}$=34.79, p<0.01). Specifically, most of the fixations located in 'Chest', followed by 'head' and 'right arm'. It can be found that 'left arm' is similar to 'right arm'. Nevertheless, significantly more fixations are put on 'right arm'.

Fixation durations of defined AOIs are shown in Figure 1(b). The ANOVA analysis report significant differences in fixation durations between six AOIs, $F_{(5,102)}$=28.49, p<0.01. Though significantly more fixations are found on 'chest' than on 'head', the total fixation duration of 'head' is almost the same with 'chest' (head=2137.50;

chest=2293.05). The results indicate that the mean fixation duration of 'head' is longer than 'chest', which may be caused by the complex design of 'head'.
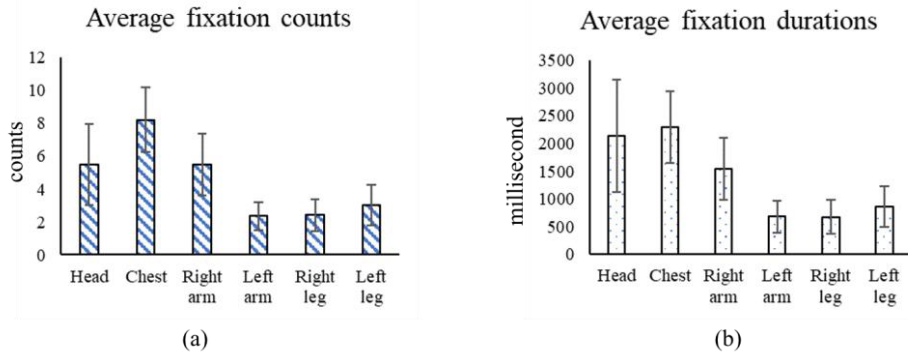

Figure 1. Fixation distributions on six AOIs (a: fixation counts, b: fixation durations).

3.2. Correlations between global eye-tracking parameters and evaluation of humanoid robot design

Table 1 shows Pearson's correlation coefficients between global eye-tracking parameters and subjective evaluation of humanoid robot design. The significant coefficients between global eye-tracking metrics and subjective evaluations can only be found in entropy of fixation count and fixation durations. Pupil dilations have high correlations with dominance while having low correlations with other aspects.

Table 1. Pearson's correlation coefficients between global eye-tracking parameters and subjective evaluation of humanoid robot design.

Eye tracking metrics	Smartness	Friendliness	Pleasure	Arousal	Dominance
Fixation count_entropy	0.011	0.238**	0.124	0.084	0.031
Fixation duration_entropy	0.037	0.214**	0.164*	0.078	0.049
Pupil dilation_right	0.112	0.019	0.009	0.108	0.186*
Pupil dilation_left	0.142	0.072	-0.016	0.053	0.164*
Pupil_ApEn_right	0.027	-0.043	0.032	-0.023	0.168*
Pupil_ApEn_left	0.033	-0.019	0.048	0.083	0.112
Gaze velocity_ApEn	0.116	0.175	0.112	0.095	0.2**
Saccade amplitude_ApEn	-0.155*	-0.039	0.043	0.075	0.124

Note: *P<0.05, **p<0.01.

The ApEn of gaze velocity reports a significant correlation with dominance, while shows no significant correlation with other aspects. Hence, the ApEn would contribute little to the evaluation of robot appearance. For the ApEn of saccade amplitude, though no significant correlation is found between the ApEn of saccade amplitude and other aspects, the coefficient between smartness and the ApEn of saccade amplitude is significant. Since smartness is highly correlated with preference [14], the ApEn of saccade amplitude may contribute to user preference detection.

