



An algorithmic approach to determine expertise development using object-related gaze pattern sequences

Felix S. Wang¹ · Céline Gianduzzo¹ · Mirko Meboldt¹ · Quentin Lohmeyer¹

Accepted: 14 June 2021 / Published online: 13 July 2021
© The Author(s) 2021

Abstract

Eye tracking (ET) technology is increasingly utilized to quantify visual behavior in the study of the development of domain-specific expertise. However, the identification and measurement of distinct gaze patterns using traditional ET metrics has been challenging, and the insights gained shown to be inconclusive about the nature of expert gaze behavior. In this article, we introduce an algorithmic approach for the extraction of object-related gaze sequences and determine task-related expertise by investigating the development of gaze sequence patterns during a multi-trial study of a simplified airplane assembly task. We demonstrate the algorithm in a study where novice ($n = 28$) and expert ($n = 2$) eye movements were recorded in successive trials ($n = 8$), allowing us to verify whether similar patterns develop with increasing expertise. In the proposed approach, AOI sequences were transformed to string representation and processed using the k -mer method, a well-known method from the field of computational biology. Our results for expertise development suggest that basic tendencies are visible in traditional ET metrics, such as the *fixation duration*, but are much more evident for k -mers of $k > 2$. With increased on-task experience, the appearance of expert k -mer patterns in novice gaze sequences was shown to increase significantly ($p < 0.001$). The results illustrate that the multi-trial k -mer approach is suitable for revealing specific cognitive processes and can quantify learning progress using gaze patterns that include both spatial and temporal information, which could provide a valuable tool for novice training and expert assessment.

Keywords Eye-tracking · Expertise development · Gaze patterns · Novice training · Sequence analysis

Introduction

Advances in the technology of eye tracking (ET) have provided us with a deeper understanding of specific cognitive processes and the development of perceptual expertise. Using traditional ET metrics, the majority of studies have focused on the analysis of visual expertise by investigating the eye movements of individuals in a wide range of domains, such as teaching (McIntyre & Foulsham, 2018a), medicine (Castner et al., 2020; Fox & Faulkne-Jones, 2017; van der Gijp et al., 2017) or aviation (Haslbeck & Zhang, 2017). Traditionally, these studies use ET as summary statistics, such as fixation duration, dwell time duration, fixation count, or time to first fixation to analyze expertise specific gaze behavior (Cristino et al., 2010; Kanan et al., 2015; Ooms et al.,

2012). These metrics, however, only capture the simplest of temporal information, while neglecting the temporal and spatial context of eye movements (Cristino et al., 2010). Studies using summary statistics have often generalized their findings, claiming that experts experience fewer fixations and shorter fixation durations, even though it has been established that expertise is highly domain-specific (Chi, 2006). Jarodzka and Boshuizen (2017) concluded that the reported measures are too reductionist to capture interesting insights into the nature of task- and stimuli-specific expert behaviors.

Researchers have since looked increasingly beyond traditional ET measures and have studied expertise by including the sequential information of gaze sequences. Sequential analysis has been shown to reveal a learner's overall learning behavior (Tsai et al., 2012) and that in-depth analysis of individuals' cognitive strategies during goal-oriented tasks (Hou et al., 2009) can be achieved. Moreover, Day et al. (2018) found that studies on expert and novice gaze behavior could benefit from identifying patterns that appear with a higher likelihood in a specific expertise group. Consequently, the comparison of gaze sequences using measures of string

✉ Felix S. Wang
wangfe@ethz.ch

¹ ETH Zurich, Leonhardstrasse 21, 8092 Zurich, Switzerland

similarity, such as the Levenstein distance (Levenshtein, 1966), has successfully advanced our understanding of expertise development by showing that experts' eye movements are more similar to each other than to those of lower expertise levels (Kanan et al., 2015; Tien et al., 2015; Watalingam et al., 2017). In these so-called *string-edit* approaches, such as ScanMatch (Cristino et al., 2010), a gaze sequence is converted into a string of letters, where each letter is assigned to a fixation onto a specific object or area of interest (AOI), thus conserving the temporal and spatial fixation information. Two sequences are then compared by counting the number of operations that are needed to convert one sequence into the other, which provides the similarity score (Anderson, Anderson, Kingstone, & Bischof, 2015a).

However, for the analysis of perceptual expertise development, two central aspects have yet to be considered. First, although the commonly applied string-edit approaches consider sequential gaze information during sequence comparison, the spatial and the temporal object relationship is lost after the similarity score calculation (Cristino et al., 2010; Day et al., 2018). Thus, after showing that differences between experts and novices exist, the question remains as to which task-specific gaze sequences can be consistently measured in expert eye movements and whether a suitable ET metric can be found that allows quantification of visual expertise development. Second, in the overwhelming majority of studies that investigate the development of expert gaze behavior, it is common practice that a single dataset of participants from different stages of development is recorded and that the gaze sequences of each group are then compared to specify expertise-related similarities (Castner et al., 2018; Eivazi et al., 2012). While this approach allows the analysis of how experts' and novices' eye movements differ at the specific time of the study, it stems from the assumption that all novices will eventually develop gaze sequences that are very similar to those of experts, with increasing training. However, to be able to truly reveal which behaviors are developed during training, multiple measurements on the same participants should be considered. Only then can we understand what behaviors drive the development of expert behavior within an individual under stable learning conditions (Gegenfurtner, 2013). Because of these limitations, a different approach to analyze gaze sequences is needed that can on one hand reveal task-specific gaze patterns and, on the other, allow the quantification of the development of expertise over time. In the field of computational biology, algorithms for quantitative DNA and protein sequence comparison approaches have been developed and optimized for decades. One well-established approach is *k*-mer analysis, which counts the frequency of subsequence patterns of neighboring elements with length *k* (2-mer, 3-mer, 4-mer, etc.) (Manekar & Sathe, 2018). The comparison of subsequences preserves the information of the

composition of said sequences, which, in this case, is the sequence of letters representing fixations on AOIs.

In this article, we demonstrate the ability of a *k*-mer pattern approach to determine the development of expertise, by applying the algorithm to object-related gaze data of an ET study involving a natural handling task. By analyzing the sub-pattern frequencies over multiple successive trials of the same expert and novice subjects, we show that *k*-mer patterns can provide a suitable metric to quantify the development of task-related expertise.

Related works

This section aims to highlight the previously conducted research in the field of visual expertise and sequence comparison approaches, and provides an overview of the *k*-mer analysis approach.

Eye movements in the study of visual expertise

Eye tracking (ET) has established itself as a popular research tool for the study of behavioral patterns (Duchowski, 2017; Land & Hayhoe, 2001) and, due to easier accessibility of the technology, has been increasingly applied to investigate visual expertise and expertise development (Brunyé et al., 2014; Crowe et al., 2018; Kelly et al., 2016). Particularly in the field of medicine, it has been of increasing interest what constitutes expertise, to increase the effectiveness and efficiency of novice training and diagnostic accuracy (van der Gijp et al., 2017). Using ET summary statistics, Wood et al. (2013) found that during the interpretation of skeletal radiographs, experts when compared to novices exhibit shorter fixation durations and are faster to fixate on the site of the fracture. In a study on laparoscopic skill acquisition, Wilson et al. (2011) discovered that experts show more fixations on task-relevant areas, while Gegenfurtner et al. (2011) showed that experts have longer saccades and, again, shorter time to fixate on task-relevant information. Zimmermann et al. (2020) measured fewer AOI transitions between task-critical objects during expert trials compared to novices during a cardiovascular intervention. Conversely, other studies have reported that novices, not experts, focused more of their attention on the surgical task (Zheng et al., 2011) and that experts visited fewer task-relevant areas (Jaarsma et al., 2015). In their review on ET study results for visual diagnostic performance in radiology, van der Gijp et al. (2017) found conflicting results regarding the relationship between the level of expertise and ET summary statistics. While in all studies the number of fixations seems to decrease with high levels of expertise, no generalization could be made on AOI fixation durations, the number of fixations on AOIs, dwell time ratios, saccade lengths, or image coverage.

Van der Gijp's results coincide with our knowledge that expertise is highly domain-specific (Beck et al., 2013; Chi, 2006; Sheridan & Reingold, 2017) and that results based on traditional ET summary statistics cannot and should not be generalized (Fox & Faulkne, 2017; Jarodzka & Boshuizen, 2017). Hence, in order to reveal more in-depth insights into the nature of perceptual expertise development, we are faced with the challenge of finding eye movement-based metrics that can help uncover task-specific, behavior-based development of expertise, while being generally applicable to a wide range of domains.

String-edit approaches

First introduced by Noton and Stark (1971), the scanpath theory postulates that fixed viewing sequences are generated top-down as a result of the specific model of a subject. Using a string editing approach, Privitera and Stark (2000) were the first to achieve scanpath comparison that compared both the temporal and the spatial information of fixations. In *string-edit* approaches, gaze sequences are converted into strings of letters, where each fixation of a different AOI is assigned a specific alphabetical character (Anderson, Anderson, Kingstone, & Bischof, 2015a). Furthermore, additional information about the length of the fixation can be included by repeating a letter based on the fixation duration (Cristino et al., 2010). Finally, by counting the number of operations needed to convert one sequence into the other, by using for example the Levenshtein distance (Levenshtein, 1966), a score is calculated to assess the similarity between eye movements in the context of a task (Foulsham et al., 2012).

One algorithm that was successfully adapted from computational biology to eye movement analysis is the Needleman-Wunsch algorithm (Kübler et al., 2017). Compared to the traditional *string-edit* approach, this algorithm allows local alignments between matching AOI patterns of two scanpath sequences.

