Recognizing Visual Attention Allocation Patterns Using Gaze Behaviors

Cheng Wu, Feng Xie$^{(\boxtimes)}$, and Changsheng Yan

School of Urban Rail Transportation, Soochow University, Suzhou 215011, China
{cwu,fengxie}@suda.edu.cn, chang-sheng.yan@hp.com

Abstract. Human Attention Allocation Strategy (HAAS) is related closely to operating performance when he/she is interacting a machine through a human-machine interface. Gaze behaviors, which is acquisited by eye tracking technology, can be used to observe attention allocation. But the performance-sensitive attention allocation strategy is still hard to measure using gaze cue. In this paper, we attempt to understand visual attention allocation behavior and reveal the relationship between attention allocation strategy and interactive performance in a quantitative manner. By using a novel Multiple-Level Clustering approach, we give some results on probabilistic analysis about interactive performance of HAAS patterns in a simulation platform of thermal-hydraulic process plant. It can be observed that these patterns are sensitive to interactive performance. We conclude that our Multiple-Level Clustering approach can extract efficiently human attention allocation patterns and evaluate interactive performance using gaze movements.

Keywords: Bioinformatic · Human Attention Allocation Strategy (HAAS) · Multiple-Level Clustering Algorithm · Human-Machine Interface

1 Introduction

One of the important issues in Human-Machine Interaction (HMI) is to evaluate whether an interface design is optimal or not. An excellent HMI interface should enhance human operators' ability, that is, extracting useful information from an amount of chaotic information and further improving interactive performance for accomplishing a complex task. It concerns with human attention, which is interpreted conventionally as the ability to select a topic of interest (a goal) for extracting information useful for a given task [1]. However, human attention is a scarce resource [2,3]. At any moment, human operators pay attention to only one or a few objects. Paying attention to something makes the operator to see these objects more clearly, more accurately, and remember better. Thus, whether allocation of human attention is reasonable is related closely to the performance of

Human-Machine Interaction. Just based on this viewpoint, allocation of attention has become one of hot topics in the research of Human-Machine Interface in recent years.

In fact, there are still some difficulties when we conduct the study about human attention allocation on human machine interaction. The root causes for this difficulties come from the inherent characteristics of human attention such as [4–6]:

(1) *Human Attention is Unobservable by Itself.* In the real world, Human attention is intangible. There is no reliable way to monitor the status of human attention [1,7].
(2) *Human Attention can Hardly be Measured.* There is neither a physical unit to describe human attention nor a method to infer the value of the specific attention on an object [8,9].

Recently, human behavior symptoms triggered by human attention are used as indicators of human attention. While interacting with the interface through viewing the interface, eye movements play a more salient role than other symptoms. Eye movements can implicitly indicates the focus area of the user's attention. There are several excellent studies on using the eye movement parameters to understand the operator's visual attention focus (or area of interest) [10]. Thus, there seems to be reasonable consensus for eye movements as indicators of visual attention focus. Theoretically, the consensus comes from the scientific rationale and advantages for the use of eye movements in observing human attention allocation focus.

(1) eye movements may yield important clues to human attention focus at a fine temporal grain size, typically on the order of 10 ms;
(2) the data about eye movement is collected non-intrusively, such that data collection in no way affects task performance;
(3) eye movement parameters can serve as the sole source of data or as a supplement to other sources like Verbal Protocols (VP), Electromyography (EMG), Electroencephalogram (EEG) etc.

The measurement of human attention is based on the above consensus that eye movements can act as indicators of visual attention focus. It is well known that information from eye movements includes eye fixations, eye saccades, pupil size, blink rate, and eye vergence etc. Eye fixations are pauses in the eye scanning process over informative regions of interest [11]. Measures of fixation include two attributes: the fixation duration (i.e. the time spent investigating a local area of the visual field; also known as dwell time) and the number of fixations (i.e. the number of times the eye stops on a certain area of the visual field). Longer fixation duration implies more time spent on interpreting, processing or associating a target with its internalized representation. Fixation duration is negatively correlated to the efficiency of task execution [12]. A larger number of fixations imply that more information is required to process a given task [13].