3.3. The cloud model of pupil dilations across apparent usability levels

The above analysis shows that the pupil dilations have low correlations with apparent usability. The results may be caused by the great variance and randomness of pupil diameters. In this section, the cloud model is adopted to visualize the distribution

of pupil diameter across two levels of apparent usability and workload. Figure 2 shows the cloud model of pupil dilations. It indicates significantly different distributions of pupil dilations across two levels of apparent usability, except friendliness. Hence, it is expected that cloud parameters of pupil dilations can be used to detect apparent usability and assess preference.
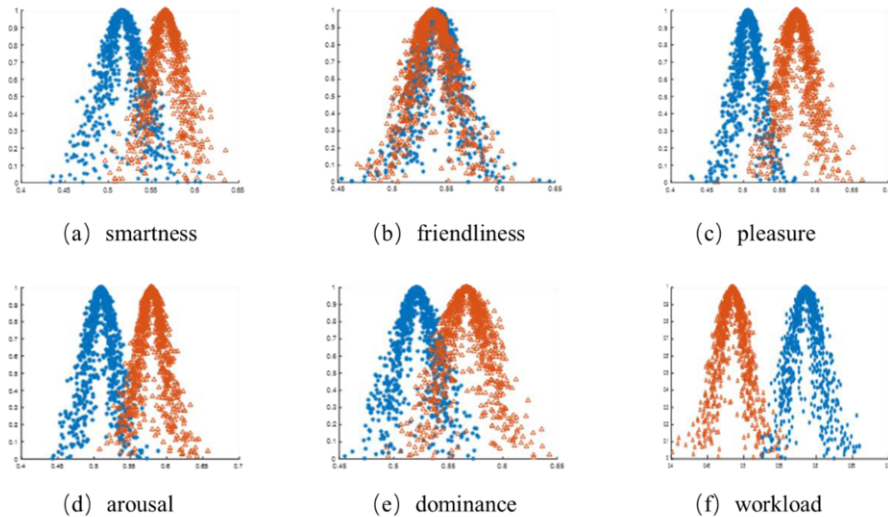

(a) smartness   (b) friendliness   (c) pleasure

(d) arousal   (e) dominance   (f) workload

Figure 2. The cloud model of pupil dilations across apparent usability (blue: low usability; red: high usability).

4. Discussion

In the present study, we aim to determine whether global eye-tracking metrics could be practical indicators of humanoid robot appearance design. In particular, two types of global eye-tracking metrics were examined in terms of their correlations with subjective evaluations. The methods for generating these global eye-tracking metrics are all widely used for other biosignals, such as heart rate and brain dynamics [24]. Therefore, the global eye-tracking metrics are expected to reflect the changes in appearance preferences and can be measured without reference to the scene to be viewed, which makes it feasible to generalize the results to other product designs and apply the findings obtained from this study into practice.

Analysis 1 reports that most of the fixations locate in 'chest', followed by 'head' and 'right arm'. Since the 'chest' AOI is in the center of the display, it is reasonable that subjects fixate on it [25]. The analysis shows that though 'chest' has most fixations, its mean fixation duration is shorter than 'head'. The results may be caused by two reasons. On the one hand, 'chest' is the center part of robot design, users must scan across it to transfer among other AOIs. Hence, some short fixations may generate. On the other hand, the 'head' AOI delivers information of facial expression and human likeness, which are two important influencing factors of human-robot interaction [19]. Since the fixation duration depends on the amount of visual information [26], subjects generate long fixations in the 'head' AOI. A user's perception of a robot can be strongly influenced by its facial appearance [27]. The results show that most of the fixations distributed on the

faces of robots, supporting that facial appearance attracts users' attention. The previous studies pointed out that the spontaneous fixation positions were normally in the center or upper left [25]. It is reasonable to find that more and longer fixations were in 'right arm' than in 'left arm.'

