Behav Res (2015) 47:848–859
DOI 10.3758/s13428-014-0507-6

Qualitative tests of remote eyetracker recovery and performance during head rotation

Roy S. Hessels · Tim H. W. Cornelissen · Chantal Kemner · Ignace T. C. Hooge

Published online: 18 July 2014

Abstract What are the decision criteria for choosing an eyetracker? Often the choice is based on specifications by the manufacturer of the validity (accuracy) and reliability (precision) of measurements that can be achieved using a particular eyetracker. These specifications are mostly achieved under optimal conditions—for example, by using an artificial eye or trained participants fixed in a chinrest. Research, however, does not always take place in optimal conditions: For instance, when investigating eye movements in infants, school children, and patient groups with disorders such as attention-deficit hyperactivity disorder, it is practically impossible to restrict movement. We modeled movements often seen in infant research in two behaviors: (1) looking away from and back to the screen, to investigate eyetracker recovery, and (2) head orientations, to investigate eyetracker performance with nonoptimal orientations of the eyes. We investigated how eight eyetracking setups by three manufacturers (SMI, Tobii, and LC Technologies) coped with these modeled behaviors in adults. We report that the tested SMI eyetrackers dropped in sampling frequency when the eyes were not visible to the eyetracker, whereas the other systems did not, and discuss the potential consequences thereof. Furthermore, we report that the tested eyetrackers varied in their rates of data loss and systematic offsets during shifted head orientations. We conclude that (prospective) eye-movement researchers who cannot restrict movement or nonoptimal head orientations in their participants might benefit from testing their eyetracker in nonoptimal conditions. Additionally, researchers should be aware of the data loss and inaccuracies that might result from nonoptimal head orientations.

Keywords Eyetracking · Head movement · Head orientation · Developmental studies · Data quality

Remote video-based eyetrackers are growing in popularity among various research disciplines (Holmqvist et al., 2011), particularly because they are easy to set up and use. When choosing which remote eyetracker to use, researchers are faced with a plethora of options, all with slightly different technical specifications. Manufacturers specify how accurate their eyetracker is (*spatial accuracy*, the average offset between the point on screen the participant looks at and what the eyetracker reports), how reliable a measurement is (*spatial precision*, the sample to sample difference while the eye remains still), and in what range of distances to the eyetracker tracking of the eyes is possible (*headbox dimensions*). In addition, manufacturers constantly improve their eyetrackers and aim for the best specifications possible. This alone leaves an individual researcher with choices of which the consequences might be difficult to grasp.

The specifications presented by the eyetracker manufacturers are, furthermore, often achieved under optimal conditions. Optimal conditions are, for instance, a fixed amount of light in the room, restricting a human participant from moving, or using an artificial eye instead of a human participant.

Electronic supplementary material The online version of this article (doi:10.3758/s13428-014-0507-6) contains supplementary material, which is available to authorized users.

R. S. Hessels · T. H. W. Cornelissen · C. Kemner · I. T. C. Hooge
Department of Experimental Psychology, Helmholtz Institute, Utrecht University, Utrecht, The Netherlands

R. S. Hessels · T. H. W. Cornelissen · C. Kemner
Department of Developmental Psychology, Utrecht University, Utrecht, The Netherlands

C. Kemner
Brain Center Rudolf Magnus, University Medical Centre, Utrecht, The Netherlands

R. S. Hessels (✉)
Heidelberglaan 1, 3584 CS Utrecht, The Netherlands
e-mail: royhessels@gmail.com

Research, in contrast, does not take place in optimal conditions. In a recent attempt to find the most suitable eyetracker for a prospective infant study, we reached the conclusion that manufacturer specifications were not informative enough. We knew beforehand that our infant participants would not be measured in the manufacturers' optimal conditions. We realized that the problem of choosing between eyetracker characteristics and measuring in nonoptimal conditions goes beyond our infant research, but applies to a much broader range of participant groups.

The problem is best illustrated with an example. Let's consider a binocular measurement (i.e., by tracking both eyes) of an infant participant using a Tobii TX300, a common eyetracker in infant research. According to the Tobii specifications,[1] a spatial accuracy of 0.4° (which is an average offset of gaze position of 0.4 cm on screen at 57 cm viewing distance) and a spatial precision of 0.09° are achieved under a specific amount of light with a participant fixed in the center of the head movement box by means of a chinrest. However, in a more realistic lab setting, both nonoptimal lighting conditions and an infant not fixed in a chinrest (i.e., an infant that is able to move) will deteriorate the accuracy of the eyetracker. Furthermore, suboptimal calibration with the infant due to large calibration stimuli (which is common in infant research) affects eyetracker accuracy even further, since it is impossible to know where exactly the infant is looking (e.g., at the top or bottom of the calibration stimulus) while calibrating. These examples apply not only to infant studies: In several research fields it is often not possible or even desirable to test a participant in optimal settings for various reasons, including ethical ones. This applies to any study in which the participant cannot be instructed to sit or be restrained in the optimal position in the eye trackers' headbox, whether the participants are infants, schoolchildren, or patients with Down's syndrome, attention-deficit hyperactivity disorder (ADHD), or muscular disorders. How then to interpret the technical specifications of an eyetracker for research in nonoptimal circumstances? Can we assume that the eyetracker with the best specifications will still perform best when pushed beyond its comfort zone?