In cognitive sciences, researchers have studied eye movements to elicit and analyze the cognitive processes in a variety of task domains. This research

has achieved notable and significant benefits in domains such as reading [14–16], arithmetic [17], and word problems [18]. In neurobiology, researchers have focused on the relation between the underlying mechanisms of eye movements and human visual processing [11,19]. There are a number of important reasons why eye movements have become popular indicators of human attention in so many research fields. Firstly, eye movements yield important clues to human behavior at a fine temporal grain size, typically on the order of 10 ms. Secondly, eye-movement data is collected non-intrusively, such that data collection in no way affects task performance. Thirdly, eye movements can serve as the sole source of data or as a supplement to other sources like verbal protocols.

In recent years, some researchers show the allocation of attention has close correlation with the performance of interaction between human operators and computers. Michael and Kim [20] presented the correlation between visual attention allocation and various measures of performance using regression analysis, and Wu et al. [8] proposed a notion of effective attention allocation as well as its measure and demonstrated that it is sensitive to operator's performance variations on different interfaces and different operators in a qualitative manner. Although these researches have wonderfully revealed that there exists strong links between attention allocation and performance, an important fact is ignored, that is, different interaction processes with similar performance from different operators often possess similar patterns of visual attention allocation. Therefore, it is imperative to extract the patterns of visual attention in relation to the performance of interaction.

Intuitively speaking, different patterns of visual attention is correlated with different performance characteristics obtained from different interaction processes on a specific interface in Human-Computer Interaction. Some patterns might lead to superior performance while others patterns might be deemed to result in bad performance. Given an interface design, evaluation of the design requires that human subjects perform task on the interface. It is possible that we can find the patterns of operators visual attention allocation which are correlated to task performances. Such a relationship can help the human-computer interface design and human-Machine interaction management in two ways. The first way is that we can select operators based on whether their attention allocation strategy patterns match with the pattern which corresponds to the best task performance. The second way is that we can optimize the interface design in terms of the best task performance achieved by the operator with a certain attention allocation strategy pattern.

Despite its importance, there is still few published work to show the relationship between the patterns of visual attention and performance characteristics, our study just starts to show the relationship in a quantitative manner. It is noted that the relationship between attention patterns and performance should be shown in the form of probabilities because of the uncertainty of human attention itself and external interference. In addition, the extraction of visual attention patterns depends to a great extend on specific task and specific interface.

The remainder of this paper is organized as follows. Section 2 proposes a novel approach - Multiple-Level Clustering Algorithm (MLCA) for extracting the patterns of HAAS. In Sect. 3, we evaluate our approach in a simulation platform of thermal-hydraulic process plant and give probabilistic analysis about interactive performance of HAAS patterns. At last, we concludes the paper and discusses the future work.

2 Multiple-Level Clustering Algorithm

2.1 Two Levels of Clustering for Solving Two Key Questions

As mentioned above, the goal of the study is to extract the patterns of HAAS from dynamic interactive process in HMI. Clearly, there are two key questions need to be solved.

(1) How to represent a HAAS?
(2) How to extract common features of HAAS?

With regard to the first question, it has already been well known that the information of eye movement can provide a possible way to represent a HAAS. The sampling data of eye movement at a moment, specifically, Eye Fixation Point Coordinate (EF-PC), Eye Fixation Dwell Time (EF-DT) reflect current attention focus in a real-time manner. A series of EF-PCs and EF-DTs reveal the allocation of human attention during a dynamic interactive process. Then further combining the detail interface display, all EF-PCs are grouped into main components on the interface display using K-means clustering algorithm. At last, an encoding procedure is issued in order to produce a feasible HAAS. K-means Clustering is chosen because cluster number (the number of main components on an interface display) can be predetermined easily.

With regard to the second question, it is difficult to extract common features from mass HAASes. These features either are independent on these HAASes or have relation with them. For the reason, unsupervised clustering procedure is applied based on its merit to probe the potential rules behind chaotic data. Considering the difficulty of foreknowing the number of clusters, hierarchical clustering becomes a reasonable choice for grouping these HAASes in our experiment. Then, the patterns of HAAS are extracted from different groups.

