Displaying Teacher's Gaze in a MOOC: Effects on Students' Video Navigation Patterns

Kshitij Sharma[1], Patrick Jermann[2], and Pierre Dillenbourg[1]

1. Computer Human Interaction in Learning and Instruction, École Polytechnique Fédérale de Lausanne, Switzrland
2. Center for Digital Education, École Polytechnique Fédérale de Lausanne, Switzrland
kshitij.sharma@epfl.ch, patrick.jermann@epfl.ch, pierre.dillenbourg@epfl.ch

Abstract. We present an eye-tracking study where we augment a Massive Open Online Course (MOOC) video with the gaze information of the teacher. We tracked the gaze of a teacher while he was recording the content for a MOOC lecture. Our working hypothesis is that displaying the gaze of the teacher will act as cues in crucial moments of dyadic conversation, the teacher-student dyad, such as reference disambiguation. We collected data about students' video interaction behaviour within a MOOC. The results show that the showing the teacher's gaze made the content easier to follow for the students even when complex visual stimulus present in the video lecture.

1 Introduction

In the present decade, off the shelf eye-trackers have became readily available for use. With the advent in the mobile eye-tracking technology, eye-tracking researchers are no longer bound to laboratory based experiments. Eye-tracking provides direct access to users' attention which is useful in situations like MOOCs, where the major questions involve student engagement and learning processes.

We know from previous eye-tracking research that speakers look at the objects they refer to just before pointing and verbally naming the objects [6]. Listeners on the other hand, look at the referred objects shortly after seeing the speaker point and refer to the objects [2]. Richardson and colleagues [9] showed that the listeners who were better at attending the references made by the speaker were also better at understanding the context of the conversation. One way to aid the listeners attending the reference in a better way is to display where the speaker is looking at. This helps the listeners in a better disambiguation of the complex references [18, 17]. In the case of complex stimulus displaying the gaze of speaker makes the disambiguation of the references even easier [19]. This motivated us to study the effect of showing the gaze of the teacher in a MOOC video on the navigation patterns of the students.

Students' navigation styles can tell us a lot about their perception about the content. Li and colleagues [16] conducted a study with over 30,000 students and 100 videos across two courses where the authors asked students to rate the

perceived difficulty of the content after the students watched the video. Based on students' rating and their video navigation behaviour [16] concluded that the students who perceived the video content as easy to understand did less frequent and shorter pauses, and replayed the video less frequently.

In this contribution, we show that displaying teacher's gaze in a MOOC video-lecture can help the students understand more easily the content of a MOOC video. Moreover, this effect remains consistent with the increasing complexity of the situation explained by the teacher.

2   Related work

The literature on the use of eye-tracking methods for MOOCs is scarce. However, there is a lot to learn from the eye-tracking research carried out in various fields like online education, usability of the video content and gaze contingency (displaying the gaze of expert or one of the collaborators) studies. In this section we give a brief overview of the eye-tracking studies done in these different but related fields.

2.1   Eye-tracking and MOOCs

There are a few studies carried out in the MOOC context that have used eye-tracking as the process data source. Sharma and colleagues [10, 11] proposed gaze measures to predict the learning outcome in MOOCs. [10] uses the low level gaze features (derived from the gaze directly on the stimulus) to predict the learning outcome; while [11] used the fact that how closely the students follow the teachers' deictic and verbal references to predict the learning outcomes [7].

2.2   Eye-tracking and online education

Use of eye-tracking in online education has provided the researchers with insights about the students' learning processes and outcomes. Van Gog [14] used eye-tracking data to provide feedback to the students about their action while troubleshooting an electrical circuit and found that the feedback improved the learning outcomes. In another contribution, Van Gog [13] found that displaying an experts gaze during problem solving guides the novices to invest more mental effort than when no gaze is displayed. Amadieu and colleagues [3] used eye-tracking data to find the affect of expertise, in a collaborative concept map task, on the cognitive load. The authors concluded that the average fixation duration was lower for the experts indicating more cognitive load on experts than novices. In an experiment, where the participants had to learn a game, [1] found that the good learners focus more on the contraption ares of the game while they think about the possible solutions. [12] found that the students spend more time on the complementary pictures in a presentation, than decorative pictures.