In over two decades, various algorithms have been proposed to further improve gaze sequence comparisons, such as MultiMatch (Dewhurst et al., 2012), SubsMatch 2.0 (Kübler et al., 2017), EyeMSA (Burch et al., 2018) or ScanMatch (Cristino et al., 2010). MultiMatch compares the similarity of scanpaths as geometric vectors, including measures of saccade length and direction, without needing to couple ET data to predefined AOIs (Dewhurst et al., 2012; Jarodzka et al., 2010). SubsMatch 2.0 classifies eye movements between groups based on k -mer subsequences, while EyeMSA allows pairwise and multiple sequence alignments. For an in-depth description see Nicola C. Anderson's et al. (2015a, b) review on scanpath comparison methods.

In the context of expertise development, the majority of ET studies have applied gaze sequence similarity in the following two ways: the evaluation of experience-related eye movement

similarities and the classification of the expertise level based on the gaze sequence. McIntyre and Foulsham (2018b) have successfully shown that the gaze sequences of subjects, within the same level of expertise, are more similar than between subjects of different expertise groups. Castner et al. (2020) proposed a model for scanpath classification that is capable of extracting expertise-specific gaze behavior during a study of dental radiograph inspection.

However, as previously mentioned in the introduction, these approaches have some known limitations. One of the biggest limitations, next to the high computational cost of pairwise sequence comparison, is that similarity calculation is an essentially reductionist approach that reduces gaze behavior to a single cumulative score. While many measures of similarity can be used as a metric to determine behavioral differences between groups (Fahimi & Bruce, 2020), it does not allow one to infer which gaze sequences are developed during the evolution of novices to experts. Measuring similarity over time would indicate that individuals behave more similarly to experts, but the question would remain as to which of the gaze sequences changed during training.

Therefore, a metric is sought that, firstly, keeps the contextual temporal and spatial information of a specific task or domain, while, secondly, allows quantitative measurement of gaze patterns and, thirdly, enables one to infer the level of expertise development. Here, we propose to apply k -mer analysis object, or AOI, related gaze sequence patterns.

k -mers

In the field of computational biology, it is a common approach to identify similarity relationships between DNA sequences, with the goal to gain a fundamental understanding of how biological organisms function (Liu et al., 2013). Next to the Needleman–Wunsch algorithm, k -mer analysis has established itself as a simple but effective sequence analysis method (Ren et al., 2017). Compared to sequence alignment and string-edit approaches, the k -mer method segments each sequence into subsequences of length k and counts their appearance within the sequence. Hence, sequences can be compared based on the k -mer count of each pattern, while the individual components that are contained within each subsequence are conserved. In DNA analysis, each DNA sequence is regarded as a string with four letters (A, G, C, T), and the choice of k determines the number of possible combinations, with $no. \text{ of combinations} = no. \text{ of AOIs}^k$ (Manekar & Sathe, 2018). Because k -mers can be applied to all sequences in character representation, they can be applied to gaze sequences in the commonly used string-edit form. Bulling et al. (2013) have applied k -mers to electrooculography (EOG) signals to recognize high-level contextual cues, and Elbattah et al. (2020) have used k -mers to describe sequence

patterns of fixations and saccades to assist the automated diagnosis of autism.

In the present study, we have used higher-level ET data that was created using fixation-to-object mapping. Each dwell on an AOI was assigned a specific letter. Consequently, each k -mer pattern both preserves the sequence of k successive looked-at AOIs and allows us to compare different expertise levels by evaluating the appearance count of frequently appearing patterns.

Method

Subjects

Thirty subjects, mostly students, participated in the study (16 male, 14 female, mean age \pm SD = 23.76 \pm 2.3 years). All participants reported normal or corrected-to-normal vision and no neurological conditions. Each participant provided informed consent prior to testing and received no monetary compensation.

Stimuli

As was mentioned in the introduction, it is desirable to record data of the same novices over multiple trials, in order to observe if the same behavioral patterns that are found in experts are developed with increasing experience. Moreover, the comparison of these behaviors across different experience levels will verify if specific tasks induce idiosyncratic gaze sequences. However, in most ET studies involving medical expertise, the investigated task can take years to master, which makes repeated data recording during the stages of development costly and, often due to time constraints of experts, unfeasible. Therefore, it was important to find a suitably simple stimulus which allows the training of task-native subjects and where the development of expert behaviors should be observable within few repetitions while being cognitively demanding. Additionally, the learning complexity should be easily adjustable. We decided to use the assembly of an aircraft model, consisting of Lego-like building blocks, as the study stimulus. In a study on learning curves, Robbins (2019) verified the suitability of such a building block assembly stimulus to portray typical learning behavior over eight repetitions.

To show that the proposed method produces comparable results for different complexity levels, ET data of two stimuli were recorded. Subjects were assigned to either the easier bicolor (NOV BC) or the more complex multicolor (NOV MC) stimulus. Figure 1 shows the aircraft model stimuli for each group. Figure 2a shows the study setup, with three separate task-relevant areas: In front of the subject, a building area was separated by duct tape (A), a tray (B) containing the building blocks was placed to the right of the building area

and a digital assembly manual with step-by-step guidance (C) was placed behind the building area. Subjects of group NOV BC were provided with 27 blocks in two different colors (see Fig. 2b). To add the complexity, subjects of group NOV MC were instead given 54 blocks in ten different colors, of which 27 were used as distracting elements (see Fig. 2c).

Study setup

The study was conducted by dividing the subjects into two groups of 15 participants. Each group was given either the simple bicolor ($n = 15$) or the more complex multicolor stimulus ($n = 15$). Due to the absence of task-native experts, one subject of each group was trained to expert level as a reference for each stimulus (group EXP). The expert training was conducted one day before data recording and was carried out until subjects acquired the ability to finish the assembly repeatedly, without the use of the manual and without making any mistakes. All 30 participants, 28 novices and 2 experts, completed a set of eight successive trials, resulting in a total of 240 recorded assemblies with 240 individual gaze sequences. Each set of trials took approximately 15–25 min to complete (including instructions). The use of the manual was not strictly required; however, all subjects were instructed to respect the building sequence in the correct and color-sensitive order. The arrangement of building blocks in each group was consistent throughout all trials and, after the last building step, a new set of building blocks was provided for the next assembly.

Recording equipment

Eye tracking data were recorded using the SMI Eye Tracking Glasses 2 and evaluated using SMI's BeGaze version 3.6 software (SensoMotoric Instruments, Teltow, Germany). The mobile ET glasses record data at a sampling rate of 60 Hz and have a reported gaze position accuracy of 0.5°. The integrated scene camera was recorded with a resolution of 1280 \times 960 pixels at 30 frames per second. The audio was simultaneously recorded with the integrated microphone. Before the start of recording, each subject completed a three-point calibration process. The eye tracking ratio (tracking accuracy) was 97.8 \pm 1.2% averaged over all participants. No participant data had to be excluded from analysis due to insufficient data quality. We conducted the calibration using the SMI recording unit and a three-point calibration, where the wearer was asked to fixate on three specific markers (top-left corner, top-right corner, and the middle of the bottom edge). During each marker fixation, the experimenter manually confirmed these marker locations on a live-view of the scene camera on the recording device. Afterward, the experimenter made sure that both eyes were clearly visible on the eye camera recordings and calibration was validated using a three-point validation of specific

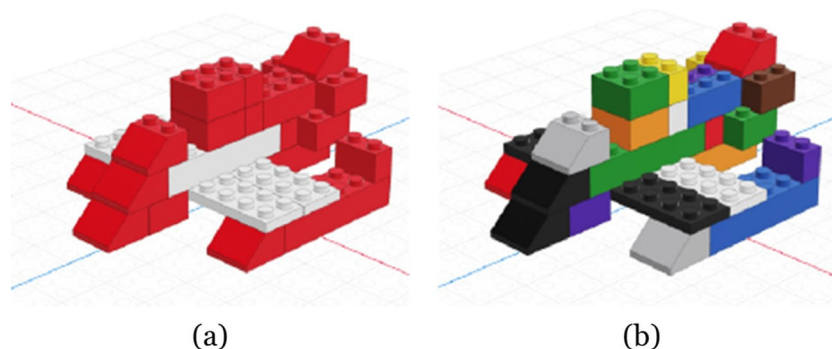


Fig. 1 Two building block aircraft models were chosen as assembly stimuli for **a** group NOV BC and **b** group NOV MC. Stimuli differed in the color combination of the model and the number of blocks to choose from

points within the task environment. If the calibration accuracy was not sufficient, calibration and validation were repeated.

k-mer sub-pattern analysis

Figure 3 shows the analysis approach for the extraction of *k*-mer gaze sequence patterns. As is the usual practice in ET analysis, the first step is to detect eye movement “events,” particularly eye fixations, defined as a moment in time where the eyes are relatively still at a given point of the visual field. For the fixation detection, we used SMI’s event detector algorithm with default settings (required fixation duration of 50 ms and a peak saccade velocity of 40°/s). Thus, we obtained a sequence of eye fixations for each recording, each fixation characterized by its starting time and duration. We then defined four task-related objects as AOIs for semantic mapping of each fixation. Semantic gaze mapping was conducted manually. Figure 2 shows the AOIs defined for data analysis, namely *building area* (A), *bricks* (B), and *manual* (C). All fixations that did not fall into these AOIs were assigned to white space (W). Using these specific letters, the gaze sequences of each task were converted into *string-edit* representation. Here, we chose the collapsed form of string representation, where successive fixations on the same AOI, also called dwell, are collapsed into a single letter. Consequently, the number of occurrences for each *k*-mer of length $k=1$ is equal to the AOI dwell count, whereas the occurrence of each *k*-mer of length $k=2$ is equal to the number of AOI transitions between two AOIs. Consequently,

the results for $k=1$ and $k=2$ were evaluated as traditional AOI metrics *dwell count* and *AOI transition*, along with the average *fixation duration* and *dwell time*.

