A Methodology for Capturing Joint Visual Attention Using Mobile Eye-trackers
--Manuscript Draft--

Article Type:	Invited Methods Article - Author Produced Video
Manuscript Number:	JoVE60670R2
Full Title:	A Methodology for Capturing Joint Visual Attention Using Mobile Eye-trackers
Section/Category:	JoVE Behavior
Keywords:	Educational Research, Collaborative Learning, Mobile Eye-tracking, Joint Visual Attention, Multimodal Learning Analytics.
Corresponding Author:	Bertrand Schneider Harvard University Cambridge, UNITED STATES
Corresponding Author's Institution:	Harvard University
Corresponding Author E-Mail:	bertrand_schneider@g.harvard.edu
Order of Authors:	Bertrand Schneider
Additional Information:	
Question	**Response**
Please indicate whether this article will be Standard Access or Open Access.	Standard Access (US$1200)

1

TITLE:

A Methodology for Capturing Joint Visual Attention Using Mobile Eye-Trackers

AUTHORS AND AFFILIATIONS:

Bertrand Schneider[1]

[1]Learning, Innovation, and Technology Lab, Graduate School of Education, Harvard University

Corresponding Author:

Bertrand Schneider   (bertrand_schneider@gse.harvard.edu)

KEYWORDS:

educational research, collaborative learning, mobile eye-tracking, joint visual attention, multimodal learning analytics

SHORT ABSTRACT:

Using multimodal sensors is a promising way to understand the role of social interactions in educational settings. This paper describes a methodology for capturing joint visual attention from colocated dyads using mobile eye-trackers.

LONG ABSTRACT:

With the advent of new technological advances, it is possible to study social interactions at a microlevel with unprecedented accuracy. High frequency sensors, such as eye-trackers, electrodermal activity wristbands, EEG bands, and motion sensors provide observations at the millisecond level. This level of precision allows researchers to collect large datasets on social interactions. In this paper, I discuss how multiple eye-trackers can capture a fundamental construct in social interactions, joint visual attention (JVA). JVA has been studied by developmental psychologists to understand how children acquire language, learning scientists to understand how small groups of learners work together, and social scientists to understand interactions in small teams. This paper describes a methodology for capturing JVA in colocated settings using mobile eye-trackers. It presents some empirical results and discusses implications of capturing microobservations to understand social interactions.

INTRODUCTION:

JVA has been extensively studied over the last century, especially by developmental psychologists studying language acquisition. It was quickly established that joint attention is more than just a way to learn words but rather a precursor to children's theories of mind[1]. Thus, it plays a significant role in many social processes, such as communicating with others, collaborating, and developing empathy. Autistic children, for instance, lack the ability to coordinate their visual attention with their caregivers, which is associated with significant social impairments[2]. Humans need joint attention to become functional members of society, to coordinate their actions, and to learn from others. From children acquiring their first words, teenagers learning from schoolteachers, students collaborating on projects, and to groups of adults working toward

45 common goals, joint attention is a fundamental mechanism to establish common ground
46 between individuals[3]. In this paper, I focus on the study of JVA in educational research.
47 Understanding how joint attention unfolds over time is of primary importance for the study of
48 collaborative learning processes. As such, it plays a predominant role in socioconstructivist
49 settings.
50
51 The exact definition of joint attention is still debated[4]. This paper is concerned with a
52 subconstruct of joint attention (JA), namely JVA. JVA happens when two subjects are looking at
53 the same place at the same time. It should be noted that JVA does not provide any information
54 about other important constructs of interest in the study of JA, such as monitoring common,
55 mutual, and shared attention, or more generally, awareness of the cognition of another group
56 member. This paper operationalizes and simplifies JVA by combining the eye-tracking data from
57 two participants and analyzing the frequency in which they align their gazes. For a more
58 comprehensive discussion, the interested reader can learn more about the study of the JA
59 construct in Siposovaet al.[4].
60
61 Over the past decade, technological advances have radically transformed research on JVA. The
62 main paradigm shift was to use multiple eye-trackers to obtain quantitative measures of
63 attentional alignments, as opposed to qualitatively analyzing video recordings in a laboratory or
64 ecological setting. This development has allowed researchers to collect precise, detailed
65 information about dyads' visual coordination. Additionally, eye-trackers are becoming more
66 affordable: until recently, their use was reserved to academic settings or large corporations. It is
67 now possible to purchase inexpensive eye-trackers that generate reliable datasets. Finally, the
68 progressive inclusion of gaze-tracking capabilities into existing devices like high-end laptops and
69 virtual and augmented reality headsets suggests that eye-tracking will soon become ubiquitous.
70
71 Because of the popularization of eye-tracking devices, it is important to understand what they
72 can and cannot tell us about social interactions. The methodology presented in this paper marks
73 a first step in this direction. I address two challenges in capturing JVA from multiple eye-trackers:
74 synchronizing the data on 1) the temporal scale, and 2) on the spatial scale. More specifically,
75 this protocol makes use of fiducial markers placed in real-world environments to inform
76 computer vision algorithms where participants are orienting their gaze. This new kind of
77 methodology paves the way to rigorous analysis of human behavior in small groups.
78
79 This research protocol complies with the guidelines of Harvard University's human research
80 ethics committee.
81
82 **PROTOCOL:**
83
84 **1. Participant Screening**
85
86 1.1. Ensure that participants with normal or corrected-to-normal vision are recruited. Because
87 participants will be asked to wear a mobile eye-tracker, they can wear contact lenses but not
88 regular eyeglasses.

89

2. Preparation for the Experiment

2.1. Eye-tracking devices

2.1.1. Use any mobile eye-tracker capable of capturing eye movement in real world environments.

NOTE: The mobile eye-trackers used here were two Tobii Pro Glasses 2 (see **Table of Materials**). In addition to specialized cameras that can track eye movements, the glasses are also equipped with an HD scene camera and a microphone so that the gaze can be visualized in the context of the user's visual field. These glasses capture gaze data 50 times per second. Other researchers have used ASL Mobile Eye[5], SMI[6], or Pupil-labs[7], all of which provide video streams from the scene camera and eye-tracking coordinates at varying sampling rates (30–120 Hz). The procedure below might vary slightly with other eye-tracking devices.

2.2. Fiducial Markers

2.2.1. The two steps below (i.e., temporal and spatial alignments) require the use of fiducial markers. There are several computer vision libraries that provide researchers with these markers and algorithms to detect them on an image or video feed. The protocol described uses the Chilitag library[8].

2.3. Temporal alignment

2.3.1. Because the eye-tracking data are recorded on two separate units, ensure that the data are properly synchronized (**Figure 1**). Two main methods can be used. This manuscript only covers the first method, because server synchronization works differently with each brand of mobile eye-tracker.

2.3.1.1. Briefly show a fiducial marker on a computer screen to mark the beginning and the end of a session. This is similar to a visual "hand clap" (**Figure 2**).

2.3.1.2. Alternatively, use a server to synchronize the clocks of the two data collection units. This method is slightly more accurate and recommended if a higher temporal accuracy is required.