2.2 Hybrid Clustering Algorithm for Extracting HAAS Patterns

In the section, a novel hybrid clustering algorithm is proposed for extracting HASS patterns. The algorithm includes four phases:

(1) Grouping EF-PC using K-means Clustering;
(2) Encoding a HAAS;
(3) Grouping HAASes using Hierarchical Clustering;
(4) Extracting common features.

Phase I: Grouping EF-PCs: in this phase, K-means Clustering Procedure is used as the first-level clustering for grouping Eye Fixation Points. In order to apply K-means Clustering, the following parameters need to be determined in advance:

– Determining the number of cluster: it depends on the number of key components on interface display.
– Determining cluster distance: it depends on data characteristic. Because the EF-PCs represent some points on a two-dimension space, Squared Euclidean distance is a feasible choice during the course of clustering the EF-PCs. Let P_1 and P_2 be two points on a two-dimension space, then the Squared Euclidean distance D_{se} between P_1 and P_2 is:

$$D_{se} = (x_{p_1} - x_{p_2})^2 + (y_{p_1} - y_{p_2})^2$$

– Determining initial cluster centroids: in order to make clustering procedure more orientable, initial cluster centroids are set to the centers of key components.
– Avoiding local Minima: an iterative algorithm is employed to minimize the sum of distances from each point to its cluster centroid.

Phase II: Encoding HAAS: it is well known that an interface display consists of certain key components and human attention always shuttles across these components. If these components are labeled numerically and are arranged in sequence, the codeword for describing interface components distribution can be constructed. Further, if the total EF-DT of every component is determined and assigned into the corresponding component field of this codeword, the codeword will describe a HAAS. The construction of a HAAS codeword is shown in Fig. 1(b).

(a) HAAS acquisition procedure   (b) HAAS encoding procedure
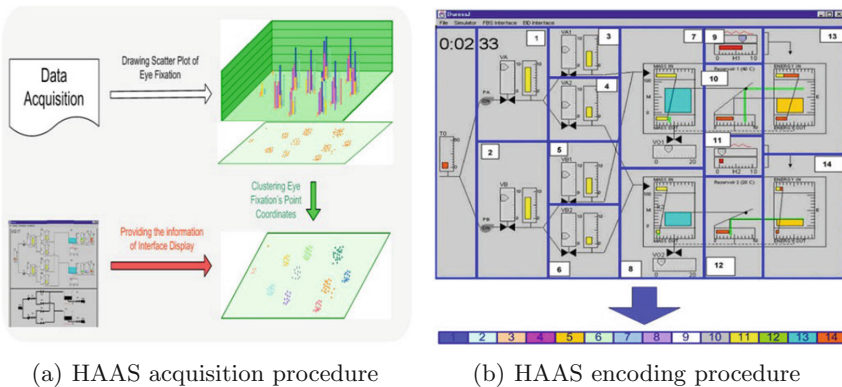
Fig. 1. The representation of a HAAS using EF-PCs. (Color figure online)

Now the problem is how to determine the total EF-DT of every component. The total EF-DT of every cluster can be achieved easily by accumulating all EF-DTs in the cluster. Then the distances between all cluster centroids and all component centers are calculated respectively. The total EF-DT of a cluster is accumulated to the nearest component and then the total EF-DTs of all components can be determined. The codeword is built using absolute EF-DTs, which is called *Codeword With Absolute EF-DTs* (CWAE). It should be seen that CWAE has its limitation. Because the duration of each trial on the experiment is different, what is most meaningful to HAAS is the percentage of the EF-DT on each component over the total duration. Hence, the *Codeword with relative EF-DTs* (CWRE) is designed by replacing absolute EF-DTs with relative EF-DTs, where

$$relative\ EFDT = \frac{absolute\ EFDT\ on\ the\ component}{total\ duration\ of\ the\ trial} \tag{1}$$