The *k*-mer sub-pattern detection was carried out using MATLAB R2019a. For each scanpath, we applied MATLAB’s *nmercount* function to retrieve all *k*-mer patterns of length $k = \{1, 2, 3, 4\}$ in string vector form. Next, we applied the *count* function, with the *k*-mer vector entries as input, to count the appearance frequencies of each pattern. The *count* function was used to count the unique appearances of sequences. The appearance frequency was calculated in each of the eight consecutive trials, allowing the analysis of how these gaze sequences change while novice subjects develop expertise in the given task.

Furthermore, we investigated the specificity of gaze sequences concerning a certain level of expertise by determining the variety of *k*-mers that appeared throughout all trials, separately for experts and novices. Here, we calculated the average relative frequency of each pattern over all trials, for the analytical evaluation of predominant gaze patterns. The equation used is shown in Eq. 1.

Relative frequency of Pattern_i

$$= \frac{\text{Frequency of Pattern}_i}{\sum_{i=1}^{\text{Number of } k\text{-mers}} \text{Frequency of Pattern}_i} \quad (1)$$

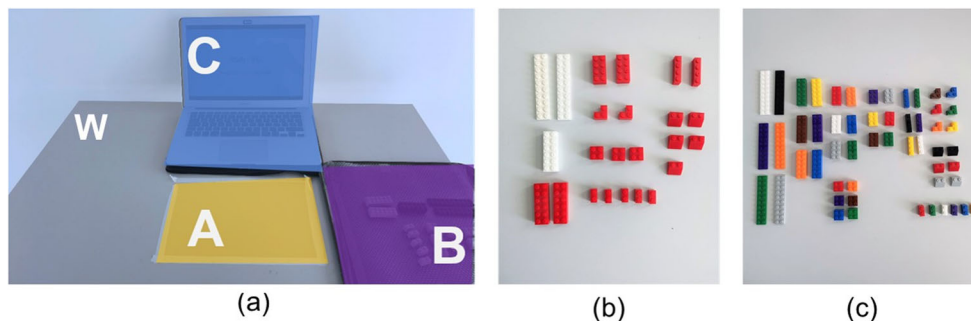


Fig. 2 Study setup with the AOIs chosen for data analysis (**a**) and placement of the building blocks for groups NOV BC (**b**) and NOV MC (**c**) at the start of each assembly. Fixations were assigned to AOIs (A) building area, (B) bricks, (C) manual, or (W) white space

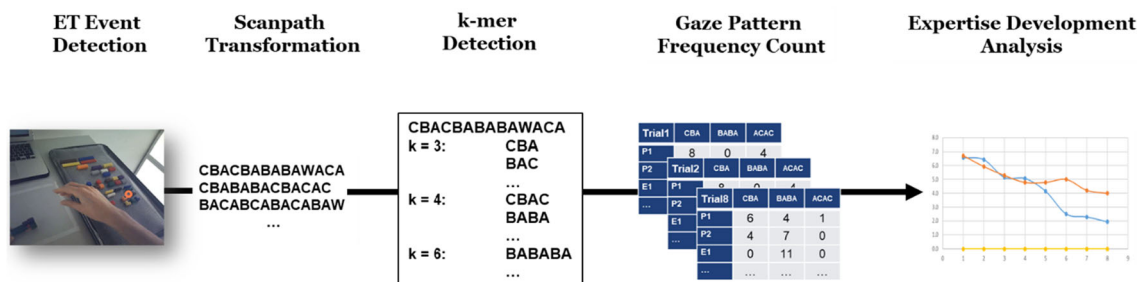


Fig. 3 Schematic of the *k-mer* analysis approach

The relative frequency of a pattern was calculated by dividing the average appearance frequency of each pattern by the sum of all pattern frequencies and expresses how dominant a pattern was in the overall gaze sequences. For our analysis, only patterns with a relative frequency of more than 1%, were deemed to be relevant gaze patterns and further considered for the analysis of the development throughout the trials.

Statistical analysis

For the statistical analysis, novice subjects were separated into two groups based on the complexity of their stimulus, group NOV BC or NOV MC. Expert subjects were not considered for statistical analysis due to the lack of significant participant numbers. The statistical analysis of the sequential gaze patterns was conducted using IBM SPSS Statistics 26 and RStudio 3.6.2. First, the data were tested for equal variances using Levene’s test of equal variances, with $p > 0.05$ as the threshold. Subsequently, changes in the appearance frequency of gaze patterns throughout an increasing number of trials were tested within groups using *t*-statistics of dependent variables. To test whether the stimulus complexity induced differences in pattern slopes and intercepts, an analysis of variance (ANOVA) of the aggregated summary statistics was applied between NOV BC and NOV MC. Due to the violation of normal distribution of data and the inequality of variances of some *k-mer* patterns, we used linear regression to estimate the slope and centered intercepts of the pattern frequencies of each participant. Differences in the slope between groups indicate different learning rates of *k-mer* patterns, while different intercepts indicate different pattern occurrences. The data were analyzed statistically, with an alpha of $\alpha = 0.05$. Additionally, summary statistics were calculated to investigate the change in pattern frequencies of each expertise group.

Results

Table 1 shows the measured mean *trial times* and observed *use errors* during the assembly for all participant groups. Novices that learn on a more complex stimulus took

significantly longer to complete the task ($p = 0.044$), but show a similar decline in the *trial times* with repeated assemblies ($p = 0.871$). The overall *use error* rate was zero for experts and below 1 for all novices.

Conventional AOI evaluation metrics

Table 2 shows the mean *fixation duration* along with the conventional AOI evaluation metrics *dwelt time*, *dwelt count* ($k = 1$), and *AOI transition* ($k = 2$). While the *fixation duration* seems to increase linearly for all participants (slope = 5.918, $p = 0.95$), the standard deviation is shown to be in a similar order of magnitude, which indicates a high variation between individual participants in each group. The *dwelt count* ($k = 1$), which is the number of gazes per AOI per task, changes similarly for novice subjects. This trend was shown for AOIs *building area* ($p = 0.081$) and *manual* ($p = 0.236$), regardless of whether the task was learned on the simpler or the more complex stimulus, but showed significant differences for AOI *bricks* ($p = 0.034$). Over time, novices look less at the *manual* (slope(C) = -4.287), while the number of times they look at AOIs *building area* (slope(A) = -2.132) and *bricks* (slope(B) = -0.489) reaches a nearly constant range after the third trial (see Table 2). Conversely, experts showed little change in the *dwelt count* for any of the three AOIs. A similar trend can be observed in the *dwelt time* on AOIs *building area*, *bricks*, and *manual*.

For $k = 2$, novices that learned on the simpler stimulus show lower overall pattern frequencies for the transitions *building area – manual* (AC), *manual – building area* (CA), and *bricks – manual* (BC), but higher average numbers for the transition *bricks – building area* (BA), compared to those participants that have learned on the more complex stimulus. With an increasing number of trials, both groups show a tendency to approach the constant number of *AOI transitions* that is observed for the expert group. Notably, only 2-mer pattern BA shows a marginal linear increase for novices in the appearance frequencies throughout eight trials (slope(BA) = 0.532, $p = 0.937$), while the other patterns have a declining tendency for patterns AC (slope(AC) = -3.049 , $p = 0.481$), CA (slope(CA) = -2.665 , $p = 0.074$), and CB (slope(CB) = -1.576 , $p = 0.663$).

Table 1 Results of the task completion time and the use errors

	Trial							
	1	2	3	4	5	6	7	8
EXP								
Task completion time [s]	111.8 ± 39.3	92.6 ± 35.3	83.7 ± 27.7	85.7 ± 30.8	77.8 ± 21.3	83.4 ± 25.3	79.0 ± 24.4	84.2 ± 22.1
Use errors [-]	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
NOV BC								
Task completion time [s]	159.9 ± 42.8	110.4 ± 21.6	104.1 ± 23.5	98.3 ± 19.5	93.8 ± 14.5	89.7 ± 16.2	89.6 ± 24.4	85.9 ± 23.2
Use Errors [-]	0.4 ± 0.8	0.4 ± 0.6	0.6 ± 1.1	0.1 ± 0.4	0.2 ± 0.6	0.2 ± 0.4	0.2 ± 0.6	0.2 ± 0.4
NOV MC								
Task completion time [s]	178.3 ± 40.8	131.2 ± 26.2	118.4 ± 22.6	113.2 ± 16.3	111.6 ± 17.1	108.9 ± 16.6	102.3 ± 20.3	104.8 ± 20.9
Use errors [-]	0.4 ± 0.8	0.4 ± 1.1	0 ± 0	0.2 ± 0.4	0.1 ± 0.4	0.3 ± 0.5	0.1 ± 0.4	0.1 ± 0.4

***k*-mer analysis of higher gaze sequences**

Next, we analyzed the *k*-mer sequences and the average relative frequency of each pattern for all trials. Here, the data is presented as mean ± standard error of the mean (SEM), unless otherwise noted.

Figure 4a and b shows the results of the evaluation of the most frequently appearing 3-mer sequences. In the gaze sequences of expert subjects, 7 of 24 possible *k*-mer patterns were measured with a relevant relative appearance frequency. Patterns BAB (40.69 ± 1.12%) and ABA (40.69 ± 1.12%) appeared most frequently. Therefore, more than 80% of the overall expert gaze behavior can be expressed through only two 3-mer patterns. In the gaze sequences of novice subjects, 12 of the 24 sequences appeared with a relevant relative frequency. Patterns ACB (16.76 ± 0.34%), BAC (13.46 ± 0.46%), CBA (12.84 ± 0.52%), and CAC (10.74 ± 1.19%) appeared most frequently. Consequently, more than 60% of the novice gaze behavior can be expressed through the five most frequently appearing 3-mer sequences. Further, more than 90% of the gaze sequence can be explained using the 10 most frequent *k*-mer patterns.

