Local Sequence Alignment for Scan Path Similarity Assessment

Asma Ben Khedher, Imène Jraidi, and Claude Frasson

Abstract—It has long been shown that there is a close relationship between eye movement, human cognition and brain activity. The present work seeks to explore this relationship by investigating the students' saccadic eye movement sequences in a problem solving task. We aim to assess students' reasoning process in a clinical problem solving task using students' visual trajectories. We use students' scan path, followed while resolving medical cases, and a local sequence alignment algorithm, to evaluate their analytical reasoning during medical case resolution. An experimental protocol was conducted with 15 participants. Eye movements were recorded while they were interacting with our learning environment. The proposed approach, based on gaze data, can be reliably applied to eye movement sequence comparison. Our findings have implications for improving novice clinicians' reasoning abilities in particular and ultimately enhancing learning outcomes.

Index Terms—Cognitive tasks, eye movements, local sequence alignment, medical reasoning, scan path similarity.

I. INTRODUCTION

Assessing students' learning experience is taken as a compulsory component in virtual learning environments. Indeed, it is extremely difficult to evaluate students' level of understanding during a learning session. Many questions arise: are they assimilating the materials? Do they pay enough attention on the learning targets? Is the reasoning process they perform is correct or incorrect? In order to overcome these issues, several studies tended to measure students' learning outcomes to evaluate their performance. However, learning performance and especially how students learn are not easily measured.

Learning environments are limited in that they do not really focus on the reasons behind the learners outcomes, meaning, why there was a failure not a success [1]. Therefore, endowing computer-based learning environments with the faculty of assessing learners' behavior during the learning process is of primary interest, to provide a more valuable learning experience.

Sensing technology (e.g. wearable cameras, bio sensors, eye tracking, etc.) has shown considerable promising results in analyzing learners' behavior and improving their learning performance [2]-[5]. Chen and his colleagues (2015) developed an attention aware system based on EEG signals to assess students' level of attention within an e-learning environment. The results proved that their system was able to identify low-attention periods that were significantly correlated with posttest scores [6]. Sawyer *et al.* (2017) introduced facial expression tools to their game-based learning environment in order to predict students learning gains. The findings suggest that facial expression tracking provide additive value compared to baseline predictive student models [7]. Some other studies proposed multimodal sensor-based methods for learning assessment [8]-[13]. Krithika and Priya (2016) proposed a system with three data sources namely, head rotation, eyelid status and facial recognition to detect learners' concentration in order to enhance learning experience. The measured level of concentration was then sent in real-time to instructors to provide students with feedbacks. Such techniques provide unobtrusive measurement of students' behavior within the learning environment. However there is a need to provide an in-depth assessment by employing a more efficient tool such as eye tracking. Tracking students' visual behavior is importance in learning environments as a natural source of information to provide pertinent real time data about the location and duration of an individual's eye fixation [14].

Eye tracking systems have flourished since the past few years due to its ease of use, high sensitivity and especially non-intrusiveness [15]-[17]. It has been a very useful tool in many research domains including visualization [18], activity recognition [19], affect detection [20] and especially, learning [21]-[24], to understand how students evolve and progress while learning. In fact, how students are reasoning while solving problems is a fundamental issue, regarding the technical and methodological challenges due to the complexity of the reasoning process. Researches had demonstrated that a relationship exists between *eye movements and cognitive processes* [25], [26]. As a result, the use of eye tracking could be beneficial as a mean to identify the link between the task-relevant information and the learner's knowledge state [27].

Current eye tracking studies are using mostly static metrics [28]. These metrics are measured from the observed data, such as fixation counts within a particular area of interest (AOI). Ben Khedher *et al.* (2016) explored eye movements in order to predict learners' level of performance during an interaction with a narrative-centered learning game. They assessed the relationship between static metrics (i.e. mean fixation duration, number of fixations, number of revisits and time to first fixation) and learners' performance. They were able to discriminate between two groups of learners in term of their scores by achieving a good accuracy using the random forest classifier [29]. Lum and his colleagues (2016) examined also the effect of the number of fixations on learners' performance. Results indicated that they were positively correlated with the mental workload meaning that when the task was not challenging enough, participants'

visual attention decreased [30]. Yet, despite the increased use of this type of metrics, they lack moment-to-moment analysis of students' visual behavior. As a result, another type of eye tracking metrics has been used to study the sequential properties of eye movements. Dynamic metrics are measured according to spatiotemporal dimensions, such as the scan path, which is a series of fixations and saccades representing the visual trajectory of a users' eye movement. These metrics provide richer information as they reveal the dynamics of the visual behavior.

This paper proposes to use eye tracking metrics to model the students' reasoning process. In particular, we use the scan path to represent the sequential visual path a student follows while resolving medical cases. We seek to assess the analytical reasoning process the learners use to yield a general conclusion from existing observations. The objective is to discriminate between the learners in a way that identifies those who follow a correct analytical reasoning that will be later referred to as *reference process*. In this approach, we use a well-known method applied in bioinformatics to quantitatively measure the similarity between two biological sequences.

An experimental protocol was conducted in order to record learners' eye movements while interacting with our virtual environment, trying to identify the matching diagnosis and treatment of six medical cases. We describe the methodology used to extract the scan paths and evaluate the students' analytical reasoning. The rest of the paper is ordered as follows: section II relates some existing work, section III presents the algorithm we will use. In section IV, we present our learning environment and the experimental protocol. In section V, we explain how we model students' reasoning process. Section VI discusses the obtained results and finally in section VII we conclude the paper with some future work directions.

