Exploring Gesture and Gaze Proxies to Communicate Instructor's Nonverbal Cues in Lecture Videos

Tobias Wagner
Ulm University, Institute of Media Informatics
Ulm, Germany
tobias.wagner@uni-ulm.de

Teresa Hirzle
University of Copenhagen, Department of Computer Science
Copenhagen, Denmark
tehi@di.ku.dk

Enrico Rukzio
enrico.rukzio@uni-ulm.de
Ulm University, Institute of Media Informatics
Ulm, Germany
enrico.rukzio@uni-ulm.de

Anke Huckauf
anke.huckauf@uni-ulm.de
Ulm University, Institute of Psychology and Education
Ulm, Germany
anke.huckauf@uni-ulm.de

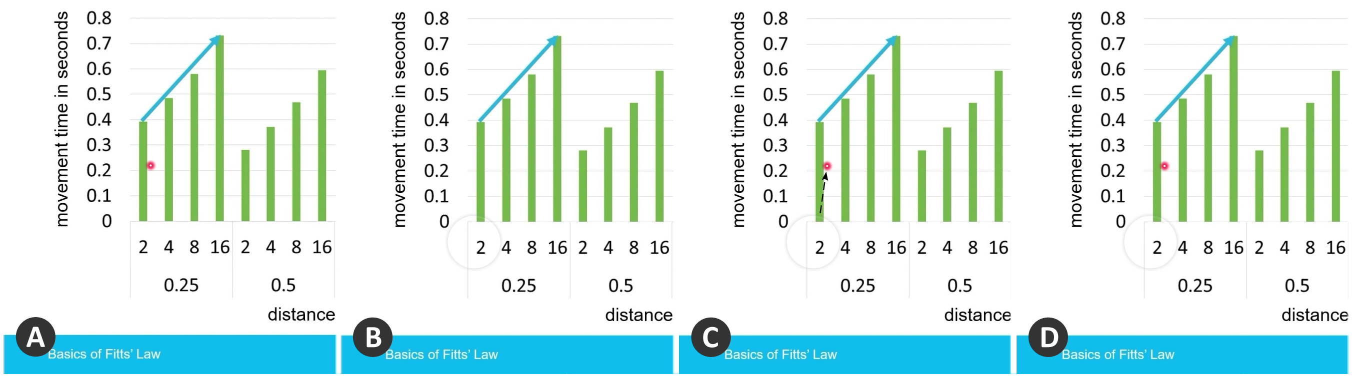

Figure 1: We implemented four proxy types onto an HCI lecture video to communicate the instructor's pointing gestures and gaze to the students: (A) gesture proxy, (B) gaze proxy, (C) alternating proxy (i.e., switching between gesture and gaze proxy), and (D) concurrent proxies (i.e., displaying gesture and gaze proxy simultaneously). ***The HCI lecture slides are licensed under CC BY 4.0 ©Niels Henze [8].***

ABSTRACT

Teaching via lecture video has become the defacto standard for remote education, but videos make it difficult to interpret instructors' nonverbal referencing to the content. This is problematic, as nonverbal cues are essential for students to follow and understand a lecture. As remedy, we explored different proxies representing instructors' pointing gestures and gaze to provide students a point of reference in a lecture video: no proxy, gesture proxy, gaze proxy, alternating proxy, and concurrent proxies. In an online study with 100 students, we evaluated the proxies' effects on mental effort, cognitive load, learning performance, and user experience. Our results show that the proxies had no significant effect on learning-directed aspects and that the gesture and alternating proxy achieved the highest pragmatic quality. Furthermore, we found that alternating between proxies is a promising approach providing students with information about instructors' pointing and gaze position in a lecture video.

CCS CONCEPTS

• **Human-centered computing → Human computer interaction (HCI)**; • **Applied computing → Distance learning**.

KEYWORDS

Gaze, Gesture, Education, Lecture video, Eye-tracking

CHI EA '23, April 23–28, 2023, Hamburg, Germany

ACM ISBN 978-1-4503-9422-2/23/04.
https://doi.org/10.1145/3544549.3585842

ACM Reference Format:
Tobias Wagner, Teresa Hirzle, Enrico Rukzio, and Anke Huckauf. 2023. Exploring Gesture and Gaze Proxies to Communicate Instructor's Nonverbal Cues in Lecture Videos. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems (CHI EA '23), April 23–28, 2023, Hamburg, Germany.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3544549.3585842

1 INTRODUCTION

In our digital and connected world, education is increasingly shifting from classrooms to online, enabling synchronous and asynchronous teaching of a spatially distributed audience. Lecture videos have become a common approach in higher education. They appear in a variety of styles, from recorded live classrooms to recordings of lectures held online via video conferencing systems (e.g., Zoom [7]) or lectures recorded via screen recording software [16, 38]. In recordings of classroom lectures, the instructional material is typically presented as slides on a whiteboard sharing the same physical environment as the instructor. Here, the instructor can refer to information on the whiteboard using pointing gestures and looking at it while explaining it [20, 33]. According to van Gog [35], the instructor's pointing gestures and gaze direction towards information in the material is a way of attention guidance based on the *signaling principle*.