2.3   Gaze contingency and reference disambiguation

Gaze contingent experiments are at the proactive side of the eye-tracking technology. These experiments consist in displaying the gaze of collaborating partners to each other; or displaying the gaze of an expert to a novice in order to teach the novice [5] . Another modality of gaze contingency is using gaze as a mode of communication. In a collaborative "Qs-in-Os" search [4] showed that the sharing gaze information between collaborating partners results in a strategy of division of labour as effective as if the partners were talking face to face. Using gaze as a communication modality [8] used gaze information to inform participants about the effectiveness of grounding process between a human and an infotainment presentation agent. In a multiparty video conference system [15] used gaze information to rotate the participants' virtual 3D representations to the persons they were talking to. Displaying the gaze of speaker helps the listener in deciphering the references [18, 17]. Moreover, gaze of speaker makes it easier for the listener in deciphering the references in situations with high ambiguity [19].

3   The present work

In this section, we present the important details about the study we carried out to explore the effects of displaying gaze of the teacher on the students' video interaction patterns. The teacher's gaze was recorded when he was recording the MOOC video. Our prime hypothesis is that displaying teachers' gaze on the video will make the reference disambiguation easy in high ambiguous situations. Moreover, displaying teacher's gaze on the video will also make the students' behaviour more linear in terms of following the content.

3.1   Experiment Setup

We asked one of the teachers to track his eyes on the MOOC video he was going to record. We used SMI mobile eye-tracking glasses to record the gaze of the teacher. The main motivation to use mobile eye-trackers was to give the teacher as ecologically valid environment as possible. The setup of the MOOC recording studio is shown in figure 1. The teacher was equipped with the eye-tracking glasses. There was a screen capture software running on the tablet with the actual content to record every move of the teacher. Also, there was a camera on the ceiling of the studio to capture the external actions (external to tablet) on the tablet. We put 9 fiducial markers on the tablet so that later we can be able to re-locate the gaze pointer of the teacher on the tablet. The video was uploaded on Coursera as one of the video lectures during one of the weeks of the course "Villes africaines: Introduction à la planification urbaine" (African cities : an introduction to urban planning). The teacher explicitly chose the parts of the video where he wanted to display his gaze. None of the authors had control over this fact.
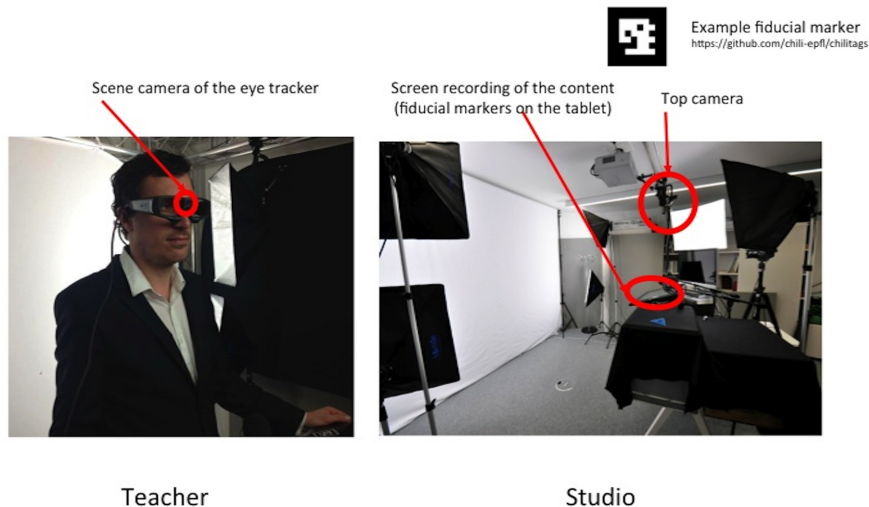
Fig. 1. Setup: The teacher is equipped with the SMI mobile eye-tracking glasses (left) and the MOOC recording studio (right) with the top camera on the ceiling and the tablet used by the teacher. The fiducial markers (top-right) are glued to the tablet to make the re-localisation of teacher's gaze on the actual content easy.