Phase III: Grouping HAASes: in this phase, Hierarchical Clustering Procedure is used as the second-level clustering for grouping HAASes. The detail steps are listed as the following.
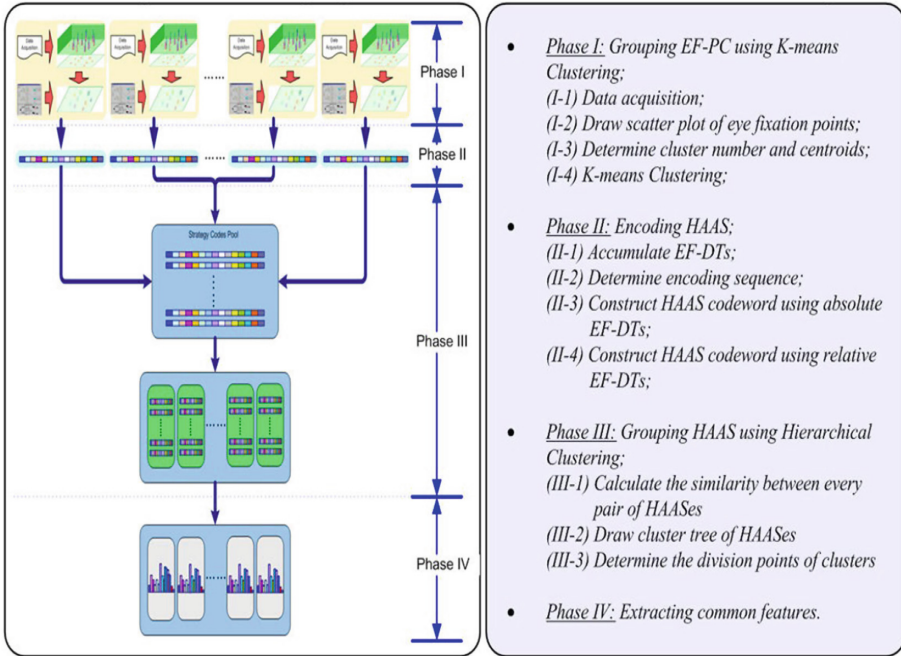
– Determining the Cluster distance and calculating the similarity between every pair of HAASes: it is standardized Euclidean distance that is used as cluster distance. The reason is to eliminate measure difference on different scales, which may result from different participants, different operation time, and different equipment calibration etc. Assume $P_1, P_2, ..., P_m$ are the points on a N-dimension space, the coordinates of the points construct a mXn matrix. To any two points P_i and P_j, the standardized Euclidean distance D_{st} is:

$$D_{st} = (p_i - p_j)D^{-1}(p_i - p_j)'$$

where D is the diagonal matrix with diagonal elements given by v^2, which denotes the variance of the variable p over m point vectors.
– Drawing cluster tree of HAASes.
– Determining division points of clusters and grouping all HAASes: according to the above cluster tree, an inconsistency coefficient threshold is specified for dividing the tree into several parts. Each part corresponds to a cluster respectively.

Phase IV: Extracting common features: In the phase, common features of each cluster are extracted, such as: which component are the operators paying more attention for? Which components are the operators concerned with during the whole interactive process? How about are the frequencies of these components being visited? Note that these common features would be deemed to have strong relations with the performance of interactive process. The Multiple-Level Clustering Algorithm (MLCA) is illustrated and outlined as shown in Fig. 2.

- *Phase I: Grouping EF-PC using K-means Clustering;*
  (I-1) Data acquisition;
  (I-2) Draw scatter plot of eye fixation points;
  (I-3) Determine cluster number and centroids;
  (I-4) K-means Clustering;

- *Phase II: Encoding HAAS;*
  (II-1) Accumulate EF-DTs;
  (II-2) Determine encoding sequence;
  (II-3) Construct HAAS codeword using absolute EF-DTs;
  (II-4) Construct HAAS codeword using relative EF-DTs;

- *Phase III: Grouping HAAS using Hierarchical Clustering;*
  (III-1) Calculate the similarity between every pair of HAASes
  (III-2) Draw cluster tree of HAASes
  (III-3) Determine the division points of clusters

- *Phase IV: Extracting common features.*

Fig. 2. Multiple-Level Clustering Algorithm for extracting HAAS

3   Experimental Evaluation

The section will give an example for demonstrating how to conduce the above algorithm, which is divided into two parts. First, HAAS Patterns extracted using MLCA are presented and some important phases for extracting these patterns are illustrated. Second, combining the performance of interactive process, an evaluation is given to prove the validity of these extracted patterns.