In computer-supported teaching, instructional material is often shared virtually together with an on-screen instructor (i.e., a video integrated into the learning material, showing the instructor's face and a part of their upper body) [37]. Since learning material and instructor are spatially separated, interpreting embodied signaling cues is challenging for students. To overcome this challenge, prior work has investigated using a mouse or gaze-aware cursor as a proxy to remotely convey pointing gestures [17] or gaze [10]. In the context of remote teaching through lecture videos, these proxies have been applied with the intention to direct students' attention towards relevant information to foster students' learning [11, 29]. However, prior work mainly investigated gesture and gaze proxies independently from each other, neglecting the relation between gesture and eye movements (i.e., hand-eye coordination) [18], which might provide important information in teaching scenarios [11]. In this work, we address this limitation by exploring different combinations of gesture and gaze proxies to communicate instructors' nonverbal cues to improve students' learning experience.

We conducted a user study with 100 students in which we measured students' *mental effort*, *cognitive load*, *learning performance*, and *user experience* while watching a five-part lecture video, where an instructor was using different gesture and gaze proxies. In each video part, one of the following proxy conditions was used by the instructor: no proxy (baseline), gesture proxy, gaze proxy, alternating proxy, and concurrent proxies. Further, we asked the students which proxy version they preferred and which they found most distracting.

The results show that the different proxy versions did not influence students' mental effort, cognitive load, and learning performance. However, students preferred the gesture proxy the most and were least distracted by the separate gesture and gaze proxies. Furthermore, using a gesture or an alternating proxy in a lecture video achieved the highest pragmatic quality, and students experienced the complementary nature of this proxy as helpful, as it provides instructor's gestural and gaze information. While they liked the concept of combining gesture and gaze information into one proxy, they perceived the current implementation of the transition between the two proxy types as distracting.

In summary, we make the following contributions: Firstly, we provide insights about the effects of gesture and gaze proxies for instructors to integrate referencing over nonverbal cues in lecture videos. Secondly, our work shows that although students like the concept of combining gesture and gaze proxy, they preferred using individual proxies in goal-directed aspects of the user experience. However, our work provides promising steps to design complementary proxies that combine gesture and gaze information.

2 RELATED WORK

Our work is related to prior research on *the signaling principle*, *gesture proxies*, and *gaze proxies*.

The theoretical foundation for the use of proxies is based on the *signaling principle*, which aims to reduce cognitive load induced by the design and complexity of the instructional material [35]. The principle recommends instructors to add visual cues that guide students' attention on essential elements or highlight the structure of the instructional material. According to the cognitive theory of multimedia learning by Mayer [22], such signaling can support students' in the initial state of active processing, namely the selection and processing of relevant information in the working memory [35].

Slide presentation software such as *PowerPoint* [23] already provide instructors tools for pointing at the shared material with gesture proxies represented as a virtual laser pointer. However, we are unaware of any empirical investigations about the effectiveness of using these proxies. Sauter et al. [29] conducted an eye-tracking study to investigate the effect of an instructor's gesture proxy, represented as a virtual laser pointer, on students' eye movements and learning outcomes when watching lecture videos. Results indicated that students whose gaze position was closer to the proxy's position achieved higher scores in a learning quiz than those with a larger distance. These results indicate that students' learning was enhanced when they paid attention to the gesture proxy.

In addition to the use of gesture proxies, gaze sharing over a gaze proxy has been well studied in the area of collaborative work [10] and education [32, 34]. For instance, Sharma et al. [31] compares the effects of a gaze proxy, represented by a scan path visualization, and a gesture proxy, represented by a pen pointer, in a lecture video on students' learning gain and degree of following the instructor. The findings show that students watching the lecture video with instructor's gaze proxy had a higher degree of following the instructor and achieved a higher learning gain than students watching videos with the gesture proxy or without any proxy. Unfortunately, an investigation of a combination of both proxies is missing here. Complementing these results, a recent study by Emhardt et al. [11] examined the effects of gesture proxies, gaze proxies, and their combination in a lecture video on students' mental effort and learning outcomes, as well as students' ease of following the instructor. In contrast to prior work, Emhardt et al. [11] investigated the effect of combining both proxies by displaying them concurrently. Their findings revealed no significant effect on students' mental effort and learning outcomes when gesture proxy, gaze proxy, or both proxies were present. However, students watching lecture videos with the gaze proxy found it easier to follow the instructor than in the other conditions. Different from the work of Emhardt et al. [11], we investigate an additional combination method that shows a single proxy that is alternating between gesture and gaze in accordance to instructors' mouse movements. With that, we aim to avoid inducing negative effects such as split-attention that could arise from the concurrent presentation of multiple proxies [4].

In summary, the aforementioned works suggest that the use of gesture and gaze proxies in educational videos can support students in following the presentation of the instructional material, thereby promoting meaningful learning. Our work aims to expand upon the body of research by exploring how two alternative methods of combining gesture and gaze proxies are evaluated compared to a single gesture or gaze proxy.