II. STATE OF THE ART

Although most of the current literature use fixation-based and saccade-based metrics for visual analysis, there is a growing interest in using the scan path as a primary metric, as it enables to monitor moment-to-moment changes in the individual's attention and focus [31], [32]. It has thus far been used in diverse domains such as medicine [33], [34], visualization [35]-[37] and affect detection [20]. Since then, the use of scan path in educational research has gained much interest, since it is closely related to information processing according to Noton and Stark [38]. Ho *et al*. (2013) examined students' visual behaviors when reading a web-based scientific report including textual and graphical contents. They aimed to investigate whether students' level of prior knowledge (high and low) has an impact on their reading process. Results showed that both types of students spent more time reading text than graphics with a higher fixation duration on the graphic for the high prior knowledge students, which implies that the other students were struggling when reading a report that integrates both text and graphic information [39]. In the same context, Lee and Wu (2016) explored students' scan paths when reading geometric descriptions. They found the same results which imply that

more than 80% of the students were text-directed [40]. In a problem solving task, the scan path of 40 students was recorded while solving simple algebraic equations in order to provide insight into how they rearrange them. The results indicated that expert students perform a well-defined and organized visual strategy compared to non-expert students who fixated more the presented solutions than the equations themselves [41].

It is obvious that the scan path provides precise and temporal information about the users' visual attention. However some crucial questions arise: how to analyze a students' scan path? What is the most suitable approach? Can the recorded scan paths be applied to assess students' reasoning process? We will be answering the first question by presenting thereafter, the main existing approaches for scan paths' analysis. Concerning, the second question, there has been a long discussion going on but no answer really exists. Although, we describe in the next section, the method we use in this paper to analyze the students' scan path. As for the last question, it is examined in the results and discussion section.

Scan path analysis methods can be classified into four categories: similarity/dissimilarity calculation, transition probability calculation, pattern detection and common scan path identification [36]. The first method compares two scan paths. The second determines the probability of a visual element to be before or after a particular element in the scan path. The third method identifies a specific visual pattern among the observed scan paths or detects the patterns with a high degree of matching. The fourth method detects a common visual pattern within all the identified scan paths. In the present work, we deal with the scan path similarity. Three main methods for scan path comparisons are described below.

String-coding based methods. These methods are the most common used techniques known in the literature. The scan path is recoded as a string of characters. The recoding consists in assigning a letter to each region of interest. The obtained sequences are then compared using string comparison methods. The Levenshtein Distance is based on the string-edit algorithm which computes the minimum number of editing operations (insertion, deletion and substitution) required to transform one scan path to another [42]. Eraslan and his colleagues (2015) developed the eMine scan path algorithm to identify a common scan path among the users when visualizing web pages in order to improve users' experience. They use the string-edit algorithm as a step in their approach to find in a list of scan paths, the two most similar ones [43]. Another popular algorithm is ScanMatch which computes the similarity between two scan paths rather than computing the distance, using the Needlman-Wunch algorithm utilized in the bioinformatics domain for comparing two DNA sequences [44].

Visual methods. Eye tracking offers a visual attention tool namely, the heat map. It highlights the frequently fixated regions by the user. The areas with a high number of fixations and/or fixations duration are called hotspots [45]. Apart from the visual aspects, in such methods, comparing two scan paths remains difficult, hence the use of another graphical method. The Dotplot is used in bioinformatics for visualizing

similarities in genetic sequences [46]. Goldberg and Helfman (2010) applied the dotplot technique to represent and identify similar scan paths. As for the string-edit methods, scan paths are coded as a string of characters and then aligned on the axis of a two dimensional matrix. Each shared character is marked with a dot and finally, the algorithm consists in finding the longest straight line in the matrix representing the common scan path between the two sequences [47].

Geometric methods. Scan paths are coded as multidimensional vectors representing fixations and saccade properties (e.g. location, fixation durations, etc.) and the objective is to compute the minimal distance mapping between the vectors. Jarodzka *et al.* (2010) proposed MultiMatch, a geometric method. Using a vector with five properties namely, length, position, direction, shape and duration, they search for the shortest route (i.e. scan path) between all the fixations based on the Dijkstra algorithm [48].

In this paper, we propose a **string-coding based method** to assess the students' reasoning process while resolving clinical cases. First, we use the scan path to model the student's analytical reasoning. Then, we seek to discriminate between learners in a way to identify those who follow a correct reasoning using a local sequence alignment algorithm namely, Smith-Waterman. We are not the first to perform scan path alignment. There is the work of Cristino and his colleagues (2010) who applied a global alignment based on the Needleman-Wunch algorithm within a web-based visual search task [44].

III. Sequence Alignment Algorithm

Comparing two nucleotides or amino acid sequences to quantify the degree of similarity between them is a commonly used technique in bioinformatics. In the human genetic system, it is fundamental to discover if a DNA sequence is conserved from one species to another or how many similar residues genes they share. Sequence alignment algorithms are intended to identify regions of similarity between the sequences and represent their evolutionary relationships [49].
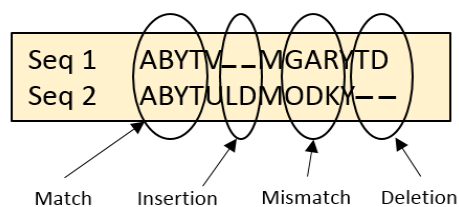

Fig. 1. Editing operations for aligning two sequences.

Aligning two sequences (pairwise alignment) or multiple sequences (multiple alignment) refers to compute the minimum number of editing operations (insertion, deletion and mutation) required to transform one sequence into another. That is, corresponding the sequences' characters to find out if they are similar. Three possible situations can occur when comparing two characters, namely, match, mismatch and gap. A match means that the two characters are identical (A-A, B-B and Y-Y). A mismatch requires a mutation, which is substituting one character for another (G-O, A-D and

R-K) and a gap implies an insertion or a deletion in one of the two sequences (a gap in sequence one leads to an insertion in sequence two and vice versa). Given an alignment between Seq 1 and Seq 2, Fig. 1 shows these different situations.
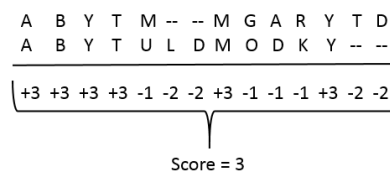

Fig. 2. An alignment of two sequences with the elementary scores for each two characters.