In order to track and record real observers eye movements, experiments have been conducted using an X50 eye tracker from Tobii Technology. This apparatus is mounted on a rigid headrest for greater measurement accuracy (less than $0.5°$ on the fixation point). Experiments were conducted in a simulation platform of a thermal-hydraulic process plant, which was called the dual reservoir simulation system (DURESS). Its interface: EID-DURESS was designed based on Ecological Interface Design (EID) interface design framework. There were 20 subjects (6 female and 14 males) involved in the experiment. All of them had an engineering background. They were asked to control the system (including adjust the openings of the valves and the heaters) until reaching the dynamic equilibrium (the demanded temperature and flow rate of the water out of the reservoirs) as quickly as possible. The participants were required to perform three replications for each trial. As a result, in total there were 60 runs for the entire experiment. The collected data corresponds to 60 different pieces of human visual attention allocation strategies.

3.1   The Extraction of HAAS Patterns

The experimental data of the scenario L01 on the EID-DURESS is analyzed to demonstrate how to extract HAAS patterns using MLCA. There are 60 trials from 20 participants, which correspond to 60 different HAASes.

In the first phase of MLCA, the number of clusters and initial cluster centroids are preset. For EID-DURESS, the number of cluster is 14 and initial cluster centroids are the centers of 14 key components shown as in Fig. 1(b). Then the scatter plot of eye fixation points is draw. Further, a K-Means clustering procedure whose cluster distance is Squared Euclidean distance is issued. After 500 iterations, the cluster result with the minimal total sum of point-to-centroid distances is chosen. As an example, the clustering result of the second trial of Participant 4 is illustrated in Fig. 3(b). The diagram in Fig. 3(c) and (d) give 3D effects of the above cluster result, whose X-Y plant represents interface surface and Z axis represents total EF-DT of a cluster.

In the second phase, an encoding sequence is determined in term of the partition of EID-DURESS. Then, HAAS codewords with absolute EF-DT and relative EF-DT are constructed.

(a) the scatter plot

(b) the cluster plot

(c) the heat map

(d) the 3-D heat map
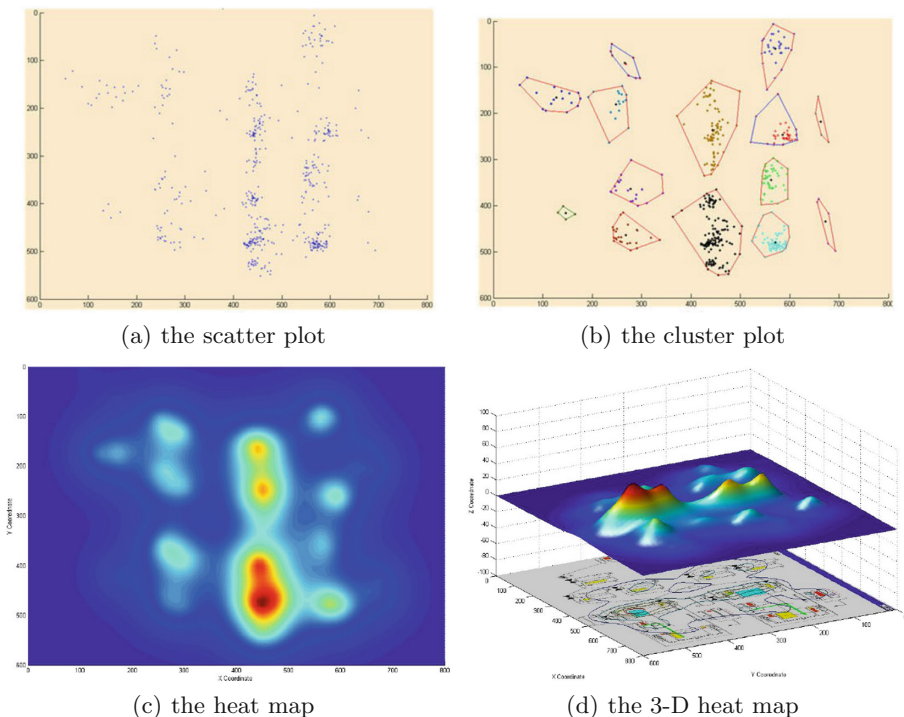
**Fig. 3.** The clustering result of the second trial of Participant 4