Figure 4c and d shows the results for the analysis of the most frequent 4-mer patterns. In the gaze sequences of expert subjects, again only 7 of 108 possible patterns showed a relevant relative frequency of over 1%. Sequences ABAB (38.24 ± 1.46%) and BABA (41.37 ± 1.42%) appeared to be predominant and showed a strong resemblance to the 3-mer patterns ABA and BAB. Again, approximately 80% of all the visual behavior of experts can be expressed through two *k*-mer gaze sequences. In novice eye movements, the fraction of 4-mer sequence patterns was more broadly distributed, showing 23 of 108 possible pattern appearances. The five most frequent sequences accounted for 37.2% and the 10 most frequent sequences for 65.6% of the gaze sequences of novice subjects.

Semantic meaning of object-related gaze patterns

Analyzing the results of conventional ET metrics, we see that experts mainly focused on AOIs *bricks* and *building area*, while the visual attention of novices was shared equally between all three task-relevant AOIs. From the bar plots, it was extracted that the predominant patterns of experts differ from those of the novices. Additionally, the same 3-mer sequence often appears in several 4-mer sequences, i.e. pattern ACB is included in ACBA, BACB, and CACB. Therefore, for the subsequent expertise development analysis, we chose to investigate *k*-mer patterns ACB, ACAC, BABA, and BABABA. Given that with increasing *k* the probability of the appearance of *k*-mer patterns due to chance is reduced, we have included gaze sequence BABABA, a 6-mer extension of BABA, in the analysis. To illustrate the gaze sequence in the study environment, each pattern is depicted as a gaze path in Fig. 5.

By analyzing each *k*-mer sequence pattern along with the ET recordings, we extracted the following semantic meanings: ACB is a triangular pattern between *building area*, *manual*, and *bricks*. Subjects who repeatedly exhibited this pattern assembled the brick in the area, looked at the assembly manual for the next building step, and finally grabbed the associated brick on the tray. Pattern sequence ACAC was observed in moments when the subject discovered a building error, resulting in multiple glances between the assembled piece in the building area and the assembly manual. Finally, the 4-mer and 6-mer patterns BABA and BABABA represent repeated consecutive gaze transitions between AOIs *bricks* and *building area*. This behavior was shown when subjects became increasingly familiar with the task and were able to carry out several building steps without the use of the manual. Here, participants' gaze movements were followed by a reach for one building brick and an assembly in the building area.

Table 2 Averages and standard deviations for traditional eye tracking metrics fixation duration, dwell time, dwell count (k -mers of $k = 1$), and AOI transitions (k -mers of $k = 2$)

ET Metric	k -mer	Trial								
		1	2	3	4	5	6	7	8	
EXP	Fixation duration [ms] Dwell time [s]	A	320.4 ± 271.2	344.1 ± 286.7	337.2 ± 293.1	325.0 ± 253.9	340.7 ± 301.0	361.5 ± 322.3	320.4 ± 271.2	320.4 ± 271.2
		B	41.6 ± 7.0	40.2 ± 8.2	33.7 ± 4.6	35.8 ± 6.4	35.0 ± 5.0	35.1 ± 4.9	33.8 ± 3.4	39.3 ± 3.7
		C	21.0 ± 4.7	18.2 ± 6.5	15.8 ± 4.1	16.4 ± 6.4	17.6 ± 6.3	17.5 ± 8.1	15.8 ± 6.3	18.5 ± 7.1
	Dwell count ($k=1$) [-]	A	5.4 ± 4.2	4.9 ± 1.4	5.8 ± 0.6	6.4 ± 0.5	8.0 ± 1.9	8.4 ± 0.2	6.1 ± 3.1	6.5 ± 0.8
		B	24.0 ± 4.0	20.5 ± 0.5	21.5 ± 2.5	21.5 ± 1.5	21.0 ± 1.0	23.0 ± 3.0	22.5 ± 2.5	22.5 ± 0.5
		C	23.5 ± 3.5	20.5 ± 0.5	21.0 ± 1.5	21.5 ± 1.0	21.0 ± 2.0	22.0 ± 2.0	22.0 ± 2.0	23.5 ± 0.5
	AOI Transitions ($k=2$) [-]	A	5.0 ± 3.0	2.5 ± 0.5	2.5 ± 1.5	3.0 ± 0.0	2.0 ± 0.5	3.5 ± 0.5	2.5 ± 0.5	1.0 ± 0.0
		B	22.5 ± 2.5	20.5 ± 0.5	21.0 ± 2.0	21.5 ± 1.5	21.0 ± 1.0	22.0 ± 2.0	22.0 ± 2.0	22.5 ± 0.5
		AC	2.5 ± 1.5	1.5 ± 0.5	1.5 ± 0.5	1.5 ± 0.5	1.0 ± 0.0	2.0 ± 1.0	1.5 ± 0.5	1.0 ± 0.0
	NOV BC	A	1.5 ± 1.5	0.0 ± 0.0	0.5 ± 0.5	0.0 ± 0.0	0.0 ± 0.0	1.0 ± 1.0	0.5 ± 0.5	0.0 ± 0.0
		B	1.5 ± 1.5	0.5 ± 0.5	0.0 ± 0.0	1.0 ± 1.0	0.0 ± 0.0	0.5 ± 0.5	0.0 ± 0.5	0.0 ± 0.0
		CB	332.9 ± 316.2	329.2 ± 314.2	332.3 ± 342.7	348.5 ± 339.9	349.7 ± 334.0	353.3 ± 341.1	359.4 ± 348.9	370.0 ± 340.2
NOV MC	Fixation duration [ms] Dwell time [s]	A	70.0 ± 23.4	53.1 ± 14.1	54.5 ± 17.3	52.3 ± 14.2	48.6 ± 9.5	47.7 ± 12.2	48.8 ± 14.6	47.4 ± 17.0
		B	13.5 ± 3.4	13.5 ± 3.3	14.3 ± 4.0	14.7 ± 5.2	14.7 ± 4.7	14.4 ± 3.9	15.6 ± 6.2	14.8 ± 5.2
		C	50.6 ± 19.1	26.4 ± 7.6	22.0 ± 5.7	18.6 ± 3.2	16.4 ± 2.9	13.8 ± 5.3	12.3 ± 5.2	11.9 ± 5.3
	Dwell count ($k=1$) [-]	A	35.4 ± 10.5	23.3 ± 4.3	22.3 ± 5.5	20.4 ± 3.5	19.9 ± 3.3	19.8 ± 2.8	19.6 ± 3.5	19.4 ± 4.2
		B	21.6 ± 3.4	19.2 ± 3.2	19.9 ± 4.5	19.0 ± 2.7	18.9 ± 2.8	18.6 ± 2.8	19.3 ± 2.8	19.3 ± 3.3
		C	42.6 ± 11.7	28.4 ± 5.3	23.8 ± 7.2	19.2 ± 4.9	18.3 ± 6.0	14.9 ± 7.2	12.9 ± 7.8	10.7 ± 7.1
	AOI Transitions ($k=2$) [-]	BA	12.3 ± 5.0	12.6 ± 4.8	13.6 ± 3.8	14.9 ± 3.2	14.4 ± 3.9	15.0 ± 4.3	15.8 ± 4.1	16.0 ± 4.7
		AC	31.1 ± 10.6	19.8 ± 4.1	16.4 ± 4.7	13.6 ± 4.7	12.2 ± 4.4	9.8 ± 4.5	8.6 ± 5.3	7.6 ± 5.2
		CA	23.1 ± 11.4	10.7 ± 5.2	8.6 ± 5.0	5.4 ± 2.4	5.5 ± 2.3	4.7 ± 3.1	3.7 ± 3.5	3.4 ± 3.0
	NOV MC	A	17.6 ± 1.4	16.0 ± 2.1	13.9 ± 3.2	12.1 ± 4.1	11.1 ± 4.3	8.5 ± 4.5	7.5 ± 4.9	6.3 ± 2.0
		B	287.8 ± 274.3	295.2 ± 278.0	299.3 ± 276.4	311.0 ± 310.2	312.9 ± 306.6	319.7 ± 313.0	320.6 ± 307.4	325.3 ± 311.9
		C	70.8 ± 23.3	63.8 ± 22.3	52.7 ± 13.4	51.8 ± 10.9	50.0 ± 12.6	50.3 ± 11.1	47.8 ± 13.0	50.3 ± 14.2
NOV MC	Fixation duration [ms] Dwell time [s]	A	25.8 ± 4.9	21.5 ± 4.1	20.2 ± 2.2	20.4 ± 4.4	19.5 ± 4.0	19.7 ± 4.7	19.4 ± 3.6	18.6 ± 3.3
		B	52.3 ± 10.9	34.5 ± 18.2	23.4 ± 6.2	21.0 ± 5.6	20.3 ± 5.3	19.5 ± 6.4	16.2 ± 4.9	16.3 ± 5.8
		C	42.9 ± 11.3	34.6 ± 17.8	26.9 ± 5.9	26.0 ± 5.6	23.6 ± 4.7	23.6 ± 4.0	21.6 ± 4.5	22.6 ± 4.2
	Dwell count ($k=1$) [-]	A	26.5 ± 2.4	23.1 ± 5.4	20.2 ± 2.6	20.7 ± 2.1	19.9 ± 1.6	20.3 ± 2.3	19.4 ± 2.1	20.4 ± 2.6
		B	58.6 ± 12.1	45.6 ± 19.1	34.0 ± 9.2	32.4 ± 10.5	27.9 ± 9.2	27.3 ± 10.8	22.9 ± 10.6	22.4 ± 10.5
		C	9.6 ± 4.9	9.7 ± 4.3	10.9 ± 3.8	10.9 ± 4.1	11.7 ± 4.2	12.5 ± 3.8	12.7 ± 4.0	12.9 ± 5.4
	AOI Transitions ($k=2$) [-]	BA	39.9 ± 10.9	30.4 ± 16.1	22.7 ± 6.2	20.1 ± 6.9	18.0 ± 6.5	17.2 ± 7.7	14.7 ± 7.0	15.1 ± 6.7
		AC	33.4 ± 11.6	24.7 ± 16.4	16.0 ± 7.8	15.1 ± 7.8	11.8 ± 6.0	11.0 ± 5.9	8.9 ± 6.1	9.6 ± 5.9
		CB	23.8 ± 3.3	19.5 ± 4.0	16.4 ± 3.5	15.4 ± 3.6	14.3 ± 4.2	14.2 ± 6.2	12.5 ± 6.0	11.9 ± 5.8