The correlation analysis shows the possibility of evaluating humanoid robot design using global eye-tracking metrics. The entropy of fixation duration, the entropy of fixation counts, and ApEn of saccade amplitude are highly correlated with subjective evaluations. The positive correlations between the entropy of fixation counts and friendliness indicate that the randomness of fixation distribution increases with levels of friendliness shown by the humanoid robot design. In other words, subjects would fixate on some specific AOIs if they think the robot is friendly. If subjects dislike the robot design, their fixation distribution would be dispersed. The ApEn of saccade amplitude indicates the disorder and randomness of saccade amplitude [9]. The negative correlation is identified between the ApEn of saccade amplitude and the smartness level of robot design. The results indicate that subjects would randomly scan the robot if they feel that the robot is not smart. However, pupil dilations, ApEn of pupil dilations, and ApEn of gaze velocity do not report any correlations with subjective feelings except "dominance". This result is contradictory to some studies [4, 28], which indicate the possibility of using pupil size to detect preference, while it is consistent with some other studies, that stimuli category and gender difference may cause different variations in pupil size [8].

Considering the randomness and great variances in pupil diameters, the authors propose to classify pupil data into two groups according to the subjective rating of apparent usability and then generate a corresponding cloud model. It can be found that the cloud model shows the distinguishable distribution of pupil size across two levels of apparent usabilities, such as smartness, pleasure, arousal, and dominance. The results show the benefits of effectiveness, efficiency, and flexibility of the cloud model.

5. Conclusion

Users' preference of product design is normally established by the first impression. This study intends to investigate into the correlations between users' evaluation of humanoid robot design and global eye-tracking metrics. A case study is conducted with data from an experiment of evaluating humanoid robot design. The results show that the entropy of fixation count and fixation duration are closely correlated with the subjective evaluation. Hence, it is expected that they can be used to assess humanoid robot design in future works. The results show that pupil dilations have high correlations with dominance induced by the robot design while no significant correlation with other subjective feelings.

This study points out the benefits of global eye-tracking metrics in evaluating humanoid robot appearance design and provides some fundamental results. Nevertheless, only several aspects of humanoid robot design are studied. For future works, more influencing factors of appearance evaluation should be considered. An algorithm to evaluate humanoid robot design using the global eye-tracking metrics should be developed.

Acknowledgement

References

[1]   M.E. Creusen and J.P. Schoormans, The different roles of product appearance in consumer choice. *Journal of product innovation management*, 22(1), 2005, pp. 63-81.

[2]   C.-F. Chi, R.S. Dewi, P. Samali and D.-Y. Hsieh, Preference ranking test for different icon design formats for smart living room and bathroom functions, *Applied ergonomics*, Vol. 81, 2019, 102891.

[3]   R. Mugge and J.P. Schoormans, Product design and apparent usability. The influence of novelty in product appearance, *Applied ergonomics*, 43(6), 2012, pp. 1081-1088.

[4]   F. Guo, Y. Ding, W. Liu, C. Liu and X. Zhang, Can eye-tracking data be measured to assess product design?: Visual attention mechanism should be considered, *International Journal of Industrial Ergonomics*, 53, 2016, pp. 229-235.

[5]   E.L.-C. Law and P. Van Schaik, Modelling user experience–An agenda for research and practice, *Interacting with computers*, 22(5), 2010, pp. 313-322.

[6]   L.H. Chew, J. Teo and J. Mountstephens, Aesthetic preference recognition of 3D shapes using EEG, *Cognitive neurodynamics*, 10(2), 2016, pp. 165-173.

[7]   M. Kitabata, Y. Niigaki and Y. Horita, Consideration of Relationship between Human Preference and Pulse Wave Derived from Brain Activity, *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, 102(9), 2019, pp. 1250-1253.

[8]   Q.-X. Qu and F. Guo, Can eye movements be effectively measured to assess product design?: Gender differences should be considered, *International Journal of Industrial Ergonomics*, 72, 2019, pp. 281-289.

[9]   M.G. Glaholt, *Eye tracking in the cockpit: a review of the relationships between eye movements and the aviators cognitive state*, Defence Research and Development Toronto (Canada), 2014.

[10]   S. Kukkonen, Exploring eye tracking in design evaluation, *Joining Forces*, 2005, pp. 119-126.

[11]   Q.-X. Qu, L. Zhang, W.-Y. Chao and V. Duffy, User experience design based on eye-tracking technology: a case study on smartphone APPs, in Advances in applied digital human modeling and simulation, In: V.G. Duffy (ed.) *Advances in Applied Digital Human Modeling and Simulation*, Springer International Publishing, Cham, 2017, pp. 303-315.