3 METHOD

We conducted a within-participant online study with $N = 100$ students to explore the influence of four different proxy types used by instructors in a lecture video on students' *mental effort*, *cognitive load*, *learning performance*, and *user experience*. Additionally, we asked students to rank the proxies in terms of *preference* and *distraction*. The type of proxy is our independent variable (IV) with five levels: using (a) no proxy, (b) gesture proxy, (c) gaze proxy, (d) alternating proxy, and (e) concurrent proxies.

3.1 Study Material

The proxy conditions were implemented using *Adobe After Effects* [1] to design visual pointers based on existing pointing tools in *Microsoft PowerPoint* and shared gaze visualizations from prior work [9, 10] (see Figure 1). Then, we recorded two instructors giving a 10-min lecture about *Fitts' Law*, based on copyright Niels Henze's [8] lecture on models in HCI, while their mouse and eye movements (*Pupil Core* [15] eye-tracking headset) were recorded. Eye movement data were smoothed by downsampling and using a dispersion-based fixation filter [21] and low pass filter before being applied to the gaze proxy in *Adobe After Effects*. The mouse movements were applied directly to the gesture proxy visualization. For the alternating proxy, we implemented an algorithm that displays the gaze proxy unless the instructor moved their mouse. In the concurrent proxy condition, gesture and gaze proxy were displayed simultaneously. Finally, proxy conditions were rendered onto the lecture video and it was divided into five two-minute parts.

3.2 Measures

Mental Effort and Cognitive Load. To assess mental effort and cognitive load, we used the 9-point mental-effort rating scale by Paas et al. [25] (*1 = very, very low mental effort* to *9 = very, very high mental effort*) and the 7-point subjective rating scale developed by Klepsch et al. [19] to measure students' intrinsic, extraneous, and germane load differently (*absolutely wrong* to *absolutely right*).

Prior-Knowledge and Learning Performance. According to Mayer [22], students' prior knowledge can influence the effect on cognitive load when their attention is guided with visual cues [2, 35]. Therefore, we designed a prior knowledge test consisting of knowledge-, comprehension-, and transfer-based questions based on *Bloom's Taxonomy for Learning, Teaching and Assessing* [3]. Tests after each video part to assess students' learning performance were designed similarly.

User Experience. We used the short version of the *User Experience Questionnaire* (UEQ-S) [30] to assess students' impression towards their user experience of the proxies. The questionnaire includes eight pairs of contrasting attributes representing the proxies *hedonic* and *pragmatic* quality (e.g., *obstructive - supportive*). Students rated on a 7-point scale to which extent one of the attributes matched their impression.

Preference, Distraction, and Qualitative Feedback. For preference and distraction assessment, students' had to rank proxy conditions according to their preference (*1 = most preferred* to *4 = least preferred*) and experienced distraction (*1 = most distracting* to *4 = least distracting*). To obtain qualitative feedback, students had to explain their ranking decision and provide their opinion about the advantages, disadvantages, and improvements of each proxy.

3.3 Task

Participants watched a pre-recorded lecture video that was divided into five 2-minute segments, each overlaid with one of the five proxy conditions. The conditions were counter-balanced by a 5x5 Latin square to prevent order effects [28]. During the task, they were not allowed to take any notes. After each part, participants were tested on their learning performance through a short test consisting of knowledge, comprehension, and transfer questions.

3.4 Procedure

We conducted the study using *Prolific* [27] and *LimeSurvey* [14]. On *Prolific*, we recruited enrolled students with a current education level of undergraduate or graduate degree and who were fluent in English. After accepting the study invite, the participants first signed an informed consent. They then received instructions on their task and completed a demographic questionnaire and prior knowledge assessment. They then watched the first part of the five-part lecture video and indicated their invested mental effort, learning performance, and cognitive load. For video parts that include a proxy, participants rated their experience of the proxy. They repeated this process until they completed the entire lecture video. In the end, they ranked the proxies towards preference and distraction, and provided feedback. The study lasted approximately 60 minutes, and participants received an expense allowance of €10 on study completion.

3.5 Participants

We recruited 100 participants with study experience (45 identified as female, 41 as male, five as non-binary, and nine preferred not to answer) with an average age of 22.46 ($SD = 2.89$, $min = 19$, $max = 37$). Regarding participants' education level, 46 of the participants hold a high school degree, 46 a bachelor's degree, five a master's degree, and three a secondary degree. On average, participants scored 6.54 points of 15 possible points in the prior knowledge test ($SD = 2.4$, $min = .6$, $max = 11.85$).

4 RESULTS

In the following, we present the analysis of $N = 100$ responses. First, we tested differences in students' *mental effort*, *cognitive load*, *learning performance*, and *user experience*, as well as *preference* and *distraction ranking* over the IVs with a quantitative analysis. Differences were considered significant for $p < .05$. Secondly, we summarize students' qualitative feedback on the *preference* and *distraction ranking*, as well as *advantages*, *disadvantages*, and *improvements* of the proxy conditions. We present a summary of the findings in Table 1.