AOI building area is transformed to letter A, AOI bricks to letter B, and AOI manual to letter C. Equally, AOI transition bricks – building area is given as BA

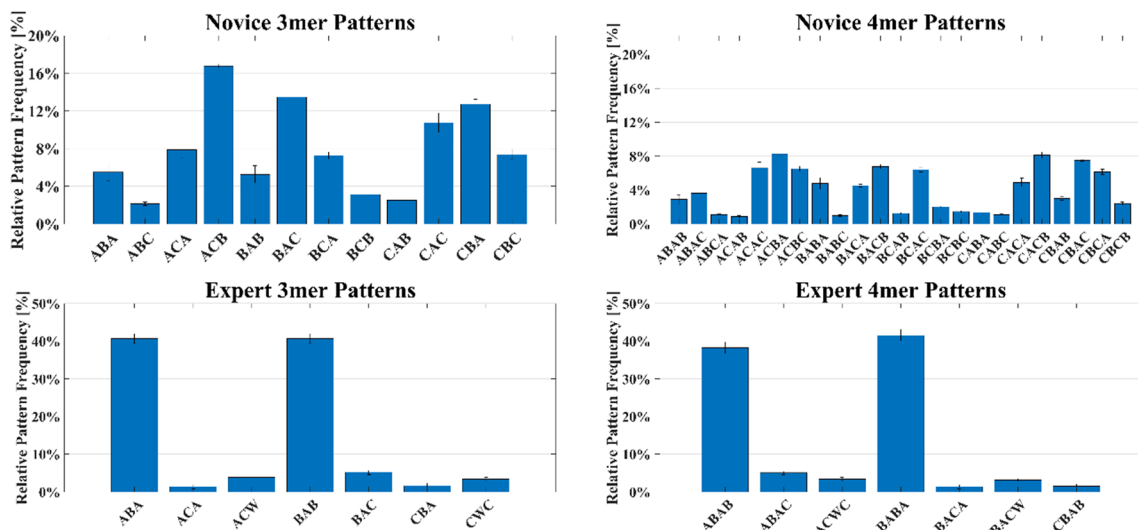


Fig. 4 Bar plots displaying k -mer patterns with a mean relative appearance frequency of $>1\%$ over all trials for $k = \{3, 4\}$

Consequently, in the subsequent paragraphs, 3-mer pattern ACB will be referenced as *monitor assembly*, 4-mer pattern ACAC as *consult manual*, BABA as *familiarizing steps*, and 6-mer pattern BABABA as *internalized steps*.

Gaze pattern frequency

Next, we examined how the frequency of selected larger k -mer gaze patterns, i.e. $k = \{3, 4, 6\}$, have developed over repeated assemblies. Figure 6 shows the average appearance frequencies per trial for experts and novices. As expected, expert subjects did not show a learning curve for any pattern, but rather constant pattern appearance frequencies. Patterns *monitor assembly* or *consult manual* were absent for most expert trials, whereas patterns *familiarizing steps* and *internalized steps* appeared with constantly high frequencies. For novice subjects, the curve of the pattern *consult manual* resembles that of a traditional learning curve, where the decrease of pattern frequencies between trials becomes smaller with increasing practice. Conversely, the curves of patterns *familiarizing steps* and *internalized steps* indicate linear growth, while the curve of pattern *monitor assembly* indicates linear decline. Notably, the frequency of each of the four gaze patterns converges towards the expert reference appearance frequency count with increasing task practice.

When comparing the k -mer frequencies of k s of different sizes, the results for $k = 1$ showed a similar *dwelt count* on AOI *bricks* (B) for novice and expert subjects over eight trials. For $k = 2$, a difference in gaze strategies between the two expertise levels, for example for AOI *transition bricks – building area* (BA), becomes more evident. For even larger values of k , constant appearances of expert gaze patterns such as 4-mer pattern *familiarizing steps* (BABA) and 6-mer pattern *internalized steps* (BABABA), can be measured less and less frequently during novice trials. As these pattern frequencies are close to zero when novice subjects start their learning progress, but increase over time, learning of this specific expert gaze strategy over time is inferred.

Table 3 shows the mean appearance frequency and SEM, for the first and the last trial, as well as the results of a statistical comparison of the pattern frequencies between the first and last trial, using a t -test for dependent variables. All subjects showed highly significant differences for all analyzed k -mer gaze patterns. The effect sizes greatly exceed Gignac and Szodorai (2016) reported value of $r = 0.3$ for large effects. Thus, within each novice group, the sequence-pattern-based gaze behavior changes significantly from the first to the last assembly.

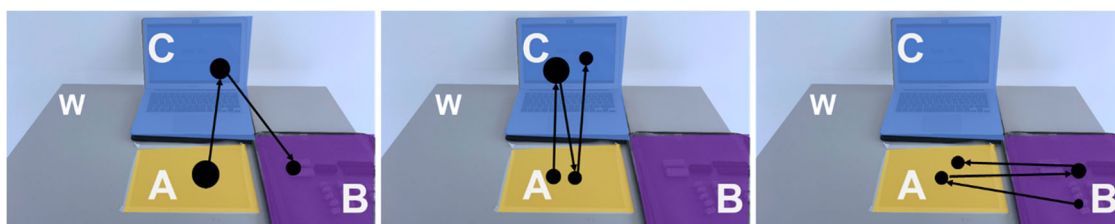


Fig. 5 Visualization of the four sequence gaze patterns, ACB (a), ACAC (b), and BABA and BABABA (c) on the study setup

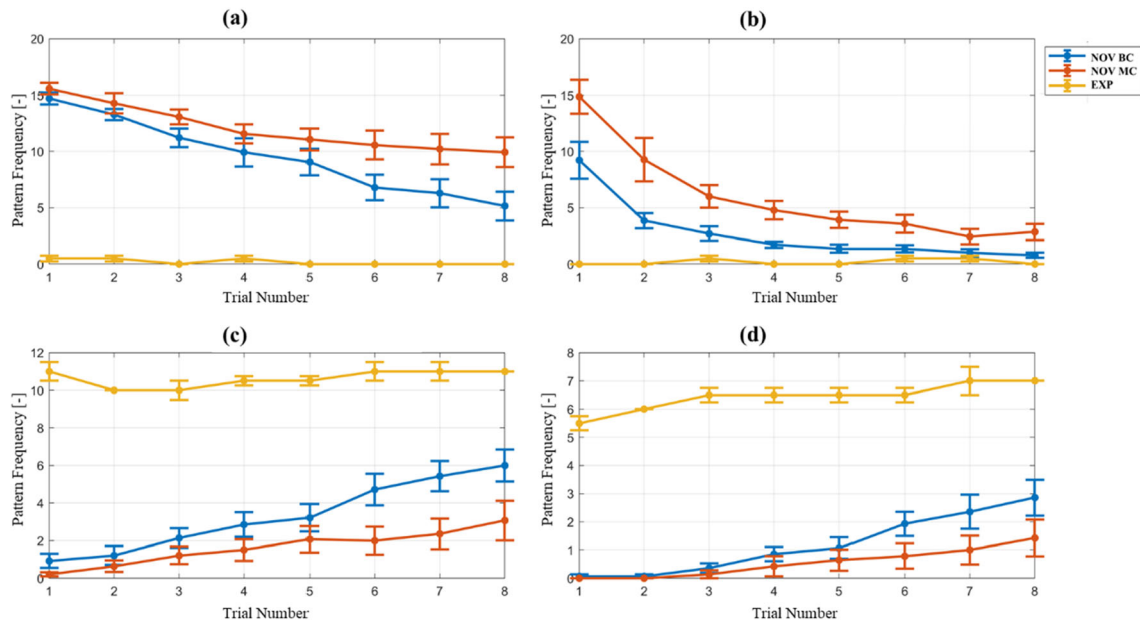


Fig. 6 The mean and SEM for patterns *a* ACB, *b* ACAC, *c* BABA, and *d* BABABA are shown over eight successive assembly trials

The effect of stimulus complexity on expertise development

Finally, we investigated the effect of the stimulus complexity on the development rate of perceptual expertise, using a statistical *k*-mer pattern frequency analysis. The slopes and intercepts of the change in the appearance frequencies of the four

investigated patterns were compared between novices that learned on the simple (Nov BC) and the more complex (Nov MC) assembly stimulus (see Fig. 7). The null hypothesis considers the slopes of the gaze patterns over time to be parallel, resulting in equal development rates of patterns over time.