2.4. Spatial alignment

2.4.1. To find if two participants are looking at the same place at the same time, map their gazes to a common plane. This plane can be a picture of the experimental setting (see the left side of **Figure 3**). Carefully design this image before the experiment.

2.4.2. Size of the fiducial markers: The general size of the fiducial markers depends on the algorithm used to detect them from the eye-tracking video. Surfaces close to the participants can

133 have smaller fiducial markers, while surfaces further away from them need to be larger, so that
134 they look similar from the participants' perspective. Try different sizes beforehand to make sure
135 that they can be detected from the eye-tracking video.

136

137 2.4.3. Number of fiducial markers: To make the process of mapping gaze points into a common
138 plane successful, make sure to have several fiducial markers visible from the participants' point
139 of view at any given time.

140

141 2.4.4. Location of the fiducial markers: Frame relevant areas of interest with strips of fiducial
142 markers (e.g., see the laptop screen on **Figure 3**).

143

144 2.5. Finally, run pilots to test the synchronization procedure and determine the optimal location,
145 size, and number of fiducial markers. Eye-tracking videos can be processed through a computer
146 vision algorithm to see if the fiducial markers are reliably detected.

147

148 **3. Running the experiment**

149

150 3.1. Instructions

151

152 3.1.1. Instruct participants to put on the eye-tracking glasses as they would a normal pair of
153 glasses. Based on the participants' distinct facial features, nose pieces of different heights may
154 need to be used to preserve data quality.

155

156 3.1.2. After turning on the eye-tracker, have the participants clip the recording unit to themselves
157 to allow for natural body movement.

158

159 3.2. Calibration

160

161 3.2.1. Instruct the participants to look at the center of the calibration marker provided by Tobii
162 while the calibration function of the software is enabled. Once the calibration is complete,
163 recording can be started from within the software.

164

165 3.2.2. Instruct participants to not move the mobile eye-trackers after calibration. If they do, the
166 data are likely to be inaccurate and the calibration procedure will need to be performed again.

167

168 3.3. Data monitoring

169

170 3.3.1. Monitor the data collection process during the study and ensure that the eye-tracking data
171 are being collected properly. Most mobile eye-trackers can provide a live stream on a separate
172 device (e.g., a tablet) for this purpose.

173

174 3.4. Data export

175

176 3.4.1. After the recording session is complete, instruct the participant to remove the eye-tracking

177   glasses and the data collection unit. Turn off the unit.

178

179   3.4.2. Extract data using another software, Tobii Pro Lab, by removing the SD card from the data
180   collection unit importing the session data. Tobii Pro Lab can be used to replay the video, create
181   visualizations, and export the eye-tracking data as comma-separated (.csv) or tab-separated (.tsv)
182   files.

183

184   **4. Preprocessing the dual eye-tracking data**

185

186   4.1. Sanity checking eye-tracking data

187

188   4.1.1. Check the eye-tracking data visually after data collection. It is not uncommon for some
189   participants to have missing data. For example, some particular eye physiology can pose
190   problems to eye-tracking algorithms, the glasses might shift during the experiment, the data
191   collection software might crash, etc.

192

193   4.1.2. Use descriptive statistics to check how much data were lost during each session and
194   exclude sessions that have significant amounts of missing or noisy data.

195

196   4.2. Temporal alignment

197

198   4.2.1. Trim the data from each mobile eye-tracker to only include interactions between the
199   participants. This can be achieved by using the method described above (i.e., presenting two
200   special fiducial markers to participants at the start and the end of the session). These fiducial
201   markers can then be detected from the eye-tracking video to trim the datasets.

202

203   4.3. Spatial alignment

204

205   NOTE: To detect whether two participants are looking at the same place at the same time, it is
206   necessary to remap the participants' gaze onto a common plane (i.e., an image of the
207   experimental setting). A computational method for achieving this goal is a homography (i.e., a
208   perspective transformation of a plane). From a technical perspective, two images of the same
209   planar surface in space are related by a homography matrix. Based on a common set of points,
210   this matrix can be used to infer the location of additional points between two planes. In **Figure
211   3**, for example, if a computer vision algorithm knows where the fiducial markers are on the
212   handout, it can remap the gaze of the participant onto the common plane on the left side. The
213   white lines connect the two sets of points shared by the video feed of each participant and the
214   scene, which are then used for building the homography to remap the green and blue dots on
215   the left side.

216

217   4.3.1. Use the Python version of OpenCV, e.g., to compute the homography matrix from the
218   fiducial markers and then to remap the eye-tracking data to the scene of the experimental setting
219   (or any other suitable library in your language of choice). OpenCV provides two useful functions:
220   **findHomography()** to get the homography matrix, and **perspectiveTransform()** to transform the

221   point from one perspective to the other.

222

223   4.3.2. To use **findHomography()**, run with two arguments: the X,Y coordinates of the source
224   points (i.e., the fiducial markers detected from the participants' scene video, shown on the right
225   in **Figure 3**) and the corresponding destination points (i.e., the same fiducials markers detected
226   on the scene image, shown on the left in **Figure 3**).

227

228   4.3.3. Feed the resulting homography matrix into the **perspectiveTransform()** function, along
229   with a new point that needs to be mapped from the source image to the destination image (e.g.,
230   the eye-tracking data shown as a blue/green dot on the right side of **Figure 3**). The
231   **perspectiveTransform** function returns the new coordinate of the same point on the scene image
232   (i.e., the blue/green dots shown on the left side of **Figure 3**).

233

234   NOTE: For more information, the OpenCV official documentation provides sample code and
235   examples   to   implement   the   homography:
236   docs.opencv.org/master/d1/de0/tutorial_py_feature_homography.html.

237

238   4.4. Sanity checking the homography

239

240   4.4.1. Complete section 4.3 for the entire session, and perform a homography on each frame of
241   the mobile eye-tracking video to check the quality of the homography. While there are no
242   automated ways to estimate the accuracy of the resulting eye-tracking data, videos like the one
243   shown in **Figure 4** should be used to manually sanity check each session.

244

245   4.4.2. If the quality is lower than expected, consider additional parameters to improve the results
246   of the homography:

247

248   4.4.2.1. Number of fiducial markers detected: Only perform the homography if enough fiducial
249   markers can be detected from the video stream. This number can be determined by examining
250   the video produced above.

251

252   4.4.2.2. Location of the fiducial markers: If different markers are at different depths and
253   orientations, the quality of the homography usually increases when the markers closest to the
254   gaze coordinates are selected, given that there are enough markers to build a robust
255   homography.

256

257   4.4.2.3. Orientation of the fiducial markers: Combining fiducial markers that have different
258   orientations (e.g., horizontal and vertical) will produce inaccurate homographies. It is
259   recommended to first detect which plane or areas of interests (AOIs) the participant is looking at
260   (e.g., the computer screen, the cheat sheet, the table, see **Figure 3**) and then use the fiducial
261   markers on this plane for the homography.