4.1 Quantitative Analysis

We used non-parametric tests for the differences of the dependent variables (DV) between the proxy conditions as Shapiro-Wilk test results indicated that the variables' distributions differed significantly from a normal distribution.

For students' *mental effort*, *cognitive load*, *learning performance*, and *user experience*, we fitted generalized linear mixed models (GLMM) with each DV as the outcome, the proxy conditions (*no proxy* as reference) and video part as fixed effects, and the students, as well as their prior knowledge as random effects. Differences were analyzed using Type-II Wald χ^2 tests, and pairwise comparisons were conducted using Tukey's post-hoc tests. The differences in the *preference and distraction ranking* were analyzed using a Friedman rank-sum test and following pairwise comparison using post-hoc Wilcoxon signed-rank tests with Holm's sequential Bonferroni correction. The statistical analysis was conducted using *R* (version 4.2.1) [13].

Mental Effort. The analysis of deviance table using Type-II Wald χ^2 test on a fitted cumulative GLMM indicated no significant main effect of proxies on students' mental effort ($\chi^2(4, N = 100) = .935$, $p = .9195$), but a significant main effect of the video parts on students' mental effort ($\chi^2(4, N = 100) = 75.657$, $p < .001^{***}$, $r = -.36$). A following pairwise comparison using Tukey's post-hoc test indicated a significant difference between video parts A and B, A and C, A and D, A and E, as well as between B and C, B and D, B and E.

Cognitive Load. Cognitive load is analyzed separately for extraneous, intrinsic, and germane load by averaging the corresponding items. The analysis of deviance tables using Type-II Wald χ^2 test on fitted gamma (inverse) GLMMs showed no significant effect of proxies on intrinsic ($\chi^2(4, N = 100) = 1.08$, $p = .898$) and germane load ($\chi^2(4, N = 100) = 3.05$, $p = .55$), but a significant main effect on extraneous load ($\chi^2(4, N = 100) = 10.46$, $p = .0333^{*}$, $r = -.095$). However, the pairwise comparisons of proxy conditions using Tukey's post-hoc test were not statistically significant.

Learning Performance. Students' answers to the learning performance test after each video part were scored according to an assessment scheme and then normalized between 0.0, meaning no points, and 1.0, meaning full points. The analysis of deviance table on a fitted beta (logit) GLMM showed no significant effect of proxies on the learning performance ($\chi^2(4, N = 100) = 3.99$, $p = .41$).

User Experience. The analysis of deviance tables using Type-II Wald χ^2 test on fitted (inverse) GLMMs indicated a statistically significant effect on *pragmatic quality* of proxies ($\chi^2(3, N = 100) = 11.34$, $p = .01^{**}$, $r = -.13$), but no statistically significant effect on *hedonic quality* ($\chi^2(3, N = 100) = .32$, $p = .96$). Students rated the *pragmatic quality* of the gesture ($M = 1.01$, $SD = 1.57$) and alternating proxy ($M = 0.85$, $SD = 1.67$) significantly higher than concurrent proxies ($M = 0.23$, $SD = 1.82$). This was confirmed through a pairwise comparison using Tukey's post-hoc test that indicated a significant difference between gesture proxy and concurrent proxies ($Z = -3.247$, $p < .01^{**}$) and between concurrent proxies and alternating proxy ($Z = 2.599$, $p < 0.05^{*}$). There was no significant difference of the *pragmatic quality* between the gaze proxy ($M = 0.71$, $SD = 1.72$) and the other proxy conditions.

Preference and Distraction Ranking. The ranking data were analyzed using Friedman rank-sum tests. Regarding the preference ranking, significant differences between all proxy conditions were revealed by the Friedman test ($\chi^2(3, N = 100) = 65.68$, $p < .001^{***}$) and a pairwise comparison using post-hoc Wilcoxon signed-rank tests with Holm's sequential Bonferroni correction. Students preferred the gesture proxy the most ($M = 1.78$, $SD = 0.99$), followed by the gaze proxy ($M = 2.26$, $SD = 1.02$), alternating proxy ($M = 2.81$, $SD = 1.01$), and concurrent proxies ($M = 3.15$, $SD = 0.95$).

In terms of the distraction of proxies, the Friedman test showed significant differences between the proxy conditions ($\chi^2(3, N = 100) = 66.04$, $p < .001^{***}$). The following post-hoc Wilcoxon signed-rank tests with Holm's sequential Bonferroni correction showed a significant difference between the proxy conditions except for the difference between gesture and gaze proxy ($Z = 1.45$). Students ranked the concurrent proxies as the most distracting ones ($M = 1.80$, $SD = 0.96$), followed by the alternating proxy ($M = 2.21$, $SD = 0.95$). Gesture ($M = 3.13$, $SD = 1.05$) and gaze proxy ($M = 2.86$, $SD = 1.01$) were ranked as the least distracting ones.