In the third phase of MLCA, hierarchical clustering procedure is performed for grouping the HAASes produced in Phase 2 into different clusters. Once the inconsistency coefficient threshold is set to 0.225, the HAASes are divided into four clusters: Cluster 1, Cluster 2, Cluster 3 and Cluster 4. In Cluster 4, there are three sub-clusters: Cluster 41, Cluster 42 and Cluster 43. In the last phase, a statistic analysis is done and some patterns are extracted from each cluster. For example: for Cluster 1, the HAASes adopted by the operators embody the following patterns: (1) the most attention focus is allocated on H2, whose percentage is near or beyond 25 %; (2) M-VOL1, M-VOL2, H1 and Principle2 pull some attention focus on themselves. For Cluster 2, the patterns of the HAAS are: (1) the
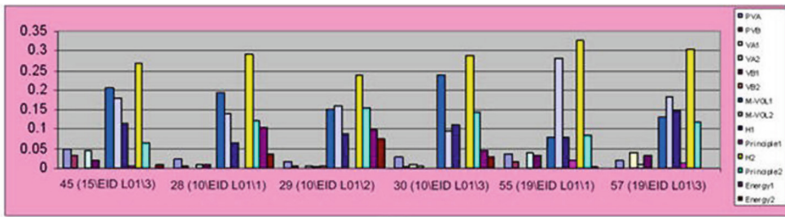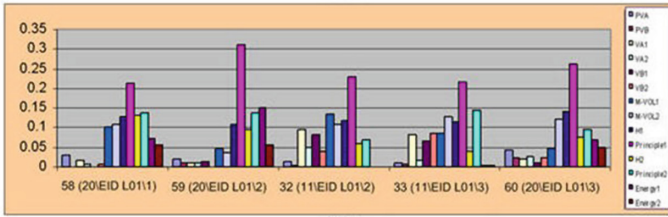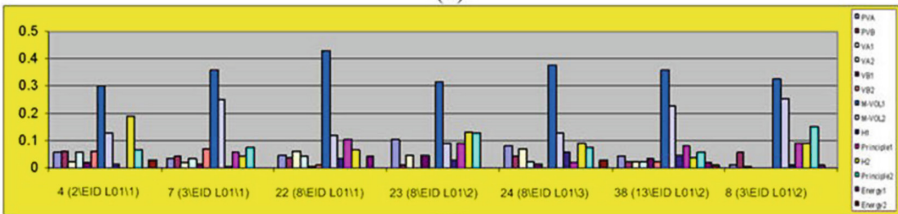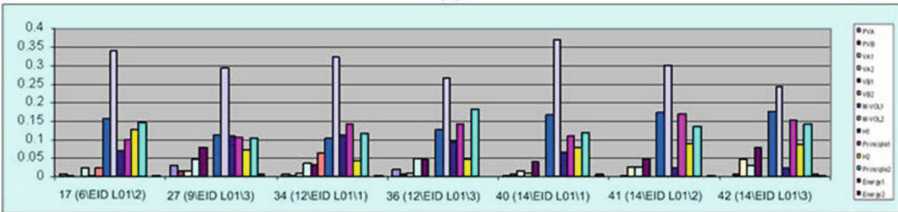

Fig. 4. This is the caption of the figure displaying a white eagle and a white horse on a snow field

most attention focus is allocated on Principle1, whose percentage is beyond 20 %; (2) part of attention is attracted by H1 and Principle2. For Cluster 4, a common pattern of HAAS is that attention focus has a large distribution on M-VOL2, whose percentages are near 20 % for Cluster 41 and beyond 25 % for Cluster 42 and Cluster 43 differently. But Cluster 41, Cluster 42, and Cluster43 have still their own particular patterns. For Cluster 41, M-VOL1 has the same or even more attraction than M-VOL2 to the operators, and at the same time, H1, H2, Principle1 and Principle2 are also pulling some attention focus. For Cluster 42, although most of the HAASes' attention focus percentages on M-VOL2 are clearly higher than those on the other components, M-VOL1, H1, H2, Principle1 and Principle2 play the role as the second attractants. The above patterns about the HAASes for scenario L01 on EID-DURESS can be achieved easily by observing the diagrams shown in Fig. 4.