262

263   4.4.2.4. Quality of the video stream: Sudden head movements can blur video frames and make
264   the data unusable, because fiducial markers cannot be reliably detected (**Figure 4**). The

265 methodology of this paper is not appropriate for experiments that involve a lot of sudden head
266 movements.
267
268 **5. Analyzing the dual eye-tracking data**
269
270 5.1. Missing data
271
272 5.1.1. In order to make sure that the data were properly remapped onto the reference image,
273 produce visualization graphs (e.g., **Figure 5**, **Figure 6**) and descriptive statistics to check how
274 much data are missing.
275
276 5.2. Cross-recurrence graphs
277
278 5.2.1. Use cross-recurrence graphs[9] to represent visual synchronization between two
279 participants (**Figure 6**), where the X-axis represents time for the first participant, and the Y-axis
280 represents time for the second participant. Black squares indicate that participants are looking
281 at the same area, a black diagonal line describes two subjects looking at the same thing at exactly
282 the same time, and black squares off the diagonal line describes when two subjects looking at
283 the same thing with a time lag. Finally, differentiating between missing data (white square) and
284 existing data with no JVA (gray squares) helps identify problematic sessions. This provides
285 researchers with a visual sanity check.
286
287 5.3. Computing JVA
288
289 5.3.1. After filtering for missing data, compute a metric for JVA by counting the number of times
290 that participants' gazes are in the same radius in the scene (defined below) in a -2/+2 s time
291 window. Divide this number by the number of valid data points that can be used to compute JVA.
292 The result of the division represents the percentage of time that two subjects were jointly looking
293 at the same place. This last step is necessary to avoid inflating the scores of groups with more
294 data after the homography.
295
296 NOTE: Two parameters need to be set before JVA can be computed, the minimal distance
297 between two gaze points, and the time window between them (**Figure 7**): 1) Time window: An
298 early foundational study[10] used a single eye-tracker to measure JVA between a listener and a
299 speaker. The researchers asked a first set of participants ("speakers") to talk about a television
300 show whose characters were displayed in front of them. A second set of participants ("listeners")
301 then watched the same show while listening to audio recording of the speakers. The eye
302 movements of the speakers and listeners were compared, and it was found that a listener's eye
303 movements closely matched a speaker's eye movement with a delay of 2 s. In subsequent work[11],
304 researchers analyzed live dialogues and found that a delay of 3 s best captured moments of JVA.
305 Because each task is unique and might exhibit different time lags, it is also suggested to explore
306 how different time lags affect the results of a given experiment (see Schneider et al.[12], for
307 example). Overall, it is common to look for JVA in a $\pm$ 2/3 s time window depending on the
308 experimental task and then explore how different time lags might change the results. 2) Distance

309 between gazes: there is no empirically defined distance between two gazes for them to count as
310 JVA. This distance is dependent on the research questions defined by the researchers. The
311 research questions should inform the size of the targets of interest. In the example seen in **Figure**
312 **7**, a radius of 100 pixels on the scene image (blue/green circles) was chosen for the analysis
313 because it is sufficient to capture when participants are looking at the robot in the maze, as well
314 as at similar user interface elements on the computer screen, which are the two main areas of
315 interest for this experimental task.
316
317 **REPRESENTATIVE RESULTS:**
318 The methodology presented above was used to study students who were following a vocational
319 training program in logistics (n = 54)[12]. In this experiment, pairs of students interacted with a
320 Tangible User Interface (TUI) that simulated a small-scale warehouse. The fiducial markers placed
321 on the TUI allowed the research team to remap students' gazes onto a common plane and
322 compute levels of JVA. Findings indicated that groups who had higher levels of JVA tended to do
323 better at the task given to them, learned more, and had a better quality of collaboration[13] (**Figure**
324 **8**, left side). Dual eye-tracking datasets also allowed us to capture particular group dynamics like
325 the free-rider effect. We estimated this effect by identifying who was likely to have initiated each
326 moment of JVA (i.e., whose gaze was there first) and who responded to it (i.e., whose gaze was
327 there second). We found a significant correlation between learning gains and the students'
328 tendency to equally share the responsibility of initiating and responding to offers of JVA. In other
329 words, groups in which the same person always initiated moments of JVA were less likely to learn
330 (**Figure 8**, right side) and groups where this responsibility was equally shared were more likely to
331 learn. This finding shows that we can go beyond merely quantifying JV, and actually identify group
332 dynamics and productivity using dual eye-tracking data.
333
334 **FIGURE AND TABLE LEGENDS:**
335 **Figure 1. Each participant generates two video feeds with the X,Y coordinates of their gaze on**
336 **each video frame:** This methodology addresses synchronizing the data temporally and spatially
337 between the participants.
338
339 **Figure 2. A methodology for synchronizing the two datasets:** Briefly showing unique fiducial
340 marker on a computer screen to tag the start and the end of the activity.
341
342 **Figure 3. Using fiducial markers disseminated in the environment to remap participants' gazes**
343 **onto a common plan (left side):** White lines indicate fiducial markers that have been detected in
344 both images.
345
346 **Figure 4. Examples of poor data quality:** Left: A blurred frame from the eye-tracking video caused
347 by a sudden head movement. Fiducial markers could not be detected in this image. Right: A failed
348 homography where the fiducial marker data were not properly synchronized with the video feed.
349
350 **Figure 5. Heatmaps:** Left: A heatmap of the eye-tracking data remapped onto the experimental
351 scene. This visualization was used as a sanity check for the homography. Right: A group that had
352 too much missing data and had to be discarded.

353

Figure 6. Cross recurrence graph generated from three dyads to visualize JVA: P1 represents time for the first participant, P2 represents time for the second participant. Black squares show JVA; gray squares show moments where participants are looking at different places; white squares show missing data. Square along the main diagonal indicates moments where participants looked at the same place at the same time. This visualization was used as a sanity check for measures of JVA from the combined eye-tracking data.

Figure 7. A video frame where JVA was detected between two participants (red dots): Richardson et al.[11] recommend looking at a time window of +/-2 s. when computing JVA. Additionally, researchers need to define the minimal distance between two gaze points to count as JVA. A radius of 100 pixel was chosen on the middle image above.

Figure 8. Examples of results: Data from Schneider et al.[12] where the percentage of time looking at the same place at the same time was correlated with participants' quality of collaboration: $r(24) = 0.460$, $P = 0.018$ (left side) and imbalances in initiating/responding to offers of JVA was correlated with their learning gains: $r(24) = -0.47$, $P = 0.02$ (right side).

DISCUSSION:

The methodology described in this paper provides a rigorous way to capture JVA in colocated dyads. With the emergence of affordable sensing technology and improved computer vision algorithms, it is now possible to study collaborative interactions with an accuracy that was previously unavailable. This methodology leverages fiducial markers disseminated in the environment and uses homographies as a way to remap participants' gazes onto a common plane. This allows researchers to rigorously study JVA in colocated groups.

This method includes multiple sanity checks that need to be performed at various point of the experiment. Because this is a complex procedure, researchers need to make sure that the resulting datasets are complete and valid. Finally, it is recommended to conduct pilot studies before the actual experiment, and to reconstruct participants' interactions though a video after data collection is completed (**Figure 3**, **Figure 4**, **Figure 5**, **Figure 6**).