4.2 Qualitative Feedback

We asked participants to explain their preference and distraction ranking decision and about advantages, disadvantages, and improvements for each proxy condition. For the analysis, we followed an informal process inspired by thematic analysis [6] and affinity diagramming. We first labeled students' statements using adjectives that summarize students' impression towards the proxy to be evaluated (e.g., "using two proxies simultaneously is very *distracting*") Then, adjectives with a similar assertion were combined to top-level adjectives, and the number of associated statements was counted (e.g., *distracting* and *disturbing* → *distracting*). This process was conducted by the first author of this submission, but the resulting top-level construct was discussed and refined with the other authors of this paper.

Gesture Proxy. On the one hand, students stated that the single gesture proxy is *meaningful* ($N = 32$), *attention-drawing* ($N = 25$), and *clear* ($N = 23$). On the other hand, they were discontent with its color, described as *glaring* ($N = 23$) *"The [gesture proxy] is fine, but it moves really quickly and has a bright color, which at times can be an obstacle. [...]"* [P89], thus, suggesting a lighter ($N = 16$) and less bright ($N = 14$) color for the proxy.

Gaze Proxy. In contrast to the gesture proxy, students indicated that the gaze proxy was *imperceptible* ($N = 42$), but then again *unobtrusive* ($N = 32$). They found it beneficial that the gaze proxy has a *focus on an area* of information ($N = 22$) and establishes *gaze-awareness* ($N = 25$) as stated by [P80]: *"I like the [gaze proxy] because it lets you follow the thought of the [instructor], as you are looking*

DV	Findings
Mental effort	No significant main effect of the different proxies on mental effort, but a significant main effect of the video parts on mental effort.
Cognitive load	No significant main effect of the proxies on intrinsic and germane load, but a significant main effect of the proxies on extraneous load (but without group differences).
Learning performance	No significant main effect of the proxies on learning performance.
User experience	Significant main effect of the proxies on pragmatic quality, with the gesture and the alternating proxy being more pragmatically relevant than concurrent proxies. No significant effect of the proxies on hedonic quality.
Preference ranking	gesture > gaze > alternating > concurrent
Distractibility ranking	gesture and gaze < alternating < concurrent
Qualitative feedback	Students were most positive about the gesture and gaze proxy. The alternating proxy and concurrent proxies were perceived as distracting. However, students liked the alternating proxy as it provides where the instructor is pointing and looking at. Regarding improvements, students consistently asked for changes in the proxies' appearance.

Table 1: Summary of findings.

at the same things.". Students suggested to *improve the visibility* ($N = 22$) of the proxy by adding color or by increasing its opacity.

Alternating Proxy. Towards the alternating proxy, students indicated that it was *distracting* ($N = 42$), in particular, the transition ($N = 33$) as stated by [P77]: *"It was confusing when the [alternating proxy] interplayed. [...]"*. However, they found it still *useful* ($N = 28$) and liked that it was *complementary* ($N = 15$) as it combines instructor's gestures and gaze but in a single proxy ($N = 14$). They suggested revising the transition by making it less distracting and performing it in the right moment ($N = 7$) as stated by [P17]: *"Be mindful of the exchange between one and the other: do not do it mid explanation or sentence."*. Furthermore, they want to change the proxy's appearance ($N = 11$).

Concurrent Proxies. Regarding the concurrent proxies, students disliked that it was *distracting* ($N = 42$) and *divided their attention* ($N = 27$). However, they liked that it allows *multi-pointing* on different locations simultaneously ($N = 14$) and that it is *complementary* ($N = 12$) and visualizes instructor's *hand-eye coordination* ($N = 10$), for example, stated by [P5]: *"[Gesture proxy] might help highlight something within the area selected by the [gaze proxy]."*. The main suggestion for improvement was to show a single proxy instead of two simultaneously, for example, by hiding the one that is currently not used ($N = 23$).

5 DISCUSSION

Our work explored the effect of proxies representing instructor's pointing gestures and gaze in lecture videos on student's *mental effort*, *cognitive load*, and *learning performance*. We could not find an effect of the different proxies on students' *mental effort* and *cognitive load*, nor on their *learning performance*. These findings are consistent with prior work conducted by Emhardt et al. [11], who also did not find an effect on students' *mental effort* and *learning performance* when gesture, gaze, or concurrent proxies were present in a lecture video.

Contrary to the assumption that signaling can reduce students' extraneous load by preventing temporal split-attention between the instructor's verbal narration and the presented instructional material [4, 35], we did not find any differences in *extraneous load* between the proxy conditions. One possible explanation may be the well-designed presentation we used, which utilized fade-in effects to display texts and images step-by-step. These might have been sufficient to structure the presentation and highlight relevant information for the students, making the effect of using proxies minimal. In the presentation, signaling may only be effective for complex information, such as diagrams shown and explained in video parts C, D, and E [2]. For these video parts, students required the most *mental effort*. An alternative explanation may be that the complexity of the learning material and the learning test were unsuitable for inducing an effect on student's learning performance [12].