To quantify the alignment between two sequences, a score is computed. It can measure either the similarity or the distance between the sequences. An elementary score is assigned for each position of the alignment where two characters are compared and the final score is the sum of these elementary scores [50]. Let us consider the two previous sequences Seq 1 and Seq 2, we associate for example *+3* if match, *-1* if mismatch and *-2* if gap. Fig. 2 shows the alignment of the two sequences with the associated scores. In this example, there are five matches, four mismatches and four gaps, given a total score of $3 = 5*(+3) + 4*(-1) + 4*(-2)$.

A. Local Alignment

Thus far, biologists attempt to perform an end-to-end alignment, meaning over the whole length of both sequences. However, years later, they found it more interesting to focus on isolated regions of similarity between the sequences, which is not feasible by a global alignment; hence, the new definition of local similarity, known as local alignment. A local alignment of two sequences *s* and *t* is an alignment of a subsequence of *s* and a subsequence of *t*. It is more appropriate than the global and this for different reasons. The first reason is that global alignment compares two sequences that have approximately the same length and are quite similar, whereas in the local alignment arbitrary-length segments of each sequence can be aligned with no penalty for the rest of the sequence regions. In fact, the sequences we use in this study are different in size and are quite different from one another since no participant has the same eye movements as the others. Second, the global alignment, as we have presented earlier, performs an alignment over the entire length of the sequences whereas, what really matters for us, is the regions of the two sequences that coincide the best. That is to find the sub-sequences that have a high level of similarity. Finally, it is more relevant to use local alignment as it is faster and above all, it can be easily considered as global alignment if the two sub-sequences are similar.

There exist another sequence alignment method known as pattern matching which consists in searching for all the approximate occurrences of a particular pattern in a given sequence. In a previous work, we applied this algorithm to our experimental data [51], [52]. We ran bivariate correlations to assess the relationships between the alignment score and students' results in the medical cases. However, no statistically significant results were found. Hence the need to change the algorithm and to test the local alignment and more specifically the Smith Waterman algorithm, which is the

most often used algorithm in biology when looking for local similarities between two sequences.

B. Smith-Waterman Algorithm

The Smith-Waterman alignment algorithm is based on dynamic programming to compare two sequences. Dynamic programming is a method by which the original problem is divided into smaller sub-problems. In the alignment context, it can be used to compute the optimal alignment scores between two given sequences. Basically, the concept is that each elementary score in the alignment is computed using the previously calculated scores (i.e. the characters that were aligned before the current alignment).

The algorithm creates first a similarity-based matrix where a score is computed for each two characters' alignment, and second selects the optimal alignment from all the possible alignments found. Let s and t two sequences of length m and n respectively. We create a (m+1) x (n+1) matrix where the first row and column are initialized with zeros. The remainder cells are filled using the following recursion relation shown in (1).

$$V(i, j) = \max (0, V(i-1, j) + g, V(i, j-1) + g, V(i-1, j-1) + p(s_i, t_j)) \quad (1)$$

where $V(i, j)$ represents the similarity score between the ith element of sequence s and the jth element of the sequence t, g is the gap penalty and $p(s_i, t_j)$ is a match if $s_i = t_j$ and a mismatch if $s_i \neq t_j$. The zero in the recursion allows to ignore a possible negative alignment scores. If the score for a given position in the matrix is negative, it is replaced by the zero value and the alignment may be interrupted. It means that we can restart a local alignment at any position.

Once the matrix is completely filled, we find the cell with the highest score that represents where the alignment ends. Then we do a trace back through the matrix until reaching the first zero, which represents the beginning of the alignment. The constructed path represents one of the optimal local alignments since many alignments may exist for two sequences if we have the same maximum value throughout the matrix. Fig. 3 illustrates an example of an optimal local alignment for two sequences s = TTCATA and t = TGCTCGTA.
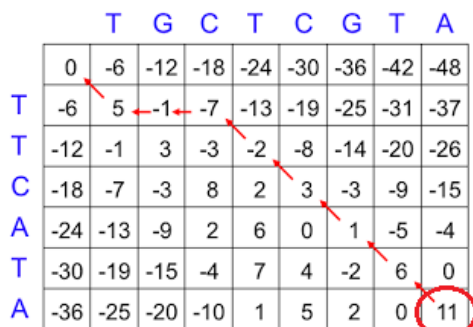

Fig. 3. Similarity matrix using the local alignment (the red arrows represent the optimal alignment path).

IV. EXPERIMENTAL PROTOCOL

We conducted an experiment to record the students' eye movement activity while interacting with our virtual environment called Amnesia. This system was specially designed for the experiment.

A. Amnesia

Amnesia is a medical serious game that assesses the cognitive abilities of novice medical students through clinical decision-making. The curriculum underlying the environment was designed by a medical professional and approved by a doctor. The game features a virtual hospital where the user plays the role of a doctor who was mistakenly diagnosed with amnesia and found himself trapped within the hospital. The player has to prove that he does not suffer from this disease, by resolving in a first step some cognitive tasks (such as memory, attention and logic tests). Memory and attention tests involve quite funny and simple exercises such as card matching, divided attention (simultaneously responding to multiple tasks) and visual memory. Logic tests include more complex tasks that involve inferential skills on numerical and verbal reasoning such as number series and analogies. Then, in a second step, the player has to demonstrate his clinical skills by resolving medical cases.

B. Medical Cases

The game includes six medical cases. For each case, the students are asked to find out the correct diagnosis and the appropriate treatment. To this end, the players are instructed to analyze a series of observations including the patient's demographic information, antecedents, symptoms and clinical data.