There are several limitations associated with this method:

Number of participants. While this methodology works well for two participants, analysis become more complicated with larger groups. Fiducial markers can still be used to remap gazes onto a ground truth but knowing how to identify JVA becomes a more nuanced process. Should JVA be defined as the times when everyone is looking at the same place at the same time, or when two participants are gazing at the same place? Additionally, visualizations like the cross-recurrence graph become impractical with more than 2–3 people.

Settings. The method described in this paper is appropriate for small, controlled settings (e.g., laboratory studies). Open-ended settings, such as outdoors or large spaces, are usually too complicated to instrument with fiducial markers and thus can limit the usefulness of the eye-

397 tracking data. Additionally, the fiducial markers can be distracting and clutter the environment.
398 In the future, better computer vision algorithms will be able to automatically extract common
399 features between two perspectives. There are already algorithms that exist for this purpose, but
400 we found that the level of accuracy was not yet acceptable for the type of experiment described
401 above.
402
403 **AOIs.** Related to the point above, computing homography and the cross-recurrence graph work
404 well with a stable number of areas of interest, but corrections have to be made when comparing
405 different tasks with different numbers of areas of interest.
406
407 **Use of equipment.** Mobile eye-trackers can be obtrusive, affecting participants' behavior or
408 failing to work with particular eye physiology.
409
410 In conclusion, the methodology described in this paper is a promising way to study colocated
411 interactions. It allows researchers to capture a precise metric for JVA, which is a critical construct
412 in the social sciences[1]. Additionally, it is possible to detect more fine-grained indicators of
413 collaborative learning through this methodology[12] compared to traditional qualitative analyses.
414 In short, it is a more efficient and accurate way to study social interactions.
415
416 Potential application of this method includes designing interventions to support collaboration
417 through real-time eye-tracking data. Some pioneering work has produced shared gaze
418 visualizations using remote eye-trackers, which has been shown to benefit collaborative learning
419 from a distance[14]. Dyads who could see the gaze of their partner in real time exhibited more JVA,
420 collaborated better and achieved higher learning gains compared to a control group. Future work
421 will examine whether this kind of intervention can support collaborative processes in colocated
422 settings (e.g., through virtual or augmented reality headsets).
423

433 **REFERENCES:**
434 1. Tomasello, M. Joint attention as social cognition. In: Moore, C., Dunham, P. J. (eds.) Joint
435   attention: Its origins and role in development. pp. 103–130. Lawrence Erlbaum Associates,
436   Inc, Hillsdale, NJ, England (1995).
437 2. Mundy, P., Sigman, M., Kasari, C. A longitudinal study of joint attention and language
438   development in autistic children. *Journal of Autism and Developmental Disorders*. **20**, 115–
439   128 (1990).
440 3. Clark, H. H., Brennan, S. E. Grounding in communication. In: Resnick, L. B., Levine, J. M.,

441   Teasley, S. D. (eds.) Perspectives on socially shared cognition. pp. 127–149. American
442   Psychological Association, Washington, DC, US (1991).

4. Siposova, B., Carpenter, M. A new look at joint attention and common knowledge. *Cognition*.
   189, 260–274 (2019).

5. Gergle, D., Clark, A. T. 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, New York, NY, USA (2011).

6. Renner, P., Pfeiffer, T., Wachsmuth, I. Spatial References with Gaze and Pointing in Shared
   Space of Humans and Robots. In: Freksa, C., Nebel, B., Hegarty, M., Barkowsky, T. (eds.)
   Spatial Cognition IX. pp. 121–136. Springer International Publishing (2014).

7. Shvarts, A. Y. Automatic detection of gaze convergence in multimodal collaboration: a dual
   eye-tracking technology. *The Russian Journal of Cognitive Science*. **5**, 4 (2018).

8. Bonnard, Q. et al. Chilitags: Robust Fiducial Markers for Augmented Reality [software].
   https://github.com/chili-epfl/qml-chilitags (2013).

9. Jermann, P., Mullins, D., Nüssli, M.-A., Dillenbourg, P. Collaborative Gaze Footprints:
   Correlates of Interaction Quality. Connecting Computer-Supported Collaborative Learning
   to Policy and Practice: CSCL2011 Conference Proceedings., Volume I - Long Papers, 184–191
   (2011).

10. Richardson, D. C., Dale, R. Looking To Unde
rstand: The Coupling Between Speakers' and Listeners' Eye Movements and Its Relationship to
   Discourse Comprehension. *Trends in Cognitive Sciences.* **29**, 1045–1060 (2005).

11. Richardson, D. C., Dale, R., Kirkham, N. Z. The Art of Conversation Is Coordination Common
   Ground and the Coupling of Eye Movements During Dialogue. *Psychological Science*. **18**,
   407–413 (2007).

12. Schneider, B. et al. Using Mobile Eye-Trackers to Unpack the Perceptual Benefits of a Tangible
   User Interface for Collaborative Learning. *ACM Transactions on Computer-Human
   Interaction*. **23**, 1–23 (2016).

13. Meier, A., Spada, H., Rummel, N. A rating scheme for assessing the quality of computer-
   supported collaboration processes. Int. J. Comput.-Support. Collab. Learn. 2, 63–86 (2007).

14. Schneider, B., Pea, R. Real-time mutual gaze perception enhances collaborative learning and
   collaboration quality. *Journal of Computer-Supported Collaborative Learning*. **8**, 375–397
   (2013).