[12]   B. Fu, N.F. Noy and M.-A. Storey, Eye tracking the user experience–an evaluation of ontology visualization techniques, *Semantic Web*, 8(1), 2017, pp. 23-41.

[13]   T. Vuori, M. Olkkonen, M. Pölönen, A. Siren and J. Häkkinen, Can eye movements be quantitatively applied to image quality studies? In: R. Raisamo (ed.) *Proceedings of the third Nordic conference on Human-computer interaction*, Association for Computing Machinery, New York, 2004, pp. 335-338.

[14]   Y. Liu, F. Li, L.H. Tang, Z. Lan, J. Cui, O. Sourina and C.-H. Chen, Detection of Humanoid Robot Design Preferences Using EEG and Eye Tracker. In: *2019 International Conference on Cyberworlds (CW)*, IEEE, Piscataway, 2019, pp. 219-224.

[15]   A. Olsen, *The Tobii I-VT fixation filter.* Tobii Technology, 2012.

[16]   A. Mehrabian, Framework for a comprehensive description and measurement of emotional states, *Genetic, social, and general psychology monographs*, 1995, 121(3), 339–361.

[17]   S.M.F. Hosseini, S. Hilliger, J. Barnes, M. Jeon, C.H. Park and A.M.Howard, Love at first sight: Mere exposure to robot appearance leaves impressions similar to interactions with physical robots, In: *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, Piscataway, 2017, pp. 615-620.

[18]   P.A. Hancock, D.R. Billings, K.E. Schaefer, J.Y. Chen, E.J. De Visser and R. Parasuraman, A meta-analysis of factors affecting trust in human-robot interaction, *Human factors*, 53(5), 2011, pp. 517-527.

[19]   H.R. Lee and S. Sabanović, Culturally variable preferences for robot design and use in South Korea, Turkey, and the United States, In: *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*, ACM, New York, 2014, pp. 17-24.

[20]   B. Farnsworth, *Eye Tracking: The Complete Pocket Guide,* 2018, https://imotions.com/blog/eye-tracking/, accessed June, 1 2020.

[21]   K. Harezlak, D.R. Augustyn and P. Kasprowski, An Analysis of Entropy-Based Eye Movement Events Detection, *Entropy*, 21(2), 2019, 107.

[22]   X. Wang, R. Berg, N. McGeorge, D. LaVergne, N. Benda and A. Bisantz, Assessing Interaction Strategies for Health IT: An Entropy Based Approach. In: *Proceedings of the International Symposium on Human Factors and Ergonomics in Health Care*. SAGE Publications Sage India: New Delhi, India, 2018, pp. 82-86.

[23]   Y. Yasui, J. Tanaka, M. Kakudo and M. Tanaka, Relationship between preference and gaze in modified food using eye tracker, *Journal of prosthodontic research*, 63(2), 2019, pp. 210-215.

[24]   F. Li, C.-H. Chen, G. Xu, L.P. Khoo and Y. Liu, Proactive mental fatigue detection of traffic control operators using bagged trees and gaze-bin analysis, *Advanced Engineering Informatics*, 42, 2019, 100987.

[25]   L. Kaufman and W. Richards, Spontaneous fixation tendencies for visual forms, *Perception & Psychophysics*, 5(2), 1969, pp. 85-88.

[26]   I.T.C. Hooge and C.J. Erkelens, Control of fixation duration in a simple search task, *Perception & Psychophysics*, 58(7), 1996, pp. 969-976.

[27]   M. Blow, K. Dautenhahn, A. Appleby, C.L. Nehaniv and D. Lee, The art of designing robot faces: Dimensions for human-robot interaction. In: *HRI '06: Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*, 2006, pp. 331-332.

[28]   C.-H. Ho and Y.-N. Lu, Can pupil size be measured to assess design products?, *International Journal of Industrial Ergonomics*, 44(3), 2014, pp. 436-441.