The different diseases to be identified are respectively flu, bacterial pneumonia, measles, ebola, mumps and whooping cough. For each diagnosis and treatment, different response alternatives are given. For the diagnosis, there is only one correct answer, which is the current disease. The students are given three attempts to find out the correct answer. Once the diagnosis is established, the student has to identify the right treatment, the same scheme is used as for the diagnosis. The students are given three attempts to respond; after two errors the game is over and there are up to three possible correct responses. Between the attempts, the student can re-analyze the patient in order to correct his answer.

C. Participants and Apparatus

Fifteen participants (8 males) aged between 20 and 27 (M = 21.8 ±2.73) were recruited for the experiment. Participants were all undergraduate medicine students at the University of Montreal. Upon arrival at the laboratory, they were briefed about the experimental objectives and familiarized with the material. They were asked to sign a consent form and placed in front of the eye tracker. They were informed that free head movements were allowed.

A Tobii T×300 eye tracker was used with a sampling rate of 300 Hz. It includes infrared sensors and a camera, all were integrated within a 23-inch computer monitor (1920 x 1080) resolution. The eye tracking system was placed at a distance of approximately 65 cm from the participants' eyes. A 9-point calibration grid was used to evaluate the quality of the students' measured gaze points. This calibration process is essential since it influences the reliability of the data. Once the calibration process is established, the session begins with the game's introductory scene where the players are instructed to discover the environment and meet the different cognitive tasks. Guidance messages were used to help them move forward in the hospital.

At the end of the game session, the participants were shown with the medical cases they were resolving and prompted to recall their reasoning process. They were asked to self-report their visual sequence (i.e. the order of the areas they looked at). They were also asked to give their opinion about the game design and usability in order to have feedbacks for potential corrections.

V. DATA ANALYSIS

Our main objective was to demonstrate how a sequence alignment algorithm could be applied to eye movements in order to assess students' analytical reasoning while resolving medical cases. After completing the recording process as described in the previous section, participants' eye movement sequences were extracted from the collected eye gaze data. First, we aimed to inspect whether we are able to visually model students' reasoning process. Second, given the reasoning sequences, we were interested in assessing their correctness.
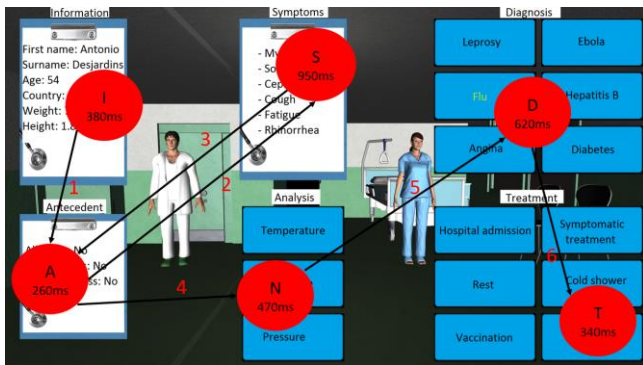

Fig. 4. An example of a scan path over the six AOIs.

A. Students' Reasoning Process

In this study, we need first to represent the visual steps the students use to resolve the medical cases as a sequence. To that end, we used the recorded eye movements to extract, for each medical case, the sequential visual path the students followed to yield their conclusions from the existing observations. First, the obtained data were filtered to fixations. We considered only the regions on which the learner focuses with a fixation duration above 250 ms; this threshold was chosen according to previous studies suggesting that eye fixations should last over 200 ms to be considered as meaningful [53]. Then, to be able to represent the students' visual scan path as a sequence, we needed to divide our environment into areas of interest (AOI).

We defined six areas and assigned a letter to each one, namely (I) Information, (A) Antecedent, (S) Symptoms, (N) aNalysis, (D) Diagnosis and (T) Treatment (see Fig. 4); each fixation occurring within a specific area is then coded with the corresponding letter. In this way, we obtain a scan path recoded as a sequence of characters as follows "IIASSSSSANNDDDTT". Each character corresponds to a visited AOI, and redundant characters refer to areas that have been visited many times successively. Finally, redundancies are removed to maintain only one occurrence of each visited area and create at the end this sequence of eye movement "IASABDT" representing students' reasoning path.

B. Reference Sequence

Once we modelled the students' reasoning process, our next investigation was to assess whether the learner is performing the correct reasoning. More precisely, we need to model this correct reasoning in order to check whether the student's followed scan path, matches the correct sequence. To this end, we used the hypothetico-deductive clinical reasoning shown in Fig. 5 (Adapted from [54]) in order to represent the correct sequence, which will be called hereafter reference sequence.

In this type of analytical reasoning, the clinician generates initial hypotheses at the beginning of the medical interview based on the symptoms exposed by the patient. Then additional clinical data (analysis, radiography, medical antecedents, etc.) are collected and interpreted to evaluate these hypotheses that will be validated or discarded until reaching a correct diagnosis and treatment. From this analytical process, we draw a sequential representation of the different steps a student needs to perform to have a correct reasoning. We obtain thus, this scan path: I→S→A→N→D→T, which yields the following reference sequence: "ISANDT".
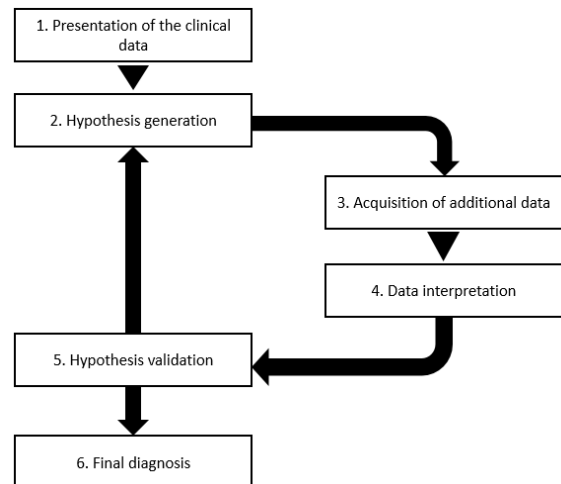

Fig. 5. The hypothetico-deductive reasoning process.