Participant 1
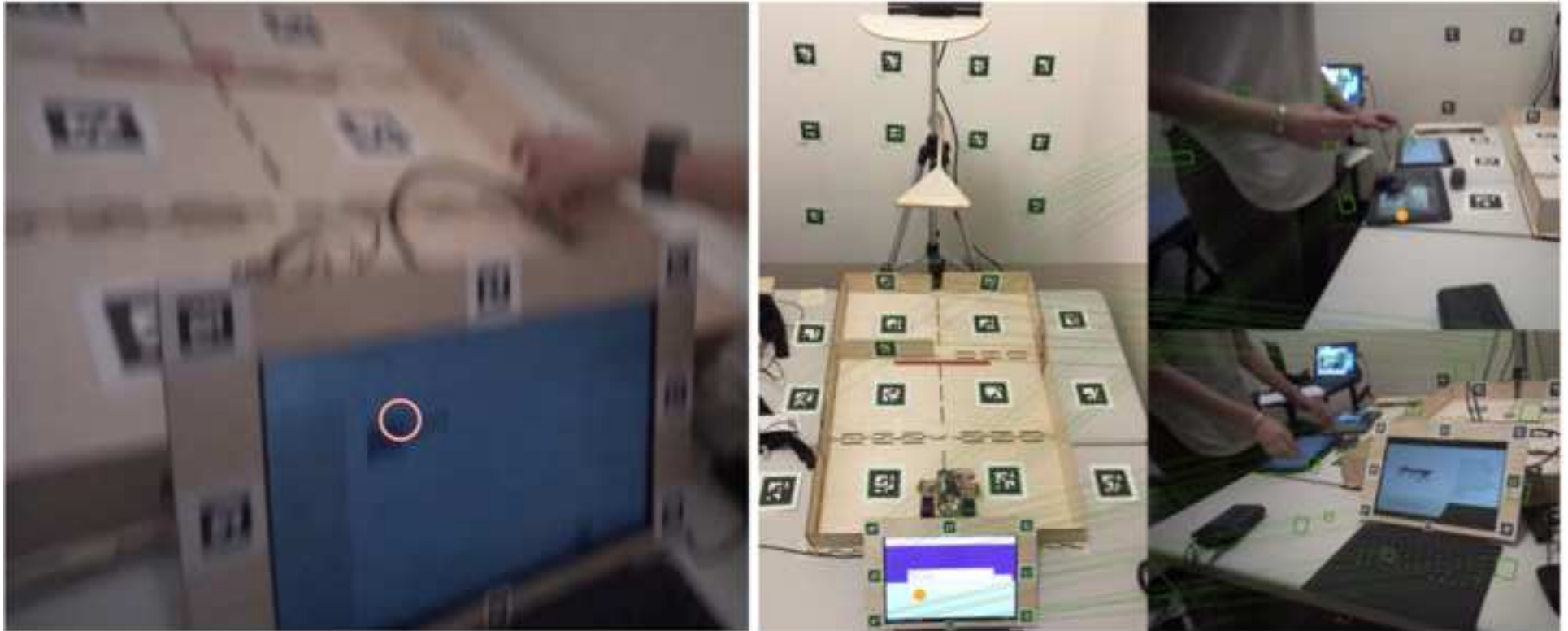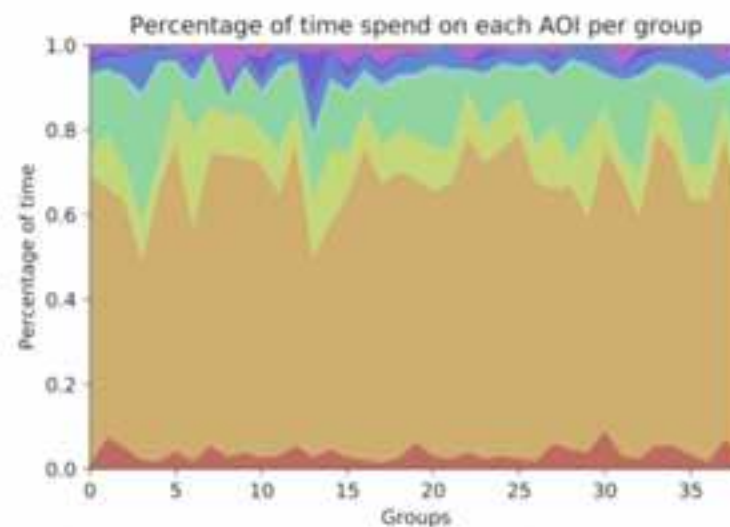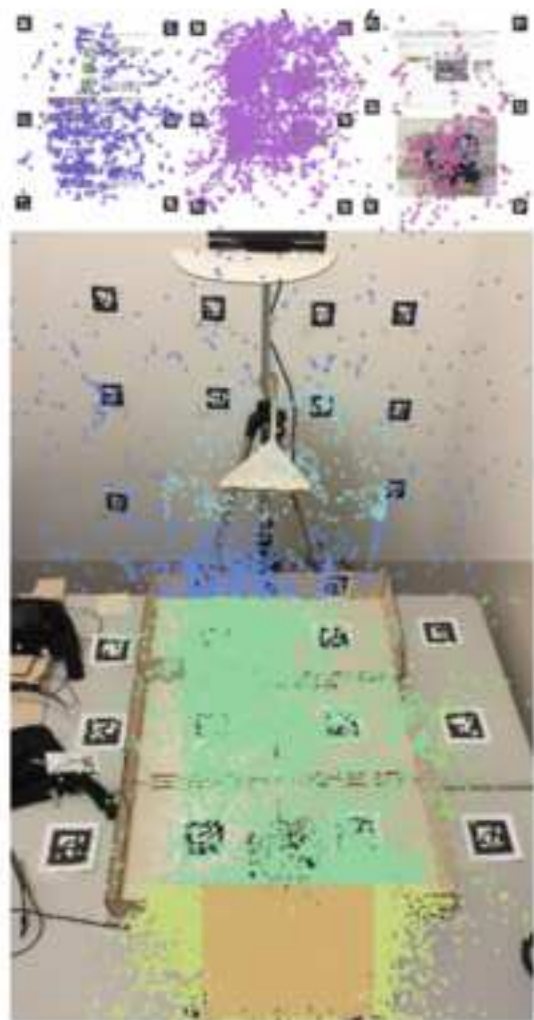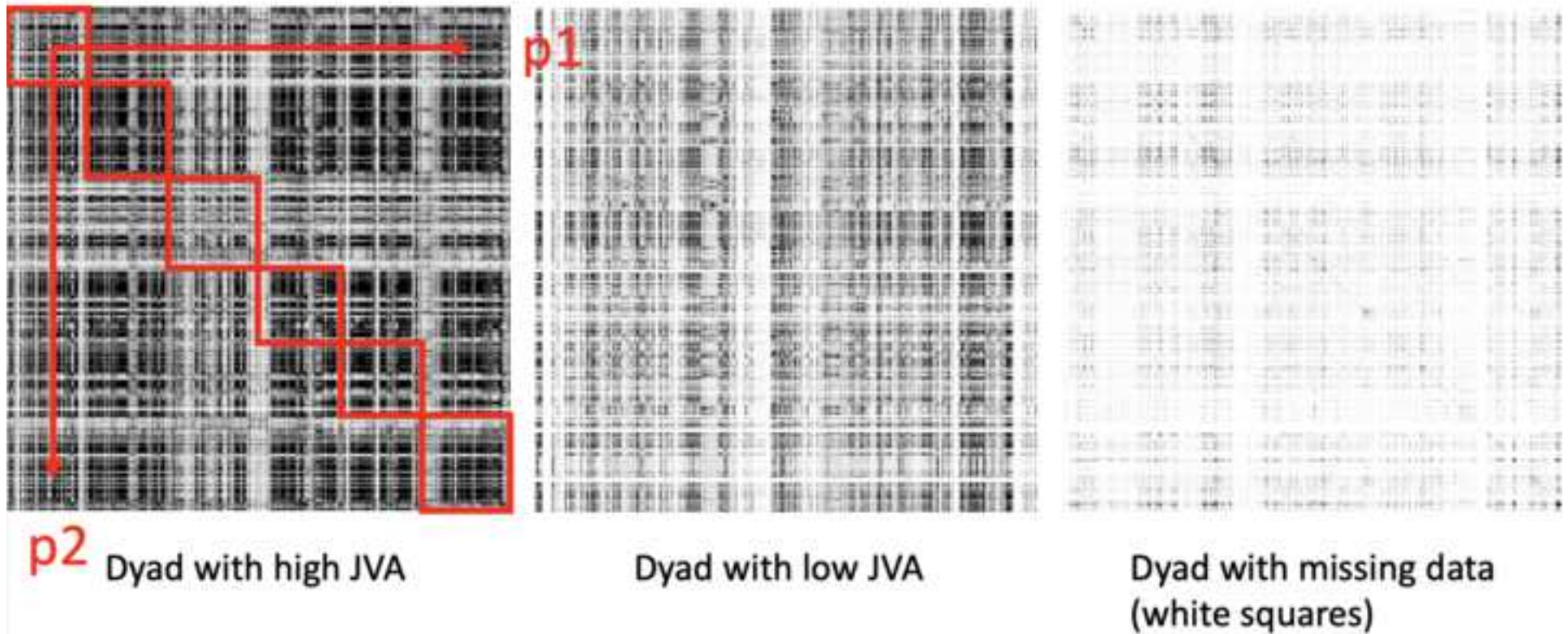
25Hz