3.2   Evaluation of HAAS Patterns

By analyzing mass HAASes using MLCA, the patterns of the HAASes have been extracted successfully. However, whether these extracted patterns are valid, that is, have a strong relation with the performance of interaction between human and machine system is still a pending question.

As mentioned in Sect. 2.2, different patterns of HAAS correspond to different interactive performances and the kind of corresponding relationship exists in the form of probabilities. Based on the view, a probabilistic analysis about all interactive performances from the HAASes in each cluster is issued to solve the above question, that is, to prove these patterns' validity. Table 1 describes the statistical result of the HAASes for scenario L01 on EID-DURESS.

Table 1. My caption

Pattern Source		Weight	Interactive Performance		
			Reach Dynamic Balance	Below Average Duration	Above 80% of Maximum Duration
Cluster 1		10.00%	83.33%	80.00%	20.00%
Cluster 2		8.33%	60.00%	100%	0%
Cluster 3		11.67%	28.57%	50.00%	50.00%
Cluster 4	Cluster 5	20.00%	88.67%	50.00%	37.50%
	Cluster 6	35.00%	90.48%	63.16%	15.79%
	Cluster 7	11.67%	85.71%	33.33%	33.33%

From Table 1, it is clear that the patterns with large weight (e.g. over 10 %) have their distinctive performance characteristics. First, the percentages of the HAASes of reaching dynamic balance over the total HAASes concentrate on two contrasting positions, that is, the one from Cluster 3 is very small (only 28.57 %) and the ones from Cluster 1 and Cluster 4 are all very large (beyond 80 %). Specially, the pattern extracted from Cluster 42 with the largest weight

has over 90 % probability to make its HAASes reach dynamic balance. Second, among the clusters that can reach dynamic balance, the pattern from Cluster1 corresponds to the highest performance because most of its HAASes (80 %) is below average duration and only 20 % of the HAASes have the large duration. The pattern from Cluster 42 is secondary and its HAASes? percentages below average duration and above large duration are about 63.16 % and 15.79 % of the totals respectively. The pattern from Cluster 43 is the worst and its performance is deemed to be unsatisfactory because of the lowest percentage of "below average duration" and the higher percentage of "above large duration".

4   Conclusion and Discussion

In this paper, an experiment-oriented study is issued to extract the patterns of HAAS from dynamic interactive process of HMI. First, a novel approach - Multiple-Level Clustering Algorithm is proposed and illustrated. Further, a probabilistic analysis about interactive performance of HAAS patterns is performed, which provides evidence for the validity of the extracted patterns.

With an eye to the necessity of extracting the patterns of HAAS on HMI, the paper has purposed a preliminary study on its algorithm and application. Indeed, still a lot of works need to be further explored. First, the algorithm itself about extracting HAAS patterns may be modified using more complex clustering processes. Second, although eye movement can provide main clues about human attention allocation, more clues from hand movement, head movement and fMRI etc. perhaps should be fused for more accurate and more comprehensive inference. Third, the application about this study would be extended to more complicated cases, such as attention analysis of drivers and pilots, human behavior prediction on intelligent adaptive interface, etc.