C. Sequence Comparison

The aim of this paper is to investigate whether the reference sequence matches the student's reasoning sequence by quantitatively comparing two eye movement sequences using the Smith-Waterman alignment algorithm. The algorithm uses a similarity-based substitution matrix seeking to find the best alignment with the highest score [55], which implies a strong similarity between the two sequences, meaning that the learner followed the correct reasoning process.

Fig. 6 presents two examples of the local alignment algorithm execution for two different participants. As it can be seen in the figure, we describe the similarity matrix as well as a possible optimal alignment we can obtain for the two sequences. The scoring metrics are set as follows: 2 for match and -1 for mismatch and gap, since we claim to decrease the number of observed differences between the two sequences and increase the alignment score. Once the scores in the array are computed, we search for the maximum value and where it comes from. The cell with the highest score will be used as a

starting point to search for the optimal local alignment which is constructed by tracing back through the matrix until we reach an entry with a score of zero. On the left side (Fig. 6 (a)), an example of a participant who performed a correct reasoning process is shown, as we can notice from the obtained alignment score equal to six. This means that the input reference sequence is perfectly aligned with the participant sequence. In other words, this implies that the student followed a good reasoning process. At the opposite, the learner's reasoning process can deviate from the reference sequence and hence errors (mismatches and gaps) can occur during the alignment. Fig. 6 (b) shows an alignment with two errors where the obtained score is equal to 4, due to the two obtained gaps.

We note in this last example that the alignment was not done over the entire length of the first sequence but only on the beginning. It is, in fact, a local alignment characteristic, which consists in selecting the best possible alignment with the subsequence offering the highest score. It can be easily noticed in the table that there is another maximum value equal to four, but the algorithm did not select it as a starting point for the alignment, as it gives a very low score with a lot of gaps which is not optimal.

Using this proposed approach, four alignment metrics were extracted: score, number of matches (NbMa), number of mismatches (NbMis) and number of gaps (NbGap). These metrics were correlated with the performance-based features: success in the medical case (Yes or No), number of attempts in the diagnostic (NbAtmpDiag) and number of attempts in the treatment (NbAtmpTreat) to investigate the relationships between the learners' performance during medical cases resolution and eye movement alignment results. One-way analyses of variance (ANOVA) were also performed to check if these alignment metrics are significantly different among diagnostic and treatment attempts.

VI. RESULTS AND DISCUSSION

We start first by considering the relationships between the alignment metrics and second their relationships according to students' performance recorded across the attempts in order to highlight the reliability of using a sequence alignment algorithm in assessing students' reasoning.

Statistically significant Pearson's correlation coefficients (r) were found between the score and the number of matches and gaps metrics respectively. We found a positive high correlation between the score and the NbMa (r = 0.637) and a high negative correlation (r = -0.815) between the score and the NbGap. These correlations were statistically significant ($p < 0.01$) suggesting that the alignment score depends on these two metrics and so the participants' reasoning. The more the alignment score is high (high number of matches and low number of gaps), the more the student performs a good reasoning.

A. Association of Alignment Metrics and the Number of Attempt Features

We ran bivariate correlations to assess the relationships respectively between the alignment metrics and the NbAtmpDiag feature and the NbAtmpTreat feature as shown

in Table I. No significant correlation was found with regards to the number of attempts in the treatment feature. However, for the number of attempts in the diagnostic, statistically significant Pearson's correlation coefficients were found.

Two alignment metrics showed a statistically moderate correlation: score (r = -0.310, p =0.024) and NbMis (r = 0.477, p =0.000) and a low positive correlation was also found (r = 0.288, p =0.036) with the number of gaps. These results are consistent since a high number of attempts in the diagnostic is likely to lead to a failure in the medical case resolution. In fact, a failure means a low alignment score and a high number of gaps.
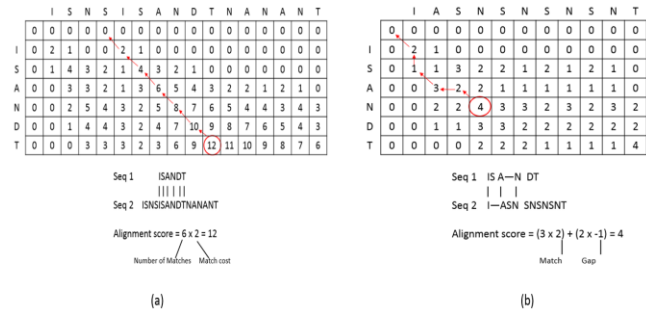

Fig. 6. Two examples of sequence alignment with the corresponding similarity matrices. (a) A perfect match between the sequences. (b) An alignment with two errors. "|" means a match and "—" means a gap.

B. Performance Assessment through the Attempts

In addition, we explored whether there were statistically significant differences between the alignment metrics when the students were reasoning across the different attempts of the medical case resolution. Statistical testing was performed using a one-way analysis of variance (ANOVA). We reported significant differences during the diagnostic attempts, however no differences were reported concerning the treatment. Table I shows the corresponding p-values according to each metric.

First, a significant difference was observed across the four attempts in terms of the alignment score (F (3, 52) = 5.193, p = 0.09). This suggests that at each failure in an attempt, the alignment score decreases which implies a reasoning sequence that deviates from the reference sequence. Similarly for the number of gaps and mismatches, statistically significant differences were found respectively (F (3, 52) = 3.644, p = 0.033) and (F (3, 52) = 24.057, p = 0.000). These findings highlight the fact that the alignment metrics are strongly linked with the students' performance when reasoning.