Despite this, it is noteworthy that even instructor's gesture proxy, a commonly used tool in presentations, did not have an effect on students although they rated it as more meaningful than the gaze proxy. This is contrary to the findings of Sauter et al. [29], who found that students who followed the proxy achieved a better learning performance, suggesting that students may not have paid attention to our proxies. However, we cannot determine if students' attention was successfully guided since we did not measure their eye movements in the study. We plan to run a follow-up eye-tracking study to analyze students' attention allocation in the lecture video with the proxies.

Several works found that instructor's pointing gestures and gaze conveyed over video enhance students' learning performance [5, 24, 26, 36]. However, we could not replicate these findings with our implementation of proxies. Reasons for this might be the well-designed presentation or a lack of attention, as discussed before.

Regarding the results of students' *user experience*, the *preference and distraction ranking*, as well as the *qualitative feedback*, students rated the *pragmatic quality* of the gesture and the alternating proxy significantly higher than the concurrent proxies' quality. This indicates that the gesture and alternating proxy support students more in goal-directed aspects than the gaze proxy and the concurrent proxies. Nevertheless, students preferred the gesture proxy most, probably because it was perceived as the least distracting one, as indicated by the ranking. The gaze proxy was ranked similarly in the distracting and second in the preference ranking. Therefore, we assume that with our implementation of the gaze proxy, we successfully reduced its distracting aspects caused by the natural fast and rapid movements of the human's eye [9], which is the main reason why previous works found gaze proxies to be difficult to follow. With the alternating proxy, we also provided a method combining gesture and gaze proxy without displaying them simultaneously, preventing split-attention for students.

The qualitative feedback of students emphasizes the benefits of the gesture or gaze proxy over the alternating proxy or concurrent proxies in terms of distraction. However, students liked the concept of the alternating proxy, as it provides them with more information from the instructor. Furthermore, students suggestions for improvement mainly address the appearance of the proxies, which indicates that our proxy design could have an impact on students learning performance.

In summary, our findings raise questions about the appropriate use and design of proxies and their effectiveness in lecture videos, particularly in poorly designed or complex presentations. Therefore, future research should investigate proxies in more complex presentations to find out when and for which type of information in videos proxies are helpful.

6 CONCLUSION

We investigated the effect of proxies conveying instructor's pointing gestures and gaze remotely in lecture videos on students' *mental effort*, *cognitive load*, and *learning performance*. Additionally, we assessed students' *user experience*, as well as which proxy type they *preferred* and which they found most *distracting*. In a user study, students watched a recorded 5-part lecture video, where each part was overlaid by either (a) no proxy, (b) a gaze proxy, (c) a gesture proxy, (d) an alternating proxy, or (d) concurrent proxies. Results showed that none of the proxy types had a positive effect on students' *mental effort*, *cognitive load*, or *learning performance*. In terms of *user experience*, the gesture and alternating proxy achieved a higher *pragmatic quality* than the concurrent proxies. Students preferred the gesture proxy the most and ranked the gesture and gaze proxy as the least distracting ones. However, students' qualitative feedback reveals that although the alternating proxy was distracting, they liked that it provided information about where the instructor is pointing or looking at. Suggestions for improvements were given to the appearance of the proxies. We conclude that, in terms of students' mental effort, cognitive load, and learning performance, using a combination of gesture and gaze proxy is not more beneficial than using an individual or no proxy. However, proxies might be benefit more complex presentations or learning material. Therefore, further research is needed to investigate the appropriate usage and design of proxies, particularly the combination of gesture and gaze proxy, to enhance students' learning with lecture videos.

ACKNOWLEDGMENTS

This project is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation, 425867974, to Anke Huckauf and Enrico Rukzio) and is part of Priority Program SPP2199 Scalable Interaction Paradigms for Pervasive Computing Environments. T. Hirzle was supported by the HumanE AI Network from the European Union's Horizon 2020 research and innovation program under grant agreement No 952026, and the Pioneer Centre for AI, DNRF grant number P1.

REFERENCES

[1] Adobe. 2023. *Adobe After Effects*. Retrieved January 18, 2023 from https://www.adobe.com/products/aftereffects.html

[2] David Alpizar, Olusola O. Adesope, and Rachel M. Wong. 2020. A Meta-Analysis of Signaling Principle in Multimedia Learning Environments. *Educational Technology Research and Development* 68, 5 (Oct. 2020), 2095–2119. https://doi.org/10.1007/s11423-020-09748-7

[3] Lorin W. Anderson, David R. Krathwohl, Peter W. Airasian, Kathleen A. Cruikshank, Richard E. Mayer, Paul R. Pintrich, James Raths, and Merlin C. Wittrock. 2000. *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives, Complete Edition*. Pearson Education, New York.