References

1. Gonalves, M.G., Luiz, A.A.F., Oliveira, R.A., Grupen, D.S., Wheeler, A.H.F.: Tracing patterns and attention: Humanoid robot cognition. IEEE Intell. Syst. Mag. **15**, 70–75 (2000)
2. Gabaix, X., Laibson, D.I., Moloche, G., Weinberg, S.E.: The allocation of attention: theory and evidence. MIT Department of Economics Working Paper No. 03–31 (2003)
3. Admoni, H., Dragan, A., Srinivasa, S.S., Scassellati, B.: Deliberate delays during robot-to-human handovers improve compliance with gaze communication, pp. 49–56 (2014)
4. Lin, Y., Zhang, W.J., Wu, C., Yang, G., Dy, J.: A fuzzy logics clustering approach to computing human attention allocation using eyegaze movementcue. Int. J. Hum. Comput. Stud. **67**(5), 455–463 (2009)
5. Brown, R., Pham, B., Maeder, A.: Visual importance-biased image synthesis animation. In: Proceedings of the 1st International Conference on Computer Graphics and Interactive Techniques, New York, pp. 63–70 (2003)

6. Wu, C.: Extracting performance-related attention pattern for human-computer interaction. J. Comput. Inf. Syst. 10(17) (2014)
7. Harr, R.: Cognitive Science. Sage Publications, New Jersey (2002)
8. Wu, C., Lin, Y., Zhang, W.J.: Human attention modeling in a human-machine interface based on the incorporation of contextual features in a Bayesian network. In: Proceedings - IEEE International Conference on Systems, Man and Cybernetics, vol. 1, pp. 760–766 (2005)
9. Horvitz, E., Kadie, C., Paek, T., Hovel, D.: Models of attention in computing and communication: from principles to applications. Commun. ACM **46**(3), 52–59 (2003)
10. Hoque, M.M., Onuki, T., Kobayashi, Y., Kuno, Y.: Effect of robots gaze behaviors for attracting and controlling human attention. Adv. Robot. **27**(11), 813–829 (2013)
11. Erkelens, C.J., Vogels, I.: The initial direction and landing position of saccades. In: Findlay, J.M., Walker, R., Kentridge, R.W. (eds.) Eye Movement Research: Mechanisms, Processes, and Applications, pp. 133–144. Elsevier Science Publishing, New York (1995)
12. Asteriadis, S., Karpouzis, K., Kollias, S.: Visual focus of attention in non-calibrated environments using gaze estimation. Int. J. Comput. Vis. **107**(3), 293–316 (2014)
13. Sorostinean, M., Ferland, F., Dang, T.-H.-H., Tapus, A.: Motion-oriented attention for a social gaze robot behavior. In: Beetz, M., Johnston, B., Williams, M.-A. (eds.) ICSR 2014. LNCS, vol. 8755, pp. 310–319. Springer, Heidelberg (2014)
14. Schilling, H.E.H., Rayner, K., Chumbley, J.I.: Comparing naming, lexical decision, and eye fixation times: word frequency effects and individual differences. Mem. Cognit. **26**, 1270–1281 (1998)
15. Pomarjanschi, L., Dorr, M., Bex, P.J., Barth, E.: Simple gaze-contingent cues guide eye movements in a realistic driving simulator, vol. 8651 (2013)
16. Pejsa, T., Andrist, S., Gleicher, M., Mutlu, B.: Gaze and attention management for embodied conversational agents. ACM Trans. Interact. Intell. Syst. **5**, 3 (2015)
17. Suppes, P.: Eye-movement models for arithmetic and reading performance. In: Kowler, E. (ed.) Eye Movements and their Role in Visual and Cognitive Processes, pp. 455–477. Elsevier Science Publishing, New York (1990)
18. Hegarty, M., Mayer, R.E., Green, C.E.: Comprehension of arithmetic word problems: evidence from students eye fixations. J. Educ. Psychol. **84**, 76–84 (1992)
19. Bampatzia, S., Vouloutsi, V., Grechuta, K., Lallée, S., Verschure, P.F.M.J.: Effects of gaze synchronization in human-robot interaction. In: Duff, A., Lepora, N.F., Mura, A., Prescott, T.J., Verschure, P.F.M.J. (eds.) Living Machines 2014. LNCS, vol. 8608, pp. 370–373. Springer, Heidelberg (2014)
20. Michael, E.J., Kim, J.V.: Attention allocation within the abstraction hierarchy. Int. J. Hum. Comput. Stud. **48**, 521–545 (1992)