Participant 2

25Hz
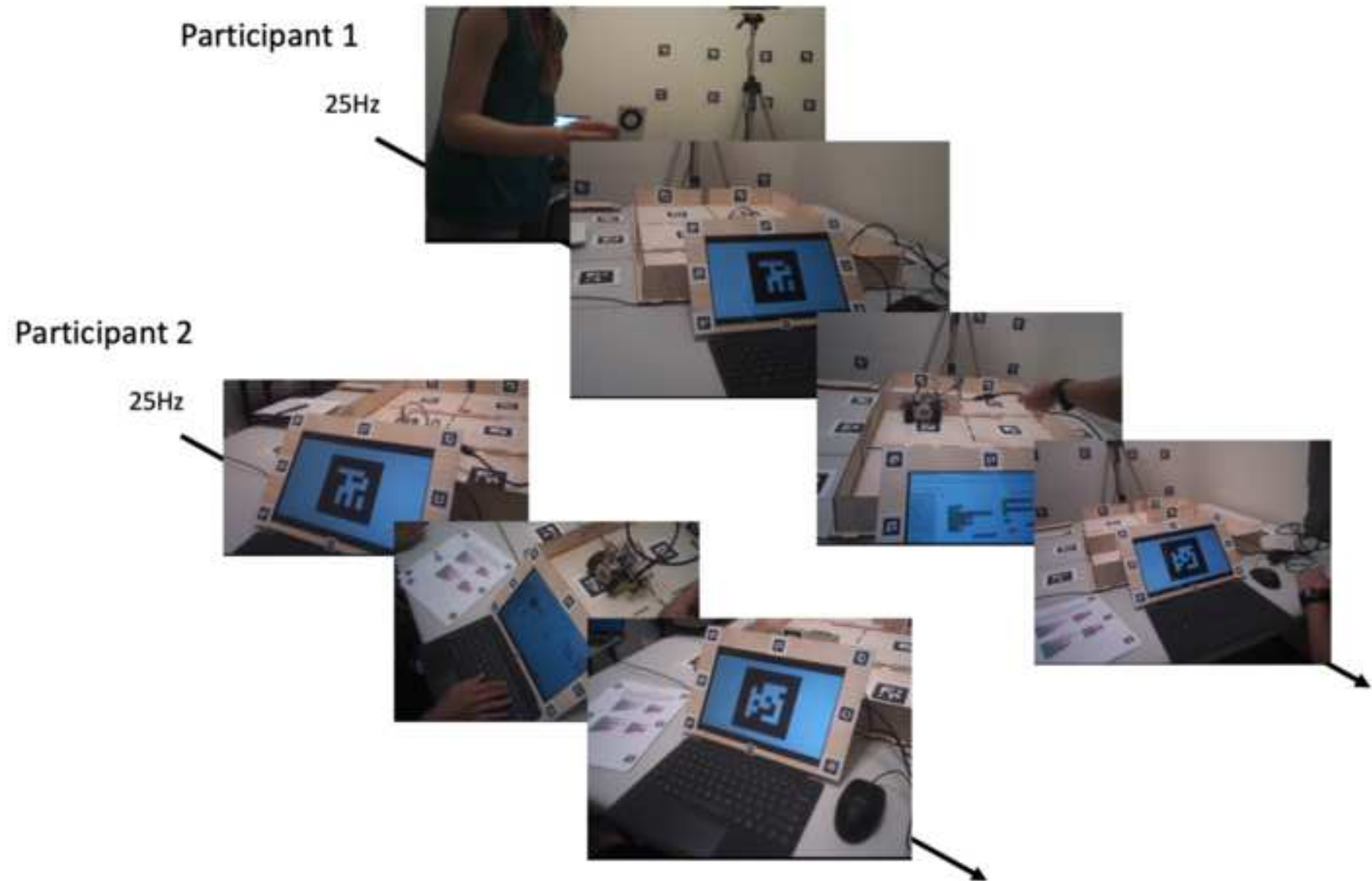
Participant 1

Participant 2

Sanity check #1: Heatmaps

Sanity check #2: Cross Recurrence Graphs

Dyad with high JVA

Dyad with low JVA

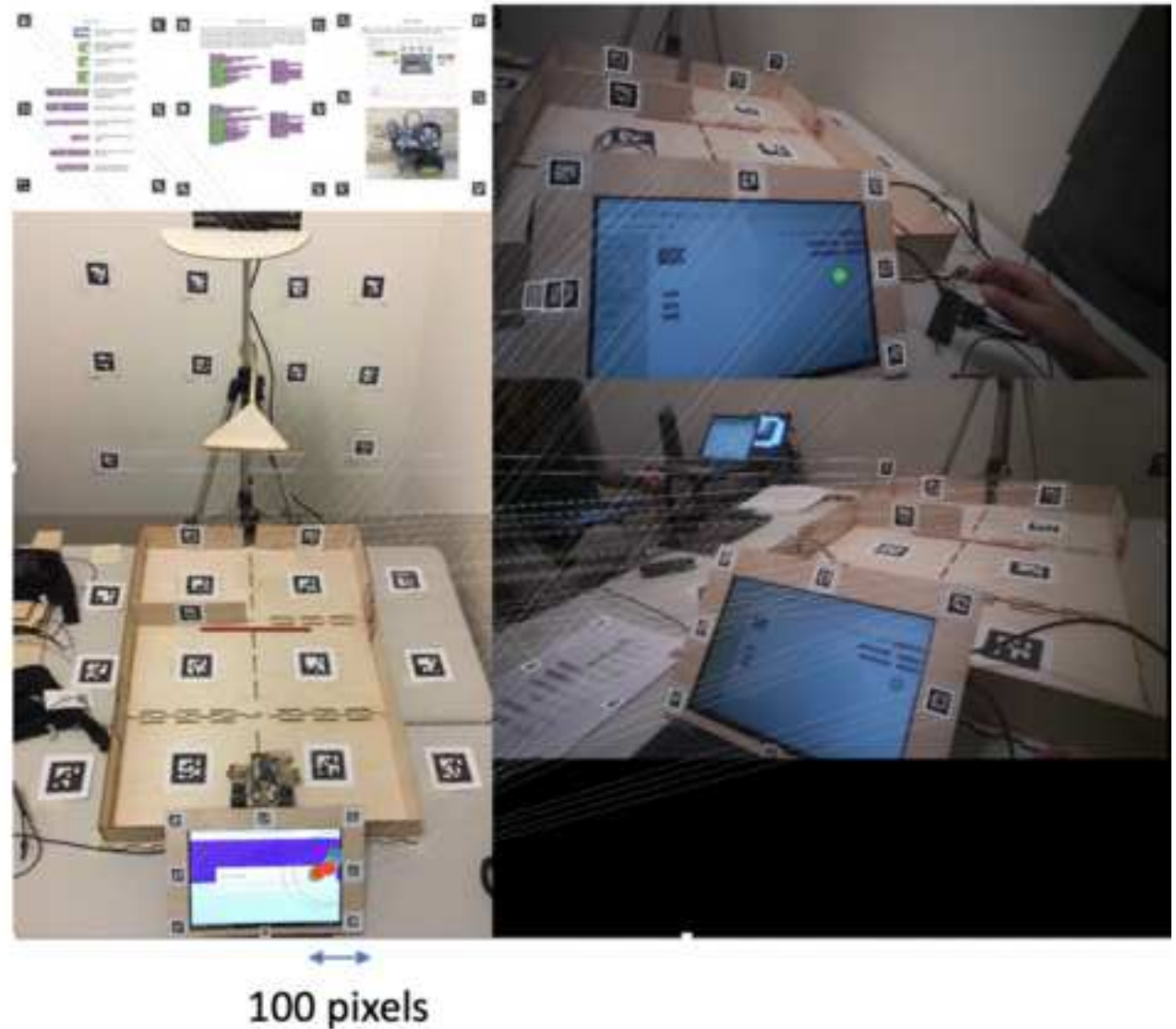Dyad with missing data (white squares)