TABLE I: BIVARIATE CORRELATIONAL RESULTS

Correlation between the NbAtmpDiag and the alignment metrics (N = 53)		
Alignment metric	*R*	*p*
Score	-0.310*	0.024
NbMa	-0.111	0.429
NbGap	0.288*	0.036
NbMis	0.477**	0.000

*Correlation is significant at the 0.05 level.
**Correlation is significant at the 0.01 level.

A scheffe post hoc test was also performed to present detailed group-based statistics for each alignment metric. The results show only three groups corresponding to the three

attempts given to students to find out the diagnosis, meaning that there is no failure in the diagnostic step. Fig.7 shows the results of the means of score, gap and mismatch metrics for the three different attempts. For the score metric, we found that the alignment score significantly decreases in the third attempt compared to the other attempts. These findings clearly show that at the end of the second attempt, the student can no longer reason properly, especially if he has not yet found the right diagnosis. This is reflected in the alignment score (Fig. 7 (a)) where it decreased in the second attempt and it has even dropped below zero in the third attempt. The same results are reflected in the other metrics (i.e. gap and mismatch). One can notice the number of gaps which almost doubled as well as the number of mismatches.
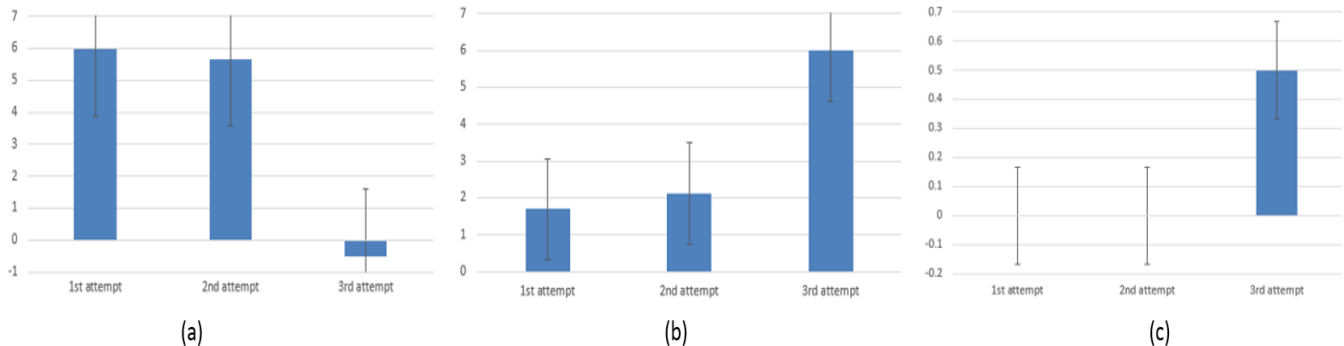

Fig. 7. Comparison of alignment metrics for the three different attempts. (a) Score (b) Gap (c) Mismatch.

VII. Conclusion

In this paper, we have presented our approach for comparing two scan paths extracted from students eye gaze while interacting with the Amnesia learning environment. The method is based on a string representation of the scan path and a local sequence alignment algorithm, namely, Smith-Waterman. The algorithm is used to compute the similarity between two sequences. An experimental protocol was conducted with 15 participants during medical case resolution tasks. They had to identify in each case the correct diagnostic and treatment. An eye tracker was used to record their eye movement and then the obtained scan paths were compared to a reference pattern to compute the similarity between them.

The objective was to evaluate the analytical reasoning process of novice medical students. The approach revealed that we were able to discriminate between the learners in a way to identify those who performed a correct/incorrect reasoning process. These findings could be applied in a more general context for human-computer applications that seek particularly to monitor the user's decision-making process within problem-solving tasks.

As future work, we have planned the classification of both alignment and eye tracking metrics to strengthen our approach and attempt to predict during the task resolution whether the student will perform a good or bad reasoning process. We also plan to integrate further physiological variables such as electroencephalography to assess both the student's visual attention and mental state.

Acknowledgment

References

[1] J. Wiemeyer, M. Kickmeier-Rust, and C. M. Steiner, *Performance Assessment in Serious Games,* Serious Games, pp. 273-302, 2016.

[2] S. S. D'Mello, P. Chipman, and A. Graesser, "Posture as a predictor of learner's affective engagement," in *Proc. the Cognitive Science Society*, vol. 29, pp. 905-910, 2007.

[3] I. Jraidi, M. Chaouachi, and C. Frasson, "A hierarchical probabilistic framework for recognizing learners' interaction experience trends and emotions," *Advances in Human-Computer Interaction*, vol. 2014, 2014.

[4] I. Jraidi, P. Chalfoun, and C. Frasson, "Implicit strategies for intelligent tutoring systems," in *Proc. 11th International Conference on Intelligent Tutoring Systems*, vol. 7315, pp. 1-10, 2012.

[5] I. Jraidi and C. Frasson, "Subliminally enhancing self-esteem: impact on learner performance and affective state," in *Proc. 10th International Conference on Intelligent Tutoring Systems*, vol. 6095, pp. 11-20, 2010.

[6] C. M. Chen, J. Y. Wang, and C. M. Yu, "Assessing the attention levels of students by using a novel attention aware system based on brainwave signals," *British Journal of Educational Technology*, vol. 48, pp. 348-369, 2017.

[7] R. Sawyer, A. Smith, J. Rowe, R. Azevedo, and J. Lester, "Enhancing student models in game-based learning with facial expression recognition," in *Proc. the 25th Conference on User Modeling, Adaptation and Personalization*, pp. 192-201, 2017.

[8] I. Jraidi, M. Chaouachi, and C. Frasson, "A dynamic multimodal approach for assessing learners' interaction experience," in *Proc. the 15th ACM on International Conference on Multimodal Interaction*, pp. 271-278, 2013.

[9] I. Jraidi and C. Frasson, "Student's uncertainty modeling through a multimodal sensor-based approach," *Educational Technology & Society*, vol. 16, pp. 219-230, 2013.