3.2   Research Questions

Through this experiment, we wanted to explore following two research questions:

1. What is the affect of displaying teachers' gaze on a MOOC lecture on students' video interaction patterns? Our hypothesis is that displaying teacher's gaze on the video will make the students' behaviour more linear in terms of following the content (*behavioural hypothesis*)
2. If there is a relation between the students' video interaction patterns and teacher's gaze, how is it effected by the ambiguity of the video? We hypothesise that displaying teachers' gaze on the video will make the reference disambiguation easy in ambiguous situations (*eye-tracking hypothesis*).

3.3   Re-localisation of teacher's gaze

We recorded three different video streams from the setup of figure 1. First, the video from scene camera of the eye-tracker. Second, from the top view camera in the studio. Finally, the third video stream comes from the screen capture software running on the teacher's tablet. We knew teacher's gaze positions in the frame of the video captured from the scene camera of the eye-tracker. The objective was to find the gaze positions on the video from the screen capture of the tablet. This is not a trivial task. Since the teacher was given full freedom to move, his field of the view of changes every instant. We compute the gaze positions on the actual content using following steps:

Step 1 We compute the relative position of the fiducial markers and the gaze positions in the video from the scene camera of the eye-tracker.

Step 2 We compute the relation between the positions of the fiducial markers in the video from the top camera and the video from the scene camera of the eye-tracker.

Step 3 Using the two relations, computed in steps 1 and 2,we compute the gaze positions on the video from the top camera. The output of this step is a video where the gaze pointers are shown on the video from the top camera.

Step 4 The video from the top camera is geometrically a distorted version of the video from the screen capture software running on the tablet. Hence, we remove the distortion from the resulting video of step 3 to get the video from the screen capture software with teachers' gaze pointers.

Fig. 2. Example of a high ambiguity image from the experimental video. The image is captured via satellite and the teacher is explaining the landscape captured. We rate these type of images because high ambiguity images as disambiguating a reference like "'the cathedral" is difficult without a visual cue.

3.4   Ambiguity in stimulus and teacher's gaze

To analyse the students' behaviour we divided the video into four episodes based on whether there was teacher's gaze present on the video and what was the level of ambiguity in the images shown in the video (high vs low ambiguity). The ambiguity in the image was determined by *how easy was it to disambiguate a simple verbal reference on any part of the image*. Simply put, how easy it was to locate what part of image/scene the speaker is talking about. Images with high ambiguity are satellite images and aerial images where the target reference are smaller in size and are not obviously present in front of the listeners' eyes.

Whereas, images with low ambiguity are street views where the target references a bigger in size and are easily detectable by the listeners. Examples of images with high and low ambiguity are shown in figures 2 and 3 respectively. Examples of simple verbal references in both the images would be "'the cathedral" (figure 2) and "the tree" (figure 3). This categorisation was done by the authors and later confirmed by the teacher himself. The main reason for this categorisation was to be able to segment the video in high and low ambiguity stimulus periods.

Fig. 3. Example of a low ambiguity image from the experimental video . The image is typical street view and the teacher is explaining the landscape captured. We rate these type of images as low ambiguity images because disambiguating a reference like "'the tree" is easy without a visual cue.

3.5   Measures

In this subsection, we present the measures of students' behaviour we used to analyse the affect of displaying the teacher's gaze in the video. We compare the measures in two ways. First, we compare the values of the variable for the experimental video and other videos (*between videos variable*). Second, we compare the values of the variable within the experimental video for different episodes in the video (*within video variable*).

Proportion of replayed video length: This is calculated by counting the number of video seconds that were played more than once. This primarily indicates the difficulty that student perceives during the video lecture. A high proportion of replayed video for a student suggest that the student was not able to understand some of the content properly in the first time going through the video. This is used only as a between video variable.