The results of the statistical ANOVA test are shown in Table 4. Novices of group NOV BC not only began to focus

Table 3 Results of the analysis of mean pattern frequencies between the *first* and the *eighth* trial of patterns ACB, ACAC, BABA, and BABABA

<i>k</i> -mer Patterns	Trial 1		Trial 8		<i>t</i> (13)	<i>p</i>	<i>r</i>
	M	SEM	M	SEM			
ACB							
Group NOV BC	14.71	0.52	5.14	1.28	8.143	< 10 ^{-3**}	0.914
Group NOV MC	15.57	0.50	9.93	0.132	4.261	.001*	0.763
Group EXP	0.50	0.25	0.00	0.00	-	-	-
ACAC							
Group NOV BC	9.21	1.63	0.79	0.21	5.282	< 10 ^{-3**}	0.825
Group NOV MC	14.86	1.50	2.86	0.73	9.474	< 10 ^{-3**}	0.934
Group EXP	0.00	0.00	0.00	0.00	-	-	-
BABA							
Group NOV BC	0.93	0.37	6.00	0.85	-6.679	< 10 ^{-3**}	0.879
Group NOV MC	0.21	0.11	3.07	1.05	-2.624	.021*	0.588
Group EXP	11.0	0.50	11.0	0.00	-	-	-
BABABA							
Group NOV BC	0.07	0.07	2.86	0.64	-4.545	.001*	0.783
Group NOV MC	0.00	0.00	1.43	0.65	-2.190	.047*	0.519
Group EXP	5.50	0.25	7.00	0.00	-	-	-

A *t*-test for paired samples was applied on novice subjects separately for groups NOV BC and NOV MC

****p* < 0.001, ***p* < 0.01, **p* < 0.05

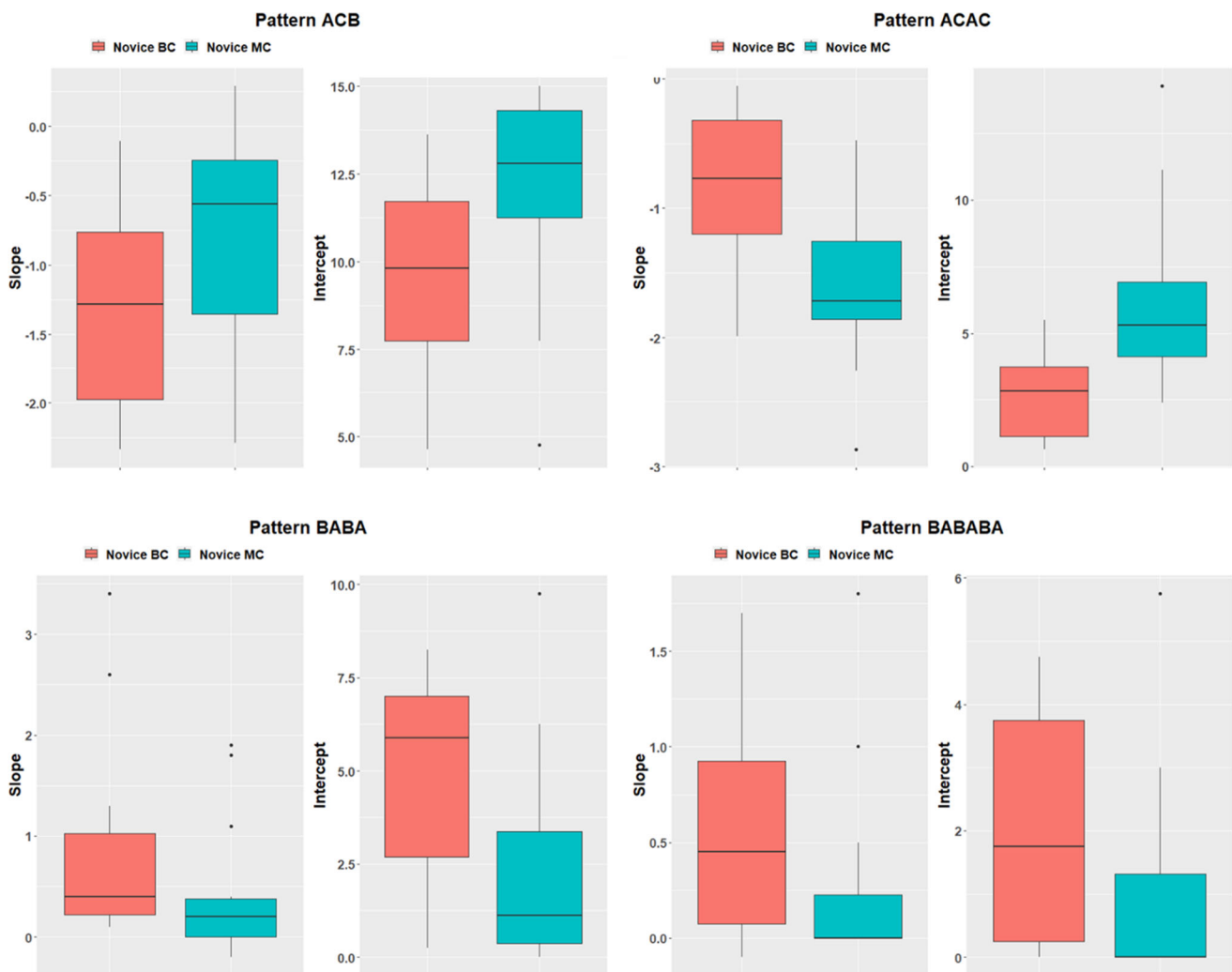


Fig. 7 Box plots of the mean slope of pattern frequencies over eight trials. The slopes indicate the rates at which k -mer gaze patterns develop over time, while the intercept show the average occurrence of the pattern in each group

their attention significantly faster between AOIs *bricks* and *building area* ($p = 0.0487$), using the k -mer gaze pattern *familiarizing steps*, but this expert pattern was also observed with a significantly higher appearance frequency ($p = 0.049$). Additionally, learning on a more complex stimulus led to a less significant decrease in gaze behavior *consult manual* ($p < 0.001$), indicating a higher dependence on the task manual during the trials. The decrease in the gaze pattern *monitor assembly* indicates no difference in either stimulus ($p = 0.543$) but shows that a simpler task led to fewer instances where the assembly was monitored ($p = 0.0374$). Similarly, novice participants developed the expert k -mer pattern *internalized steps* at the same rate for eight trials ($p = 0.119$) with similar appearance frequencies ($p = 0.119$), regardless of the complexity of the stimulus. Overall, the difference between the complexities of the investigated stimuli did not change the rate at which patterns were developed during the assembly for most

patterns, but was shown to influence the average appearance of gaze patterns depending on which stimuli task was learned on.

Discussion

In the context of the investigated task, our results for the quantitative analysis of expertise development using the k -mer approach led us to the following conclusions:

First, the use of the proposed k -mer approach allowed us to investigate the differences in perceptual expertise, by showing that the most common expert and novice fixation sequences differ in their specific AOI composition. By calculating the relative appearance frequency of these patterns, we were able to show that the gaze behavior, in the context of this assembly task, was dominated by only a few k -mer patterns. The greater

Table 4 Results of an ANOVA test of the slopes and intercepts between novices groups for the evaluation of the influence of stimulus complexity on gaze pattern development for k-mer patterns ACB, ACAC, BABA, and BABABA, with $\alpha = 0.05$

Pattern	Global mean	df	Sum Sq	Mean Sq	F value	P
ACB						
Slope	-0.804	1	2.313	2.3130	4.061	.0543
Intercept	8.634	1	43.13	43.13	4.812	.0374*
ACAC						
Slope	-0.321	1	4.626	4.626	13.7	< 10 ⁻³ ***
Intercept	2.161	1	86.63	86.63	14.4	< 10 ⁻³ ***
BABA						
Slope	0.614	1	.989	.9888	4.277	.0487*
Intercept	3.607	1	17.68	17.681	4.267	.049*
BABABA						
Slope	0.418	1	.347	.3474	2.599	.119
Intercept	1.509	1	2.893	2.893	2.606	.119

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

variety of AOI k -mer sequences in novice patterns aligns with findings of previous research conducted by Castner et al. (2020), who found that novices experience a larger variety of gaze strategies when compared to experts.

Second, the comparison with traditional AOI evaluation methods showed that the *dwell time* and the *dwell count* ($k = 1$) reveal some learning tendencies regarding the change in novice subjects' visual attention over time, while the analysis of the *fixation duration* was shown to be inconclusive. Learning behavior was shown to be much more evident for patterns of $k \geq 2$. Through the introduction of a multi-trial study setup, we were able to measure how the gaze behavior of novice participants changed with increasing task familiarity. Compared to studies that have used ET summary statistics as generalizable findings for expert behavior (Jaarsma et al., 2015; Wood et al., 2013), the proposed multi-trial k -mer approach could be used as a methodological framework, which reveals those particular gaze strategies that are attributed to a specific level of expertise. Furthermore, our results provide strong evidence that a learning behavior can be measured in the changes in k -mer pattern appearances.

Third, the extract gaze behaviors confirmed expected task-related results, showing that experts do not use the manual, while novices learn to use it less over time. Even though the tendency to rely less on the AOI *manual* could be shown in the *dwell time*, the use of larger k -mers allowed us to gain additional semantic insights, such as the use of gaze patterns *monitor assembly* and *consult manual* during task-related moments of hesitancy, as well as the level of developed task familiarity through the increased use of gaze behaviors *familiarizing steps* and *internalized steps*. Admittedly, it might

make little sense to teach students to look less at the manual during training scenarios. However, in a domain such as cardiovascular intervention, where expert surgeons were shown to exhibit fewer gaze transitions than novices between task-related AOIs (Zimmermann et al., 2020), these gaze patterns could be used to assess the current stage of expertise of each subject.