[10] M. Taub, N. V. Mudrick, R. Azevedo, G. C. Millar, J. Rowe *et al.*, "Using multi-channel data with multi-level modeling to assess in-game performance during gameplay with Crystal Island," *Computers in Human Behavior*, 2017.

[11] L. Krithika, "Student emotion recognition system (SERS) for e-learning improvement based on learner concentration metric," *Procedia Computer Science*, vol. 85, pp. 767-776, 2016.

[12] M. Chaouachi, I. Jraidi, and C. Frasson, "Modeling mental workload using EEG features for intelligent systems," in *Proc. the 19th International Conference on User Modeling, Adaption and Personalization*, pp. 50-61, 2011.

[13] M. Chaouachi, P. Chalfoun, I. Jraidi, and C. Frasson, "Affect and mental engagement: Towards adaptability for intelligent systems," in *Proc. the 23rd International FLAIRS Conference*, 2010.

[14] M. Frutos-Pascual and B. Garcia-Zapirain, "Assessing visual attention using eye tracking sensors in intelligent cognitive therapies based on serious games," *Sensors*, vol. 15, pp. 11092-11117, 2015.

[15] S. Chandra, G. Sharma, S. Malhotra, D. Jha, and A. P. Mittal, "Eye tracking based human computer interaction: Applications and their uses," in *Proc. 2015 International Conference on Man and Machine Interfacing (MAMI)*, pp. 1-5, 2015.

[16] H.-S. Yeo, B.-G. Lee, and H. Lim, "Hand tracking and gesture recognition system for human-computer interaction using low-cost hardware," *Multimedia Tools and Applications*, vol. 74, pp. 2687-2715, 2015.

[17] C.-Y. Wang, M.-J. Tsai, and C.-C. Tsai, "Multimedia recipe reading: Predicting learning outcomes and diagnosing cooking interest using eye-tracking measures," *Computers in Human Behavior*, vol. 62, pp. 9-18, 2016.

[18] D. Toker and C. Conati, "Eye tracking to understand user differences in visualization processing with highlighting interventions," in *Proc. International Conference on User Modeling, Adaptation, and Personalization*, pp. 219-230, 2014.

[19] F. Courtemanche, E. A ïmeur, A. Dufresne, M. Najjar, and F. Mpondo, "Activity recognition using eye-gaze movements and traditional interactions," *Interacting with Computers*, vol. 23, pp. 202-213, 2011.

[20] N. Jaques, C. Conati, J. M. Harley, and R. Azevedo, "Predicting affect from gaze data during interaction with an intelligent tutoring system," *Intelligent Tutoring Systems*, pp. 29-38, 2014.

[21] R. Ghali, C. Frasson, and S. Ouellet, "Towards Real Time Detection of Learners' Need of Help in Serious Games," in *Proc. the 29th International Flairs Conference*, pp. 154-157, 2016.

[22] S. Kardan and C. Conati, "Exploring gaze data for determining user learning with an interactive simulation," *User Modeling, Adaptation, and Personalization*, pp. 126-138, 2012.

[23] P. Mart ñez-G ómez and A. Aizawa, "Recognition of understanding level and language skill using measurements of reading behavior," in *Proc. the 19th International Conference on Intelligent User Interfaces*, pp. 95-104, 2014.

[24] J. Yi, B. Sheng, R. Shen, W. Lin, and E. Wu, "Real time learning evaluation based on gaze tracking," in *Proc. the 14th International Conference on Computer-Aided Design and Computer Graphics*, pp. 157-164, 2015.

[25] J. Chakraborty and M. P. McGuire, "Directional scan path characterization of eye tracking sequences: A multi-scale approach," in *Proc. Future Technologies Conference (FTC)*, pp. 51-61, 2016.

[26] M. Hunter, Q. H. Mach, and R. S. Grewal, "The relationship between scan path direction and cognitive processing," in *Proc. the Third C* Conference on Computer Science and Software Engineering*, pp. 97-100, 2010.

[27] A. Ben Khedher, I. Jraidi, and C. Frasson, "Learners' performance tracking using eye gaze data," 1st International Workshop on Supporting Dynamic Cognitive, Affective, and Metacognitive Processes (SD-CAM), part of the 13th International Conference on Intelligent Tutoring Systems, pp. 15-24, 2016.

[28] H. Takeuchi and Y. Habuchi, "A quantitative method for analyzing scan path data obtained by eye tracker," *IEEE Symposium on Computational Intelligence and Data Mining*, pp. 283-286, 2007.

[29] A. Ben Khedher and C. Frasson, "Predicting user learning performance from eye movements during interaction with a serious game," in *Proc. EdMedia: World Conference on Educational Media and Technology*, pp. 1504-1511, 2016.

[30] H. C. Lum, R. L. Greatbatch, G. E. Waldfogle, J. D. Benedict, and D. A. Nembhard, "The relationship of eye movement, workload, and attention on learning in a computer-based training program," in *Proc. the Human Factors and Ergonomics Society Annual Meeting*, vol. 60, pp. 1477-1481, 2016.

[31] E. Cutrell and Z. Guan, "What are you looking for?: An eye-tracking study of information usage in web search," in *Proc. the SIGCHI Conference on Human Factors in Computing Systems*, pp. 407-416, 2007.

[32] M. Hayhoe and D. Ballard, "Eye movements in natural behavior," *Trends in Cognitive Sciences*, vol. 9, pp. 188-194, 2005.

[33] G. Tien, M. S. Atkins, and B. Zheng, "Measuring gaze overlap on videos between multiple observers," in *Proc. the Symposium on Eye Tracking Research and Applications*, pp. 309-312, 2012.

[34] D. C. Hood and A. S. Raza, "On improving the use of OCT imaging for detecting glaucomatous damage," *British Journal of Ophthalmology*, vol. 98, pp. ii1-ii9, 2014.