Number of pauses: This is the average number of pauses that a student makes during one video. High number of pauses indicates the difficulty as well as frequent disengagements from the video. This variable is used as both a between and within video variable.

Ratio of pause time and video length: This is the ratio of the total time spent by the students while keeping the video in a pause state and the total video length. Longer pauses will result in a higher value of the ratio. Moreover, the higher ratio will indicate the difficulty in understanding the video as students will need more time to grasp the concept. This variable is used only as a between video variable.

Number of seek backs: This is the average number of backward jumps that a student makes during one video. The seek back event typically reflects two necessities from a student. First, a check for a reference that was made at a previous video point. Second, a complete section of the video being too difficult to understand and the student decided to re-watch the whole video segment. This variable is used as both the between and within video variable.

4   Results

As we mentioned in the section 3.5, there are two levels of analysis to be presented in the paper. First, we compare students' behaviour across different videos in the weeks succeeding and the preceding the week of the experimental video. Second, we compare the students' behaviour across different episodes within the experimental video. The three weeks are weeks 10, 11 and 12, which also are the last weeks of the course. The main reasons behind selecting only three weeks to compare are that the size of student population is comparable for these three weeks and that the population is comparable in terms of motivation to finish the course and the levels of engagement.

4.1   Comparing user behaviour across different weeks

In this subsection we compare the number of pauses, seek backs, seek forwards, the pause time and replay time across different videos. The experimental video has "11.1" as the label. Moreover, in the figures 4 - 7 the variables corresponding to the experimental video are shown as a thicker bar than the other videos.

Proportion of replayed video length: We observe that the proportion of the replayed length video is lowest (figure 4) for the experimental video ($F[9,4202] =2.12$, $p = .03$).

Number of pauses: We observe that the average number of pauses is second lowest (figure 5) for the experimental video ($F[9,4202] =2.89$, $p = .002$). The reason for the video 12.3 having the least number of pauses is that the video 12.3 is the "end of course" video with only a few seconds of length.
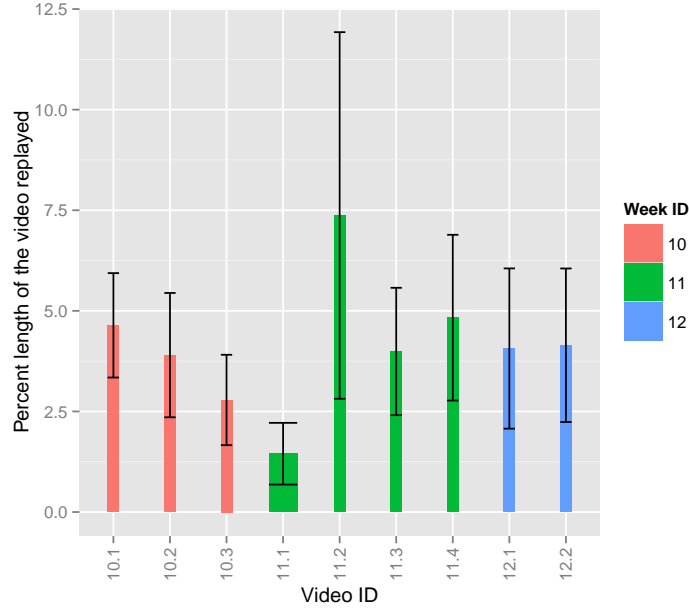
Fig. 4. Proportion of replayed video length compared across weeks 10, 11 and 12. The experimental video (thick bar, id 11.1) has significantly the lowest proportion of the replayed length among all the videos.

Ratio of pause time and video length: We observe that the ratio of pause time and video length is lowest (figure 6) for the experimental video ($F[9,4202] = 2.58$, $p = .005$).

Number of seek backs: We observe that the average number of seek backs is second lowest (figure 7) for the experimental video ($F[9,4202] = 1.92$, $p = .04$). Again, 'end of course" video (12.3) is with only a few seconds of length.