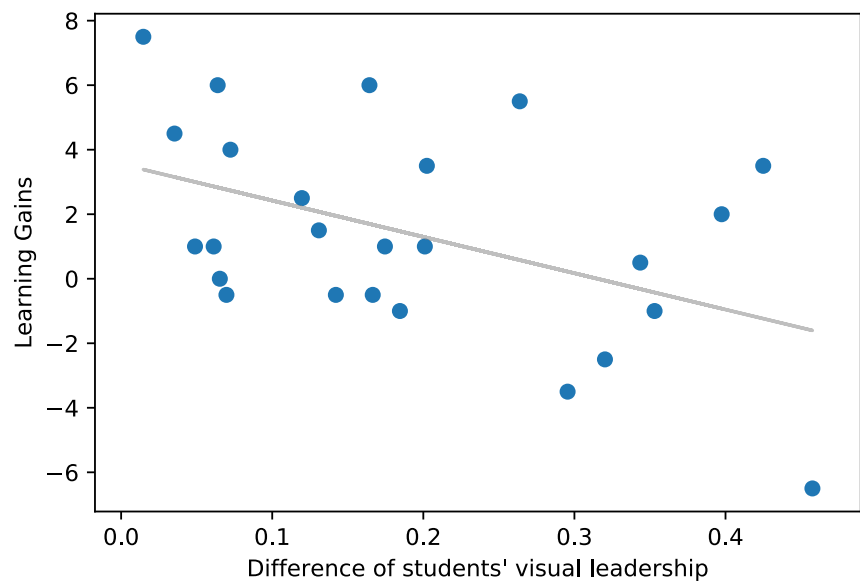# Two parameters for Computing Joint Visual Attention: time and distance
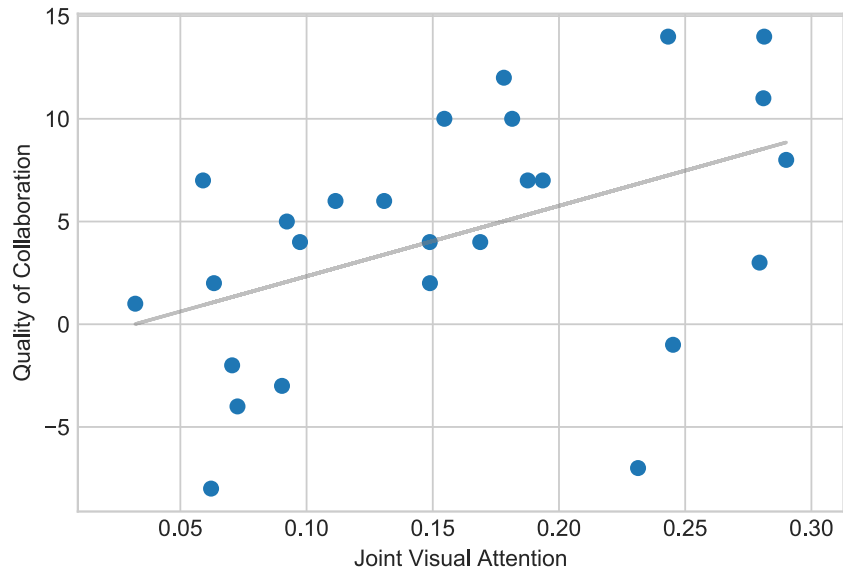
Time Window
-2sec / + 2sec
(Richardson and
Dale, 2005)

Distance (radius)

To be determined
based on the task
and setting

100 pixels

Table of Materials

Name of Material/ Equipment	Company	Catalog Number
Tobii Glasses 2	Tobii	N/A
Fiducial markers	Chili lab – EPFL, Switzerland	N/A

Comments/Description

https://www.tobiipro.com/product-listing/tobii-pro-glasses-2/

https://github.com/chili-epfl/chilitags

Professor Bertrand Schneider
Longfellow Hall, office 333
School of Education
Harvard University
13 Appian Way Cambridge, MA 02138
bertrand_schneider@gse.harvard.edu

August 12th, 2019

Dear JoVE editors,

I wish to submit a revised version of the manuscript: "A Methodology for Capturing Joint Visual Attention Using Mobile Eye-trackers" for consideration by the Journal of Visualized Experiments. I have made the following changes to the manuscript and video:
1. Section 0:46-3:00 of the video now appears in the manuscript.
2. The written protocol is in the imperative voice (when applicable).
3. Protocol step 4c: There is more detail on how the python functions are implemented (which includes additional references and/or links to example code).
4. Figure 8: the two panels are combined into one image file and both are described in the caption.
5. Showing Tobi logos in the video: as agreed by email, we haven't made any changes to citing Tobii in the video.
6. 3:01, 3:50, 6:38, 7:13, 8:23 - the white background has been extended to eliminate the pillarboxing. The background is now in a consistent color.
7. The title card is placed at the very beginning of the video as well as the end.
8. Tere are chapter title cards before each section of the video.
9. The article ID number (60670) was included in the video file name and uploaded to dropbox.

Please don't hesitate to let me know if other changes are necessary.

Sincerely,

Bertrand Schneider, PhD
Assistant Professor
Graduate School of Education
Harvard University

Supplemental File (Figures, Permissions, etc.)