[4] Paul Ayres and John Sweller. 2021. The Split-Attention Principle in Multimedia Learning. In *The Cambridge Handbook of Multimedia Learning* (third ed.), Logan Fiorella and Richard E. Mayer (Eds.). Cambridge University Press, Cambridge, 199–211. https://doi.org/10.1017/9781108894333.020

[5] Maik Beege, Manuel Ninaus, Sascha Schneider, Steve Nebel, Julia Schlemmel, Jasmin Weidenmüller, Korbinian Moeller, and Günter Daniel Rey. 2020. Investigating the Effects of Beat and Deictic Gestures of a Lecturer in Educational Videos. *Computers & Education* 156 (Oct. 2020), 103955. https://doi.org/10.1016/j.compedu.2020.103955

[6] Virginia Braun and Victoria Clarke. 2006. Using Thematic Analysis in Psychology. *Qualitative Research in Psychology* 3, 2 (Jan. 2006), 77–101. https://doi.org/10.1191/1478088706qp063oa

[7] Zoom Video Communications. 2023. *Zoom Meetings*. Retrieved January 18, 2023 from https://explore.zoom.us/en/products/meetings/

[8] ©Niels Henze. 2022. *Models for HCI*. Retrieved January 18, 2023 from https://hci-lecture.org/models-for-hci/ *This work is licensed under the Creative Commons Attribution 4.0 International License. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/*.

[9] Sarah D'Angelo and Darren Gergle. 2016. Gazed and Confused: Understanding and Designing Shared Gaze for Remote Collaboration. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. Association for Computing Machinery, New York, NY, USA, 2492–2496. https://doi.org/10.1145/2858036.2858499

[10] Sarah D'Angelo and Bertrand Schneider. 2021. Shared Gaze Visualizations in Collaborative Interactions: Past, Present and Future. *Interacting with Computers* 33, 2 (May 2021), 115–133. https://doi.org/10.1093/iwcomp/iwab015

[11] Selina N. Emhardt, Halszka Jarodzka, Saskia Brand-Gruwel, Christian Drumm, Diederick C. Niehorster, and Tamara van Gog. 2022. What Is My Teacher Talking about? Effects of Displaying the Teacher's Gaze and Mouse Cursor Cues in Video Lectures on Students' Learning. *Journal of Cognitive Psychology* 0, 0 (2022), 1–19. https://doi.org/10.1080/20445911.2022.2080831 arXiv:https://doi.org/10.1080/20445911.2022.2080831

[12] Logan Fiorella and Richard E. Mayer. 2021. Principles for Reducing Extraneous Processing in Multimedia Learning: Coherence, Signaling, Redundancy, Spatial Contiguity, and Temporal Contiguity Principles. In *The Cambridge Handbook of Multimedia Learning* (third ed.), Logan Fiorella and Richard E. Mayer (Eds.). Cambridge University Press, Cambridge, 185–198. https://doi.org/10.1017/9781108894333.019

[13] The R Foundation. 2022. *R: The R Project for Statistical Computing*. Retrieved January 16, 2023 from https://www.r-project.org/

[14] LimeSurvey GmbH. 2021. *LimeSurvey*. Retrieved January 12, 2023 from https://www.limesurvey.org/

[15] Pupil Labs GmbH. 2023. *Pupil Labs Pupil Core Eye-Tracking Headset*. Retrieved January 18, 2023 from https://pupil-labs.com/products/core/

[16] Philip J. Guo, Juho Kim, and Rob Rubin. 2014. How Video Production Affects Student Engagement: An Empirical Study of MOOC Videos. In *Proceedings of the First ACM Conference on Learning @ Scale Conference (L@S '14)*. Association for Computing Machinery, New York, NY, USA, 41–50. https://doi.org/10.1145/2556325.2566239

[17] Stephen Hayne, Mark Pendergast, and Saul Greenberg. 1993. Implementing Gesturing with Cursors in Group Support Systems. *Journal of Management Information Systems* 10, 3 (Dec. 1993), 43–61. https://doi.org/10.1080/07421222.1993.11518010

[18] Jeff Huang, Ryen White, and Georg Buscher. 2012. User See, User Point: Gaze and Cursor Alignment in Web Search. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. Association for Computing Machinery, New York, NY, USA, 1341–1350. https://doi.org/10.1145/2207676.2208591

[19] Melina Klepsch, Florian Schmitz, and Tina Seufert. 2017. Development and Validation of Two Instruments Measuring Intrinsic, Extraneous, and Germane Cognitive Load. *Frontiers in Psychology* 8 (2017), 1997.

[20] Hubert Knoblauch. 2008. The Performance of Knowledge: Pointing and Knowledge in Powerpoint Presentations. *Cultural Sociology* 2, 1 (March 2008), 75–97. https://doi.org/10.1177/1749975507086275

[21] Manu Kumar, Jeff Klingner, Rohan Puranik, Terry Winograd, and Andreas Paepcke. 2008. Improving the Accuracy of Gaze Input for Interaction. In *Proceedings of the 2008 Symposium on Eye Tracking Research & Applications (ETRA '08)*. Association for Computing Machinery, New York, NY, USA, 65–68. https://doi.org/10.1145/1344471.1344488

[22] Richard E. Mayer. 2021. Cognitive Theory of Multimedia Learning. In *The Cambridge Handbook of Multimedia Learning* (third ed.), Richard E. Mayer and Logan Fiorella (Eds.). Cambridge University Press, 57–72. https://doi.org/10.1017/9781108894333.008

[23] Microsoft. 2022. *Turn Your Mouse into a Laser Pointer.* Retrieved January 18, 2023 from https://support.microsoft.com/en-us/office/turn-your-mouse-into-a-laser-pointer-77367b36-d25b-4ed2-8c87-358bc216a1e0

[24] Kim Ouwehand, Tamara van Gog, and Fred Paas. 2015. Designing Effective Video-Based Modeling Examples Using Gaze and Gesture Cues. *Journal of Educational Technology & Society* 18, 4 (2015), 78–88.