4.2   Comparing user behaviour within the video

In this subsection, we compare the number of pauses, and seek back actions for different episodes within the experimental video (figures 8 and 9). As we explained in section 3.4, the experimental video was divided in 4 different kinds of episodes based on two facts: 1) whether teacher's gaze is present or not; and 2) whether the ambiguity in the video has high or low ambiguity. In table 1 we observe the following:

Number of pauses in "gaze-present" episodes is lower than that in "gaze-absent" episodes. Moreover, there are lower number of pauses in the high ambiguity situations than those in low ambiguity situations ($\chi^2 = 79.83$, p ¡ .001).
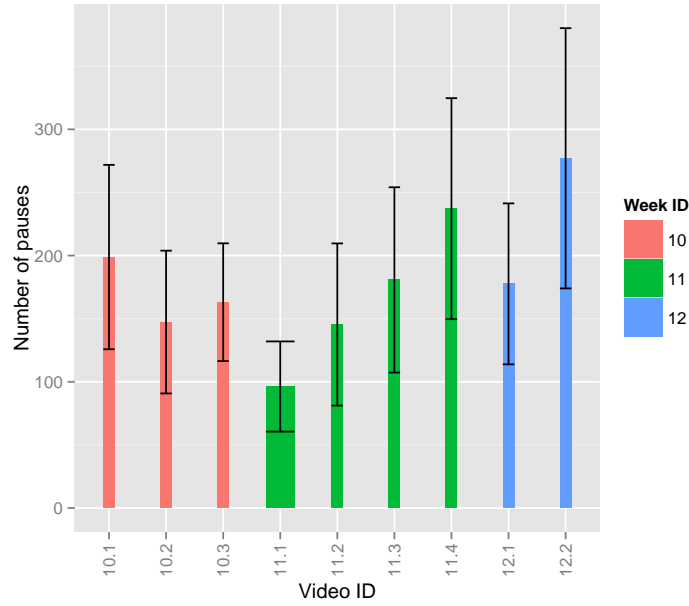
Fig. 5. Average number of pauses compared across weeks 10, 11 and 12. The experimental video (thick bar, id 11.1) has significantly lower number of pauses than other videos.

Number of seek backs in "gaze-present" episodes is lower than that in "gaze-absent" episodes. Moreover, there are lower number of seek backs in the high ambiguity situations than those in low ambiguity situations ($\chi^2 = 164.83$, p = .001).

Table 1. Number of different types of events compared within the experiment video across different episodes.

Action	Pause		Seek-back	
Ambiguity level of the scene	High	Low	High	Low
Gaze-present	16 (7.22)	64 (-4.27)	18 (-4.27)	23 (-5.71)
Gaze-absent	94 (-2.97)	232 (-2.97)	52 (1.21)	142 (8.77)

5 Discussion

The results in section 4.1 show that the behavioural hypothesis (section 3.2) comes out to be true. The fact that the students have less number of seek back
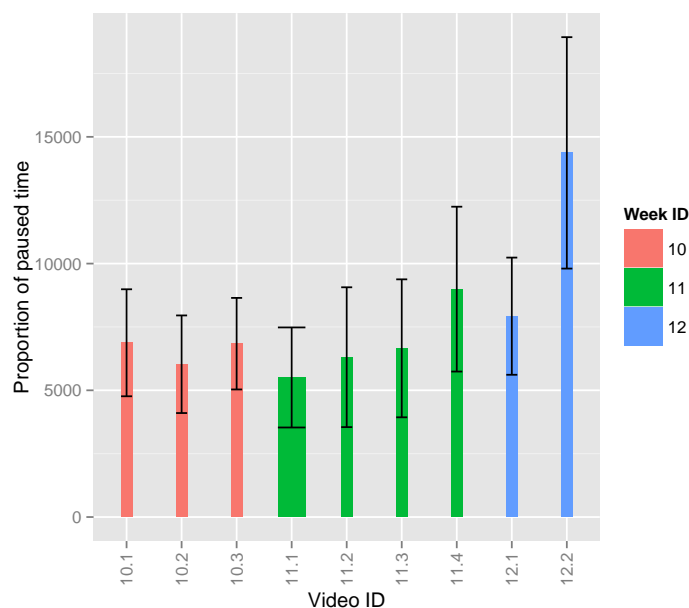
Fig. 6. Ratio of pause time and video length compared across weeks 10, 11 and 12. The experimental video (thick bar, id 11.1) has significantly the lowest ratio among all the videos.