Click here to access/download;Supplemental File (Figures, Permissions, etc.);ijCSCL ⬇
- RightsLink Printable License.pdf

**SPRINGER NATURE LICENSE
TERMS AND CONDITIONS**

Oct 04, 2019

This Agreement between Bertrand Schneider ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	4682160372499
License date	Oct 04, 2019
Licensed Content Publisher	Springer Nature
Licensed Content Publication	International Journal of Computer-Supported Collaborative Learning
Licensed Content Title	Leveraging mobile eye-trackers to capture joint visual attention in co-located collaborative learning groups
Licensed Content Author	Bertrand Schneider, Kshitij Sharma, Sebastien Cuendet et al
Licensed Content Date	Jan 1, 2018
Licensed Content Volume	13
Licensed Content Issue	3
Type of Use	Journal/Magazine
Requestor type	academic/university or research institute
Is this reuse sponsored by or associated with a pharmaceutical or a medical products company?	no
Format	electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Will you be translating?	no
Circulation/distribution	1 - 29
Author of this Springer Nature content	yes
Title of new article	A Methodology for Capturing Joint Visual Attention Using Mobile Eye-trackers

Terms and Conditions

Springer Nature Customer Service Centre GmbH
Terms and Conditions

This agreement sets out the terms and conditions of the licence (the **Licence**) between you and **Springer Nature Customer Service Centre GmbH** (the **Licensor**). By clicking 'accept' and completing the transaction for the material (**Licensed Material**), you also confirm your acceptance of these terms and conditions.

1. Grant of License

1. 1. The Licensor grants you a personal, non-exclusive, non-transferable, world-wide licence to reproduce the Licensed Material for the purpose specified in your order only. Licences are granted for the specific use requested in the order and for no other use, subject to the conditions below.

1. 2. The Licensor warrants that it has, to the best of its knowledge, the rights to license reuse of the Licensed Material. However, you should ensure that the material you are requesting is original to the Licensor and does not carry the copyright of another entity (as credited in the published version).

1. 3. If the credit line on any part of the material you have requested indicates that it was reprinted or adapted with permission from another source, then you should also seek permission from that source to reuse the material.

2. Scope of Licence

2. 1. You may only use the Licensed Content in the manner and to the extent permitted by these Ts&Cs and any applicable laws.

2. 2. A separate licence may be required for any additional use of the Licensed Material, e.g. where a licence has been purchased for print only use, separate permission must be obtained for electronic re-use. Similarly, a licence is only valid in the language selected and does not apply for editions in other languages unless additional translation rights have been granted separately in the licence. Any content owned by third parties are expressly excluded from the licence.

2. 3. Similarly, rights for additional components such as custom editions and derivatives require additional permission and may be subject to an additional fee. Please apply to Journalpermissions@springernature.com/bookpermissions@springernature.com for these rights.

2. 4. Where permission has been granted **free of charge** for material in print, permission may also be granted for any electronic version of that work, provided that the material is incidental to your work as a whole and that the electronic version is essentially equivalent to, or substitutes for, the print version.

2. 5. An alternative scope of licence may apply to signatories of the STM Permissions Guidelines, as amended from time to time.

3. Duration of Licence

3. 1. A licence for is valid from the date of purchase ('Licence Date') at the end of the relevant period in the below table:

Scope of Licence	Duration of Licence
Post on a website	12 months
Presentations	12 months
Books and journals	Lifetime of the edition in the language purchased

4. Acknowledgement

4. 1. The Licensor's permission must be acknowledged next to the Licenced Material in print. In electronic form, this acknowledgement must be visible at the same time as the figures/tables/illustrations or abstract, and must be hyperlinked to the journal/book's homepage. Our required acknowledgement format is in the Appendix below.

5. Restrictions on use

5. 1. Use of the Licensed Material may be permitted for incidental promotional use and minor editing privileges e.g. minor adaptations of single figures, changes of format, colour and/or style where the adaptation is credited as set out in Appendix 1 below. Any other changes including but not limited to, cropping, adapting, omitting material that affect the meaning, intention or moral rights of the author are strictly prohibited.

5. 2. You must not use any Licensed Material as part of any design or trademark.

5. 3. Licensed Material may be used in Open Access Publications (OAP) before publication by Springer Nature, but any Licensed Material must be removed from OAP sites prior to final publication.

6. Ownership of Rights

6. 1. Licensed Material remains the property of either Licensor or the relevant third party and any rights not explicitly granted herein are expressly reserved.

7. Warranty

IN NO EVENT SHALL LICENSOR BE LIABLE TO YOU OR ANY OTHER PARTY OR ANY OTHER PERSON OR FOR ANY SPECIAL, CONSEQUENTIAL, INCIDENTAL OR INDIRECT DAMAGES, HOWEVER CAUSED, ARISING OUT OF OR IN CONNECTION WITH THE DOWNLOADING, VIEWING OR USE OF THE MATERIALS REGARDLESS OF THE FORM OF ACTION, WHETHER FOR BREACH OF CONTRACT, BREACH OF WARRANTY, TORT, NEGLIGENCE, INFRINGEMENT OR OTHERWISE (INCLUDING, WITHOUT LIMITATION, DAMAGES BASED ON LOSS OF PROFITS, DATA, FILES, USE, BUSINESS OPPORTUNITY OR CLAIMS OF THIRD PARTIES), AND
WHETHER OR NOT THE PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. THIS LIMITATION SHALL APPLY NOTWITHSTANDING ANY FAILURE OF ESSENTIAL PURPOSE OF ANY LIMITED REMEDY PROVIDED HEREIN.

8. Limitations

8. 1. _BOOKS ONLY:_ Where **'reuse in a dissertation/thesis'** has been selected the following terms apply: Print rights of the final author's accepted manuscript (for clarity, NOT the published version) for up to 100 copies, electronic rights for use only on a personal website or institutional repository as defined by the Sherpa guideline (www.sherpa.ac.uk/romeo/).

9. Termination and Cancellation

9. 1. Licences will expire after the period shown in Clause 3 (above).

9. 2. Licensee reserves the right to terminate the Licence in the event that payment is not received in full or if there has been a breach of this agreement by you.

Appendix 1 — Acknowledgements:

For Journal Content:
Reprinted by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication)

For **Advance Online Publication papers:**
Reprinted by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**]

[**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication), advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM].)

For Adaptations/Translations:
Adapted/Translated by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication)

Note: For any republication from the British Journal of Cancer, the following credit line style applies:

Reprinted/adapted/translated by permission from [**the Licensor**]: on behalf of Cancer Research UK: : [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication)

For **Advance Online Publication** papers:
Reprinted by permission from The [**the Licensor**]: on behalf of Cancer Research UK: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication), advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM])

For Book content:
Reprinted/adapted by permission from [**the Licensor**]: [**Book Publisher** (e.g. Palgrave Macmillan, Springer etc) [**Book Title**] by [**Book author**(s)] [**COPYRIGHT**] (year of publication)

Other Conditions:

Version 1.2

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.