[25] Fred G. W. C. Paas, Jeroen J. G. van Merriënboer, and Jos J. Adam. 1994. Measurement of Cognitive Load in Instructional Research. *Perceptual and Motor Skills* 79, 1 (Aug. 1994), 419–430. https://doi.org/10.2466/pms.1994.79.1.419

[26] Zhongling Pi, Yi Zhang, Fangfang Zhu, Ke Xu, Jiumin Yang, and Weiping Hu. 2019. Instructors' Pointing Gestures Improve Learning Regardless of Their Use of Directed Gaze in Video Lectures. *Computers & Education* 128 (Jan. 2019), 345–352. https://doi.org/10.1016/j.compedu.2018.10.006

[27] Prolific. 2022. *Prolific Crowdsourcing Platform.* Retrieved January 12, 2023 from https://www.prolific.co/

[28] John T. E. Richardson. 2018. The Use of Latin-square Designs in Educational and Psychological Research. *Educational Research Review* 24 (June 2018), 84–97. https://doi.org/10.1016/j.edurev.2018.03.003

[29] Marian Sauter, Tobias Wagner, and Anke Huckauf. 2022. Distance between Gaze and Laser Pointer Predicts Performance in Video-Based e-Learning Independent of the Presence of an on-Screen Instructor. In *2022 Symposium on Eye Tracking Research and Applications (ETRA '22)*. Association for Computing Machinery, New York, NY, USA, 1–10. https://doi.org/10.1145/3517031.3529620

[30] Martin Schrepp, Andreas Hinderks, and Jörg Thomaschewski. 2017. Design and Evaluation of a Short Version of the User Experience Questionnaire (UEQ-S). 4 (2017), 103–108. https://doi.org/10.9781/ijimai.2017.09.001

[31] Kshitij Sharma, Sarah D'Angelo, Darren Gergle, and Pierre Dillenbourg. 2016. Visual Augmentation of Deictic Gestures in MOOC Videos. (July 2016).

[32] Kshitij Sharma, Patrick Jermann, and Pierre Dillenbourg. 2015. Displaying Teacher's Gaze in a MOOC: Effects on Students' Video Navigation Patterns. In *Design for Teaching and Learning in a Networked World (Lecture Notes in Computer Science)*, Gráinne Conole, Tomaž Klobučar, Christoph Rensing, Johannes Konert, and Elise Lavoué (Eds.). Springer International Publishing, Cham, 325–338. https://doi.org/10.1007/978-3-319-24258-3_24

[33] Yi Tian and Marie-Luce Bourguet. 2016. Lecturers' Hand Gestures as Clues to Detect Pedagogical Significance in Video Lectures. In *Proceedings of the European Conference on Cognitive Ergonomics*. ACM, Nottingham United Kingdom, 1–3. https://doi.org/10.1145/2970930.2970933

[34] Yeliz Tunga and Kursat Cagiltay. 2023. Looking through the Model's Eye: A Systematic Review of Eye Movement Modeling Example Studies. *Education and Information Technologies* (Jan. 2023). https://doi.org/10.1007/s10639-022-11569-5

[35] Tamara van Gog. 2021. The Signaling (or Cueing) Principle in Multimedia Learning. In *The Cambridge Handbook of Multimedia Learning* (third ed.), Logan Fiorella and Richard E. Mayer (Eds.). Cambridge University Press, Cambridge, 221–230. https://doi.org/10.1017/9781108894333.022

[36] Hongyan Wang, Zhongling Pi, and Weiping Hu. 2019. The Instructor's Gaze Guidance in Video Lectures Improves Learning. *Journal of Computer Assisted Learning* 35, 1 (2019), 42–50. https://doi.org/10.1111/jcal.12309

[37] Jiahui Wang and Pavlo D. Antonenko. 2017. Instructor Presence in Instructional Video: Effects on Visual Attention, Recall, and Perceived Learning. *Computers in Human Behavior* 71 (June 2017), 79–89. https://doi.org/10.1016/j.chb.2017.01.049

[38] Wen-Fang Wang, Chih-Ming Chen, and Chung-Hsin Wu. 2015. Effects of Different Video Lecture Types on Sustained Attention, Emotion, Cognitive Load, and Learning Performance. In *2015 IIAI 4th International Congress on Advanced Applied Informatics*. 385–390. https://doi.org/10.1109/IIAI-AAI.2015.225

Received 19 January 2023