events reflects the fact that they did not need to check back the previously told content because it was easy to understand for them once the teacher's gaze was displayed on the video. Moreover, the same fact is strongly supported with less amount of video content replayed for the experimental video. Similarly, less frequent and shorter pauses indicate that the content delivery was also easy due to the presence of additional cues to disambiguate complex references during the video. Li and colleagues [16] found similar video navigation patterns in their study for the students who perceived the video content as easy to understand.

The observation that there are less seek back and pauses during the experimental video also verifies our working hypothesis of making the learning experience more linear as compared to the video material. With less breaks in the content delivery and the less back references the student is well aligned with the video content in temporal space and hence the creation and maintenance of a mutual understanding between the teacher student dyad is effective and efficient.

The key difference between the experimental video and the videos from the other week was the augmentation of teacher's gaze on top of the video content. Since the students see where the teacher was looking and it is proved by eye-tracking research that people start looking at the point they are about to refer
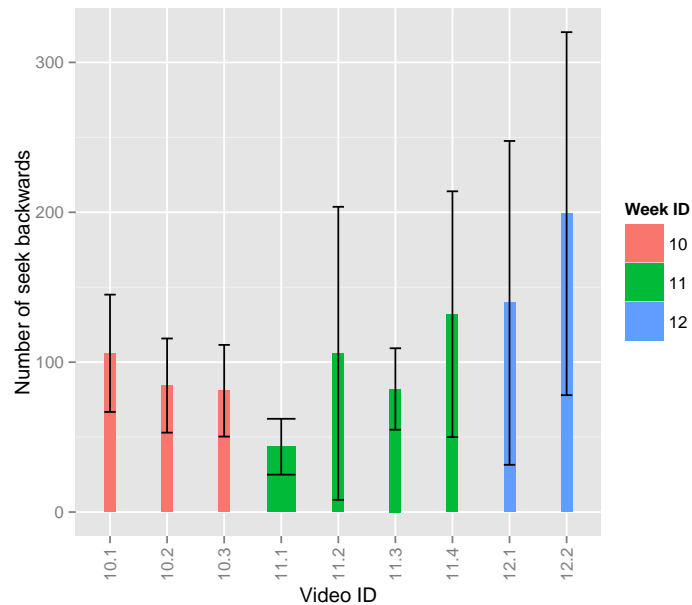
Fig. 7. Average number of seek back events compared across weeks 10, 11 and 12. The experimental video (thick bar, id 11.1) has significantly lower number of pauses than other videos.

and hence it is easy to disambiguate the point of reference for the listener when (s)he sees the gaze of referee.

The results from the section 4.2 proved our eye-tracking hypothesis (section 3.2) to be true as well. The students make less pause and less seek backs in high ambiguity situations, such as the teacher describing complex images like satellite captured image (figure 2),when the gaze is present in the video as compared to when the gaze is absent on the video. This effect is present, although less pronounced, in situations with low ambiguity (for example when the teacher is explaining a street view, figure 3). Prasov and Chai [19] also found in their study about reference disambiguation in complex stimulus that the displaying the gaze of speaker makes it easy for the listener to disambiguate the reference.