[35] K. Kurzhals, F. Heimerl, and D. Weiskopf, "ISeeCube: Visual analysis of gaze data for video," in *Proc. the Symposium on Eye Tracking Research and Applications*, pp. 43-50, 2014.

[36] S. Eraslan, Y. Yesilada, and S. Harper, "Eye tracking scanpath analysis techniques on web pages: A survey, evaluation and comparison," *Journal of Eye Movement Research*, vol. 9, 2015.

[37] M. Raschke, T. Blascheck, M. Richter, T. Agapkin, and T. Ertl, "Visual analysis of perceptual and cognitive processes," in *Proc. International Conference on Information Visualization Theory and Applications* pp. 284-291, 2014.

[38] D. Noton and L. Stark, "Scanpaths in saccadic eye movements while viewing and recognizing patterns," *Vision research*, vol. 11, pp. 929-IN8, 1971.

[39] H. N. J. Ho, M.-J. Tsai, C.-Y. Wang, and C.-C. Tsai, "Prior knowledge and online inquiry-based science reading: Evidence from eye tracking," *International Journal of Science and Mathematics Education*, vol. 12, pp. 525-554, 2014.

[40] W.-K. Lee and C.-J. Wu, "Eye movements in integrating geometric text and figure: Scanpaths and given-new effects," *International Journal of Science and Mathematics Education*, pp. 1-16, 2017.

[41] A. Susac, A. Bubic, J. Kaponja, M. Planinic, and M. Palmovic, "Eye movements reveal students'strategies in simple equation solving," *International Journal of Science and Mathematics Education*, vol. 12, pp. 555-577, 2014.

[42] V. I. Levenshtein, "Binary codes capable of correcting deletions, insertions, and reversals," *Soviet Physics Doklady*, vol. 10, pp. 707-710, 1966.

[43] S. Eraslan and Y. Yesilada, "Patterns in eyetracking scanpaths and the affecting factors," *Journal of Web Engineering*, vol. 14, pp. 363-385, 2015.

[44] F. Cristino, S. Math ôt, J. Theeuwes, and I. D. Gilchrist, "ScanMatch: A novel method for comparing fixation sequences," *Behavior Research Methods*, vol. 42, pp. 692-700, 2010.

[45] V. Sundstedt, "Gazing at games: An introduction to eye tracking control," *Synthesis Lectures on Computer Graphics and Animation*, vol. 5, pp. 1-113, 2012.

[46] Y. Huang and L. Zhang, "Rapid and sensitive dot-matrix methods for genome analysis," *Bioinformatics*, vol. 20, pp. 460-466, 2004.

[47] J. H. Goldberg and J. I. Helfman, "Scanpath clustering and aggregation," in *Proc. the 2010 Symposium on Eye-Tracking Research & Applications*, pp. 227-234, 2010.

[48] H. Jarodzka, K. Holmqvist, and M. Nystr öm, "A vector-based, multidimensional scanpath similarity measure," in *Proc. the 2010 Symposium on Eye-Tracking Research & Applications*, pp. 211-218, 2010.

[49] W. R. Pearson, "An introduction to sequence similarity ("homology") searching," *Current Protocols in Bioinformatics*, pp. 3.1. 1-3.1. 8, 2013.

[50] A. C. Melo, M. E. M. Walter, R. C. F. Melo, M. N. Santana, and R. B. Batista, "Local DNA sequence alignment in a cluster of workstations: algorithms and tools," *Journal of the Brazilian Computer Society*, vol. 10, pp. 73-80, 2004.

[51] A. B. Khedher, I. Jraidi, and C. Frasson, "Assessing learners' reasoning using eye tracking and a sequence alignment method," in *Proc. International Conference on Intelligent Computing*, pp. 47-57, 2017.

[52] A. B. Khedher, I. Jraidi, and C. Frasson, "Tracking students' Analytical reasoning using visual scan paths," in *Proc. 2017 IEEE 17th International Conference on Advanced Learning Technologies (ICALT)*, pp. 53-54, 2017.

[53] S. P. Liversedge and J. M. Findlay, "Saccadic eye movements and cognition," *Trends in Cognitive Sciences*, vol. 4, pp. 6-14, 2000.

[54] M. Nendaz, B. Charlin, V. Leblanc, and G. Bordage, "Le raisonnement clinique: donn ées issues de la recherche et implications pour l'enseignement," *P édagogie m édicale*, vol. 6, pp. 235-254, 2005.

[55] J. C. Setubal, J. Meidanis, and Setubal-Meidanis, *Introduction to Computational Molecular Biology*, 1997.

Asma Ben Khedher is a PhD student in the HERON laboratory at the University of Montreal under the supervision of Prof. Claude Frasson. She held her master's degree in computer science from the University of Tunis in 2012, Tunisia and a bachelor degree in the same field and from the same university in 2010. Her research interests include intelligent learning environments, eye tracking systems and their use to assess students' cognitive abilities. She is also a member of the LEADS (Learning Environment Across Disciplines) project in collaboration with McGill University. She was a laboratory assistant in the Jeux Vid éo Intelligents (JVI) laboratory during fall 2015 and winter 2016.

Imène Jraidi is a researcher at the HERON laboratory of the University of Montreal. She received her MS degree (2007) in Business Informatics from the University of Tunis under the supervision of Prof. Zied Elouedi. She completed her PhD (2014) in Computer Science at the University of Montreal under the supervision of Prof. Claude Frasson. Her main research interests are in artificial intelligence, human and machine learning, multimodal interaction and cognitive and affective modeling.

Claude Frasson is a professor in the Computer Science and Operations research Department at the University of Montreal since 1983. He is the head of HERON laboratory. He founded the biannual Intelligent Tutoring Systems, international conference in 1988. He also created the GRITI group (Inter-university Research group in Intelligent Tutors) involving seven universities in Quebec and grouped around the SAFARI project which is a multidisciplinary research group. His major research focuses are mainly on emotional intelligence and intelligent tutoring systems.