Fourth, using the k -mer approach, we were able to quantify both traditional ET metrics (k -mer = {1, 2}) and higher-level gaze patterns (k -mer = {3, 4, 6}). The consistency of expert pattern appearances resembled a learning plateau, which, as mentioned by Khan et al. (2014), can act as a strong gaze movement-based indicator for expert behavior. Novice subjects showed either an increasing (patterns *familiarizing steps* and *internalized steps*) or decreasing (patterns *monitor assembly* and *consult manual*) pattern development towards the expert plateau. This suggests that the learning of task-specific gaze behaviors can be adequately quantified using the presented method. We were further able to verify previous findings, which demonstrated that gaze behaviors become more similar after a period of learning (van der Gijp et al., 2017), and, using the k -mer analysis, specified some of those gaze patterns that were evoked with increased on-task experience.

Finally, the statistical analysis of the pattern slopes suggests that a simpler stimulus reduces the total amount of trials required to reach expert skill levels, while the actual learning rate between tasks was not influenced. This finding should be further investigated in the future using more complex real-world tasks.

We believe that the presented approach adds significant value to the understanding of how expertise related to gaze behavior is developed by contributing advancements to traditional AOI evaluation and string-edit approaches that compare pairwise (Ben Khedher et al., 2017) or group-wise (Burch et al., 2018) similarities of gaze sequences. While the exact AOI composition of the analyzed sequence patterns, i.e. *monitor assembly* or *consult manual*, and their appearance frequencies are expected to differ for each domain of expertise, the k -mer approach applies to eye tracking data of all domains.

Limitations

Limitations of the current work include the limited number of expert subjects within the evaluated ET study. We are aware that the expert data used in this study might not be representative, but the results strongly indicate the consistency of experts' sequential gaze behavior. Furthermore, this approach suffers from the same limitation as other sequence analysis approaches, which is that the correct or incorrect outcome of a task cannot be directly inferred solely using quantified gaze data and needs to be investigated separately. Some additional

uncertainties remain in regard to the optimal choice of k and the number of AOIs to be used for the analysis, and should be further investigated. In the future, an AOI frequency-based approach, as introduced by Arzarello et al. (2011), could be implemented to help filter out patterns that are considered identical and thus help automate the choice of analyzed sequence patterns. Furthermore, for mobile ET, conducting semantic gaze mapping manually results in increased manual labor (Vansteenkiste et al., 2015). However, recent advancements in deep convolutional neural network applications have shown that a massive reduction in the effort spent on semantic gaze mapping can be achieved, enabling the automated detection and mapping of looked-at objects (Wolf, Hess, Bachmann, Lohmeyer, & Meboldt, 2018).

Conclusion

In the present work, we introduced a novel algorithmic methodological approach for the quantification of expertise development using sequential k -mer gaze patterns. Through the evaluation of a simplified natural handling task using a unique multi-trial study design, we contribute to the understanding of the acquisition of task-specific perceptual expertise. By investigating how specific behaviors develop within the same individuals over time, evidence was given that k -mer patterns can be a suitable metric to measure and assess learning progress. Additionally, by retaining the object-related AOI identity, gaze sequence patterns are easily interpretable, while containing both temporal and spatial task information. Specifically, for novice education and skill assessment, this approach could provide an answer to the need for a measurement methodology for operator experience and allows us to advance the understanding of task-specific gaze behavior. Consequently, an assessment of the rate with which these patterns approach expert thresholds could serve as a quantitative means to verify the achievement of specific competencies.

By presenting the measurable development of gaze-based expertise using a multi-trial study design coupled with the algorithmic k -mer approach, researchers are provided with a promising new methodological framework to study domain-specific expertise and the effects of training on the development of expert gaze strategies.

Notes The data for all experiments are available at <https://doi.org/10.6084/m9.figshare.12593657.v1>. None of the experiments were preregistered.

Funding Open Access funding provided by ETH Zurich.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as

you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Anderson, N. C., Anderson, F., Kingstone, A., & Bischof, W. F. (2015a). A comparison of scanpath comparison methods. *Behavior Research Methods*, 47(4), 1377–1392. <https://doi.org/10.3758/s13428-014-0550-3>
- Anderson, N. C., Anderson, F., Kingstone, A., Bischof, W. F., Mathôt, S., Cristino, F., ... Weijer, J. Van De. (2015b). A simple way to estimate similarity between pairs of eye movement sequences. *Journal of Eye Movement Research*, 5(4), 1–15. <https://doi.org/10.16910/jemr.5.1.4>
- Arzarello, F., Ferrara, F., Holmqvist, K., Lindström, P., Robutti, O., Andrà, C., & Sabena, C. (2011). A method for quantifying focused versus overview behavior in AOI sequences. *Behavior Research Methods*. <https://doi.org/10.3758/s13428-011-0104-x>
- Beck, M. R., Martin, B. A., Smitherman, E., & Gaschen, L. (2013). Eyes-On Training and Radiological Expertise. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 55(4), 747–763. <https://doi.org/10.1177/0018720812469224>
- Ben Khedher, A., Jraidi, I., & Frasson, C. (2017). Assessing learners' reasoning using eye tracking and a sequence alignment method. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 10362 LNCS, 47–57. https://doi.org/10.1007/978-3-319-63312-1_5
- Brunyé, T. T., Carney, P. A., Allison, K. H., Shapiro, L. G., Weaver, D. L., & Elmore, J. G. (2014). Eye Movements as an Index of Pathologist Visual Expertise: A Pilot Study. *PLoS ONE*, 9(8), e103447. <https://doi.org/10.1371/journal.pone.0103447>
- Bulling, A., Weichel, C., & Gellersen, H. (2013). EyeContext: Recognition of high-level contextual cues from human visual behaviour. *Conference on Human Factors in Computing Systems - Proceedings*, 305–308. <https://doi.org/10.1145/2470654.2470697>
- Burch, M., Kurzhals, K., Kleinhans, N., & Weiskopf, D. (2018). EyeMSA: Exploring eye movement data with pairwise and multiple sequence alignment. *Eye Tracking Research and Applications Symposium (ETRA)*, 1–5. <https://doi.org/10.1145/3204493.3204565>
- Castner, N., Kasneci, E., Kübler, T., Scheiter, K., Richter, J., Eder, T., ... Keutel, C. (2018). Scanpath comparison in medical image reading skills of dental students. *Eye Tracking Research and Applications Symposium (ETRA)*, 1–9. <https://doi.org/10.1145/3204493.3204550>
- Castner, N., Kuebler, T. C., Scheiter, K., Richter, J., Eder, T., Huettig, F., ... Kasneci, E. (2020). Deep semantic gaze embedding and scanpath comparison for expertise classification during OPT viewing. *Symposium on Eye Tracking Research and Applications*, 1–10. <https://doi.org/10.1145/3379155.3391320>
- Chi, M. T. H. (2006). Two Approaches to the Study of Experts' Characteristics. In: K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge Handbook of Expertise and Expert Performance* (pp. 21–30). <https://doi.org/10.1017/CBO9780511816796.002>