Although the results support our hypothesis, more experimentation is required to find out whether displaying teacher's gaze helps in increasing the effectiveness of learning experiences. Moreover, further investigation is necessary to comment on the affect of augmenting multiple MOOC videos with teacher's gaze on the overall learning experience of students.
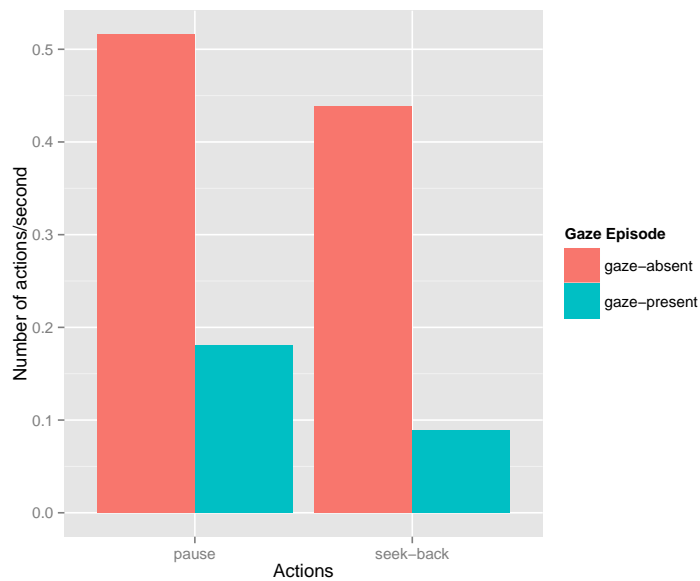
Fig. 8. Proportions of different types of events compared within the experiment video across different gaze episodes.

6   Conclusion

We presented a MOOC experiment, where we augmented the video lecture with the gaze of the teacher. We showed that displaying the gaze of the teacher on the video not only makes the video watching more linear but also makes it easy for the students to disambiguate teacher's references in complex situations. We put emphasis on the fact that it is important to help students understand where the teacher is referring in the video because these are the moments that create a shared understanding for the teacher student dyad. Having more such moments help maintain the shared understanding.

The introduction of teachers' gaze might also work as a novelty in the engagement process of the students as well. To keep the engagement up to a level which benefits the student, such novelties could prove to be effective. The results show that usually during the end of the course the students who watch the videos decrease drastically. However, once we put the experimental video online the number of students who watched the video increased from the previous week.

In a nutshell, both of our hypotheses were verified and this motivates us to continue experiments with augmenting the MOOC videos with the visual cues to help students better understand the content. Our future work includes experimentation with different eye-tracking data visualisations to augment the MOOC video and check how it affects the students' video navigation patterns and their learning processes. Also to perform a laboratory experiment to see
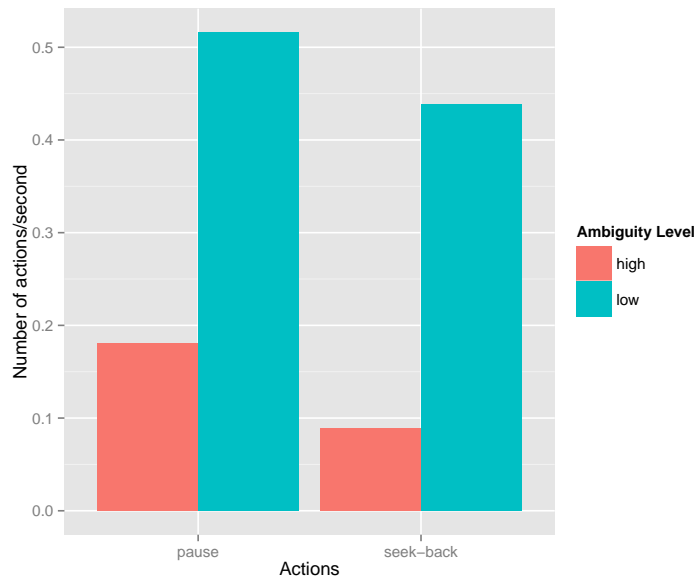
Fig. 9. Proportions of different types of events compared within the experiment video across different ambiguity episodes.

how closely the students follow the gaze pointer of the teacher and how it affects their learning outcome.

7 Acknowledgements

References

1. Alkan, S. and Cagiltay, K., Studying computer game learning experience through eye tracking. British Journal of Educational Technology, 38(3):538?542, 2007.
2. P. D. Allopenna, J. S. Magnuson, and M. K. Tanenhaus. Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. Journal of memory and language, 38(4):419?439, 1998.
3. F. Amadieu, T. Van Gog, F. Paas, A. Tricot, and C. Marin?e. Effects of prior knowledge and concept-map structure on disorientation, cognitive load, and learning. Learning and Instruction, 19(5):376?386, 2009.
4. S. E. Brennan, X. Chen, C. A. Dickinson, M. B. Neider, and G. J. Zelinsky. Coordinating cognition: The costs and benefits of shared gaze during collaborative search. Cognition, 106(3):1465?1477, 2008.

5. A. S. Chetwood, K.-W. Kwok, L.-W. Sun, G. P. Mylonas, J. Clark, A. Darzi, and G.-Z. Yang. Collaborative eye tracking: a potential training tool in laparoscopic surgery. Surgical endoscopy, 26(7):2003?2009, 2012.

6. Z. M. Griffin and K. Bock. What the eyes say about speaking. Psychological science, 11(4):274?279, 2000.

7. R. F. Kizilcec, K. Papadopoulos, and L. Sritanyaratana. Showing face in video instruction: effects on information retention, visual attention, and affect. In Proceedings of the 32nd annual ACM conference on Human factors in computing systems, pages 2095?2102. ACM, 2014.

8. H. Prendinger, T. Eichner, E. Andr?e, and M. Ishizuka. Gaze-based infotainment agents. In Proceedings of the international conference on Advances in computer entertainment technology, pages 87?90. ACM, 2007.

9. D. C. Richardson, R. Dale, and N. Z. Kirkham. The art of conversation is co-ordination common ground and the coupling of eye movements during dialogue. Psychological science, 18(5):407?413, 2007.

10. K. Sharma, P. Jermann, and P. Dillenbourg. How students learn using moocs: An eye-tracking insight. In EMOOCs 2014, the Second MOOC European Stakeholders Summit, 2014.

11. K. Sharma, P. Jermann, and P. Dillenbourg. "With-me-ness" : A gaze-measure for students? attention in moocs. In International conference of the learning sciences, 2014.

12. D. A. Slykhuis, E. N. Wiebe, and L. A. Annetta. Eye-tracking students? attention to powerpoint photographs in a science education setting. Journal of Science Education and Technology, 14(5-6):509?520, 2005.

13. T. Van Gog, H. Jarodzka, K. Scheiter, P. Gerjets, and F. Paas. Attention guidance during example study via the model's eye movements. Computers in Human Behavior, 25(3):785?791, 2009.

14. T. Van Gog, F. Paas, J. J. van Merri?enboer, and P. Witte. Uncovering the problem-solving process: cued retrospective reporting versus concurrent and retrospective reporting. Journal of Experimental Psychology: Applied, 11(4):237, 2005.

15. R. Vertegaal, I. Weevers, and C. Sohn. Gaze-2: An attentive video conferencing system. In CHI'02 Extended Abstracts on Human Factors in Computing Systems, pages 736?737. ACM, 2002.

16. N. Li, L. Kidzinski, P. Jermann and P. Dillenbourg. How do In-video interactions Reflect Perceived Video Difficulty. In EMOOCs 2015, the third MOOC European Stakeholders Summit, 2015.

17. Hanna, J. E., and Brennan, S. E. (2007). Speakers' eye gaze disambiguates referring expressions early during face-to-face conversation. Journal of Memory and Language, 57(4), 596-615.

18. Gergle, D., and Clark, A. T. (2011). See what I'm saying? using Dyadic Mobile Eye tracking to study collaborative reference. In Proceedings of the ACM 2011 conference on Computer supported cooperative work (pp. 435-444). ACM.

19. Prasov, Z., and Chai, J. Y. (2008). What's in a gaze: the role of eye-gaze in reference resolution in multimodal conversational interfaces. In Proceedings of the 13th international conference on Intelligent user interfaces (pp. 20-29). ACM.