- Cristino, F., Mathôt, S., Theeuwes, J., & Gilchrist, I. D. (2010). ScanMatch: A novel method for comparing fixation sequences. *Behavior Research Methods*, 42(3), 692–700. <https://doi.org/10.3758/BRM.42.3.692>
- Crowe, E. M., Gilchrist, I. D., & Kent, C. (2018). New approaches to the analysis of eye movement behaviour across expertise while viewing brain MRIs. *Cognitive Research: Principles and Implications*, 3(1), 1–14. <https://doi.org/10.1186/s41235-018-0097-4>
- Day, E. L., Tang, H., Kendhammer, L. K., & Pienta, N. J. (2018). *Sequence Analysis: Use of Scanpath Patterns for Analysis of Students' Problem-Solving Strategies*. <https://doi.org/10.1021/bk-2018-1292.ch005>
- Dewhurst, R., Nyström, M., Jarodzka, H., Foulsham, T., Johansson, R., & Holmqvist, K. (2012). It depends on how you look at it: Scanpath comparison in multiple dimensions with MultiMatch, a vector-based approach. *Behavior Research Methods*, 44(4), 1079–1100. <https://doi.org/10.3758/s13428-012-0212-2>
- Duchowski, A. T. (2017). *Eye Tracking Methodology*. <https://doi.org/10.1007/978-3-319-57883-5>
- Eivazi, S., Bednarik, R., Tukiainen, M., von und zu Fraunberg, M., Leinonen, V., & Jääskeläinen, J. E. (2012). Gaze behaviour of expert and novice microneurosurgeons differs during observations of tumor removal recordings. *Proceedings of the Symposium on Eye Tracking Research and Applications - ETRA '12*, 377. <https://doi.org/10.1145/2168556.2168641>
- Elbattah, M., Guerin, J. L., Carette, R., Cilia, F., & Dequen, G. (2020). NLP-Based Approach to Detect Autism Spectrum Disorder in Saccadic Eye Movement. *2020 IEEE Symposium Series on Computational Intelligence, SSCI 2020*, 1581–1587. <https://doi.org/10.1109/SSCI47803.2020.9308238>
- Fahimi, R., & Bruce, N. D. B. (2020). On metrics for measuring scanpath similarity. *Behavior Research Methods*, 1–20. <https://doi.org/10.3758/s13428-020-01441-0>
- Foulsham, T., Dewhurst, R., Nyström, M., Jarodzka, H., Johansson, R., Underwood, G., & Holmqvist, K. (2012). Comparing scanpaths during scene encoding and recognition: A multi-dimensional approach. *Journal of Eye Movement Research*, 5(4), 3. <https://doi.org/10.16910/jemr.5.4.3>
- Fox, S. E. & Faulkne-Jones B. E. (2017). Eye-Tracking in the Study of Visual Expertise: Methodology and Approaches in Medicine. *Frontline Learning Research*, 5(3), 29–40.
- Gegenfurtner, A. (2013). Transitions of Expertise. In: *Transitions in Vocational Education* (pp. 309–324). <https://doi.org/10.2307/j.ctvd0j2f.19>
- Gegenfurtner, A., Lehtinen, E., & Säljö, R. (2011). Expertise Differences in the Comprehension of Visualizations: A Meta-Analysis of Eye-Tracking Research in Professional Domains. *Educational Psychology Review*, 23, 523–552. <https://doi.org/10.1007/s10648-011-9174-7>
- Gignac, G. E., & Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. *Personality and Individual Differences*, 102, 74–78. <https://doi.org/10.1016/j.paid.2016.06.069>
- Haslbeck, A., & Zhang, B. (2017). I spy with my little eye: Analysis of airline pilots' gaze patterns in a manual instrument flight scenario. *Applied Ergonomics*, 63, 62–71. <https://doi.org/10.1016/j.apergo.2017.03.015>
- Hou, H.-T., Chang, K.-E., & Sung, Y.-T. (2009). Using blogs as a professional development tool for teachers: analysis of interaction behavioral patterns. *Interactive Learning Environments*, 17(4), 325–340. <https://doi.org/10.1080/10494820903195215>
- Jaarsma, T., Jarodzka, H., Nap, M., van Merriënboer, J. J. G., & Boshuizen, H. P. A. (2015). Expertise in clinical pathology: combining the visual and cognitive perspective. *Advances in Health Sciences Education*, 20(4), 1089–1106. <https://doi.org/10.1007/s10459-015-9589-x>
- Jarodzka, H., & Boshuizen, H. P. A. (2017). Unboxing the black box of visual expertise in medicine. *Frontline Learning Research*, 5(3), 167–183. <https://doi.org/10.14786/flr.v5i3.322>
- Jarodzka, H., Holmqvist, K., & Nyström, M. (2010). A vector-based, multidimensional scanpath similarity measure. *Eye Tracking Research and Applications Symposium (ETRA)*, 211–218. <https://doi.org/10.1145/1743666.1743718>
- Kanan, C., Bseiso, D. N. F., Ray, N. A., Hsiao, J. H., & Cottrell, G. W. (2015). Humans have idiosyncratic and task-specific scanpaths for judging faces. *Vision Research*, 108, 67–76. <https://doi.org/10.1016/j.visres.2015.01.013>
- Kelly, B. S., Rainford, L. A., Darcy, S. P., Kavanagh, E. C., & Toomey, R. J. (2016). The Development of Expertise in Radiology: In Chest Radiograph Interpretation, “Expert” Search Pattern May Predate “Expert” Levels of Diagnostic Accuracy for Pneumothorax Identification. *Radiology*, 280(1), 252–260. <https://doi.org/10.1148/radiol.2016150409>
- Khan, N., Abboudi, H., Khan, M. S., Dasgupta, P., & Ahmed, K. (2014). Measuring the surgical ‘learning curve’: methods, variables and competency. *BJU International*, 113(3), 504–508. <https://doi.org/10.1111/bju.12197>
- Kübler, T. C., Rothe, C., Schiefer, U., Rosenstiel, W., & Kasneci, E. (2017). SubsMatch 2.0: Scanpath comparison and classification based on subsequence frequencies. *Behavior Research Methods*, 49(3), 1048–1064. <https://doi.org/10.3758/s13428-016-0765-6>
- Land, M. F., & Hayhoe, M. (2001). In what ways do eye movements contribute to everyday activities? *Vision Research*, 41(25–26), 3559–3565. [https://doi.org/10.1016/S0042-6989\(01\)00102-X](https://doi.org/10.1016/S0042-6989(01)00102-X)
- Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Cybernetics and Control Theory*. Retrieved from <https://nymity.ch/sybilhunting/pdf/Levenshtein1966a.pdf>. Accessed 21 May 2020
- Liu, B., Shi, Y., Yuan, J., Hu, X., Zhang, H., Li, N., ... Fan, W. (2013). Estimation of genomic characteristics by analyzing k-mer frequency in de novo genome projects. In: *arxiv.org*. Retrieved from <https://arxiv.org/abs/1308.2012>. Accessed 21 May 2020
- Manekar, S. C., & Sathe, S. R. (2018). *A benchmark study of k-mer counting methods for high-throughput sequencing*. 7, 1–13. <https://doi.org/10.1093/gigascience/giy125>
- McIntyre, N. A., & Foulsham, T. (2018a). Scanpath analysis of expertise and culture in teacher gaze in real-world classrooms. *Instructional Science*, 46(3), 435–455. <https://doi.org/10.1007/s11251-017-9445-x>
- McIntyre, N. A., & Foulsham, T. (2018b). Scanpath analysis of expertise and culture in teacher gaze in real-world classrooms. *Instructional Science*, 46(3), 435–455. <https://doi.org/10.1007/s11251-017-9445-x>
- Noton, D., & Stark, L. (1971). Scanpaths in eye movements during pattern perception. *Science*, 171(3968), 308–311. <https://doi.org/10.1126/science.171.3968.308>
- Ooms, K., De Maeyer, P., Fack, V., Van Assche, E., & Witlox, F. (2012). Interpreting maps through the eyes of expert and novice users. *International Journal of Geographical Information Science*, 26(10), 1773–1788. <https://doi.org/10.1080/13658816.2011.642801>
- Privitera, C. M., & Stark, L. W. (2000). Algorithms for defining visual regions-of-interest: comparison with eye fixations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(9), 970–982. <https://doi.org/10.1109/34.877520>
- Ren, J., Ahlgren, N. A., Lu, Y. Y., Fuhrman, J. A., & Sun, F. (2017). VirFinder: a novel k-mer based tool for identifying viral sequences from assembled metagenomic data. *Microbiome*, 5(1), 69. <https://doi.org/10.1186/s40168-017-0283-5>
- Robbins, T. R. (2019). A Learning Curve-Based Assembly Game. *Decision Sciences Journal of Innovative Education*, 17(4), 344–362. <https://doi.org/10.1111/dsji.12190>

- Sheridan, H., & Reingold, E. M. (2017). The holistic processing account of visual expertise in medical image perception: A review. *Frontiers in Psychology*, 8, 1620. <https://doi.org/10.3389/fpsyg.2017.01620>
- Tien, T., Pucher, P. H., Sodergren, M. H., Sriskandarajah, K., Yang, G. Z., & Darzi, A. (2015). Differences in gaze behaviour of expert and junior surgeons performing open inguinal hernia repair. *Surgical Endoscopy*, 29(2), 405–413. <https://doi.org/10.1007/s00464-014-3683-7>
- Tsai, M.-J., Hou, H.-T., Lai, M.-L., Liu, W.-Y., & Yang, F.-Y. (2012). Visual attention for solving multiple-choice science problem: An eye-tracking analysis. *Computers & Education*, 58(1), 375–385. <https://doi.org/10.1016/j.compedu.2011.07.012>
- van der Gijp, A., Ravesloot, C. J., Jarodzka, H., van der Schaaf, M. F., van der Schaaf, I. C., van Schaik, J. P. J., & ten Cate, T. J. (2017). How visual search relates to visual diagnostic performance: a narrative systematic review of eye-tracking research in radiology. *Advances in Health Sciences Education*, 22(3), 765–787. <https://doi.org/10.1007/s10459-016-9698-1>
- Vansteenkiste, P., Cardon, G., Philippaerts, R., & Lenoir, M. (2015). Measuring dwell time percentage from head-mounted eye-tracking data – comparison of a frame-by-frame and a fixation-by-fixation analysis. *Ergonomics*, 58(5), 712–721. <https://doi.org/10.1080/00140139.2014.990524>
- Watalingam, R. D., Richetelli, N., Pelz, J. B., & Speir, J. A. (2017). Eye tracking to evaluate evidence recognition in crime scene investigations. *Forensic Science International*, 280, 64–80. <https://doi.org/10.1016/j.forsciint.2017.08.012>
- Wilson, M. R., Vine, S. J., Bright, E., Masters, R. S. W., Defriend, D., & McGrath, J. S. (2011). Gaze training enhances laparoscopic technical skill acquisition and multi-tasking performance: A randomized, controlled study. *Surgical Endoscopy*, 25(12), 3731–3739. <https://doi.org/10.1007/s00464-011-1802-2>
- Wolf, J., Hess, S., Bachmann, D., Lohmeyer, Q., & Meboldt, M. (2018). Automating areas of interest analysis in mobile eye tracking experiments based on machine learning. *Journal of Eye Movement Research*. <https://doi.org/10.3929/ethz-b-000309840>
- Wood, G., Knapp, K. M., Rock, B., Cousens, C., Roobottom, C., & Wilson, M. R. (2013). Visual expertise in detecting and diagnosing skeletal fractures. *Skeletal Radiology*, 42(2), 165–172. <https://doi.org/10.1007/s00256-012-1503-5>
- Zheng, B., Tien, G., Atkins, S. M., Swindells, C., Tanin, H., Meneghetti, A., ... Panton, M. (2011). Surgeon's vigilance in the operating room. *American Journal of Surgery*, 201(5), 673–677. <https://doi.org/10.1016/j.amjsurg.2011.01.016>
- Zimmermann, J. M., Vicentini, L., Lohmeyer, Q., Taramasso, M., Maisano, F., & Meboldt, M. (2020). Visual Behaviour Strategies of Operators during Catheter-Based Cardiovascular Interventions. *Journal of Medical Systems*, 44(1), 1–7. <https://doi.org/10.1007/s10916-019-1480-5>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.