Here we propose a set of tests to qualitatively assess eyetrackers' performance in nonoptimal conditions, in order to aid potential users of eyetrackers in their choice, and to indicate potential issues in interpretation of eyetracking data. The focus is not on determining the best system, but specifically on whether eyetrackers are robust to a set of head movements often seen in eyetracking research with infants. As one of the optimal conditions for an eyetracker (as described above) is that a participant is positioned in the middle of the headbox (i.e., the space in which reliable tracking of the eyes is possible), preferably moves as little as possible and looks straight at the screen, we were interested in eyetrackers' performance during the changing positions of the infant. We modeled the infants' changing head position in two behaviors: (1) looking away from the screen and back, and (2) shifting head orientations. During these movements and orientations, we investigated whether the eyetrackers still reported gaze data. If the system did report gaze data, we were interested in whether there was any indication of systematic offsets (i.e., the same offset across trials and participants) or unsystematic offsets (i.e., highly variable across trials and participants). Although these movements and orientations were inspired by infant research, they are relevant for any research field in which the participant cannot be instructed or positioned fully to the experimenters' liking. In addition to modeling these two behaviors, we investigated whether system-specific issues during these behaviors are important for data analysis—for instance, during the detection of periods in which the eye remains still (fixations) and periods of ballistic movement (saccades). We tested eight different eyetracking setups and discuss our findings with regard to their applications to eyetracking in difficult groups such as infants, children, and certain patient groups.

Method

Participants

A total of nine volunteers participated in the study. Each of eight eyetracking setups (see Table 1) was tested with five of these nine volunteers. The setups were tested in two labs: at Utrecht University, The Netherlands, and at Lund University, Sweden. Because of this, only two out of nine participants (R.H. and T.C., the first and second authors) could participate in all setups. Due to varying availability of the eyetracking setups, the testing order was not identical for each participant. Mean age was 29.2 years (*SD* = 8.15 years). All participants had normal or corrected-to-normal vision and reported no ocular deficits. Seven of the participants had previous experience with participating in and conducting eyetracking research.

Apparatus

We used eight different eyetracking setups from three different manufacturers (SMI, Tobii, and LC Technologies), all of which are in production as of this writing. This specific set was chosen for two reasons: (1) SMI and Tobii are two manufacturers of the most common eyetrackers in Northwest Europe and (2) these eyetrackers are commonly used in the labs that we are familiar with. The LC Technologies EyeFollower was

[1] www.tobii.com/Global/Analysis/Marketing/Brochures/ProductBrochures/Tobii_TX300_Technical_Specification_Leaflet.pdf

Table 1 Participation of volunteers in each eyetracking setup

Participant	Tobii X2-60	Tobii T120	Tobii TX300: 120 Hz	Tobii TX300: 300 Hz	SMI REDm: 60 Hz	SMI REDm: 120 Hz	SMI RED 250	LC Technologies EyeFollower
C.F.	X		X	X	X	X		
R.H.	X	X	X	X	X	X	X	X
T.C.	X	X	X	X	X	X	X	X
J.L.	X		X	X	X	X		
L.W.	X		X	X	X	X		
M.N.		X					X	X
I.H.		X					X	X
D.W.		X					X	
A.S.								X

included because it is specifically designed to allow a large range of movement. No conflicts of interest with any of the manufacturers were present. The eyetracker specifications provided by the manufacturers are summarized in Table 2. Although these specifications give a good overview of the general differences between the devices, we did not make any assumptions about an eyetracker's performance on that basis.

The Tobii X2-60 and SMI REDm eyetrackers were the only eyetrackers not attached to a screen. They were positioned at the bottom of a laptop display placed on a table, as they are most commonly used. As a result, participants looked slightly down at the screen with the Tobii X2-60 and SMI REDm as compared to the other setups. The other eyetrackers were integrated in a monitor and were positioned perpendicular to the table, with the middle of the screen roughly at eye height.

Stimulus presentation was done using MATLAB and Psychophysics Toolbox (Brainard, 1997). Data recording was done using the iView SDK for SMI, the Tobii SDK for Tobii, and EyeGaze for LC Technologies. Afterward, all data files were imported into MATLAB for data analysis.

Procedure

Participants were positioned in front of the eyetracker (see the Apparatus section for more info on the different setups) at the optimal tracking distance for each setup. This was either done by using the distance values reported back by the eyetracking software (i.e., Tobii SDK for Tobii, iView SDK for SMI) or by the experimenter positioning the participant at the optimal tracking distance reported by the manufacturer (for the LC Technologies EyeFollower). Hereafter, a calibration sequence was run. For the SMI systems, a five-point calibration was performed followed by a validation sequence. For the Tobii systems, a five-point calibration was run followed by an inspection of calibration results using the Tobii SDK.[2] For the LC Technologies, a nine-point calibration was run followed by a validation sequence. Calibration was repeated until the quality of calibration was judged to be good enough by the experimenters. After positioning and calibration, participants were presented with three tasks: a "recovery" task, a "yaw orientation" task, and a "roll orientation" task. If more than one eye-tracking setup was tested in one session, measurements were interspersed with short breaks.

The main interest for the recovery task was to determine what happens when an eyetracker loses track of the eyes (i.e., theoretically it cannot report gaze data anymore) and when it restarts reporting gaze data. The focus here was on *how* an eyetracker recovers, not *when*. The main interest for the yaw orientation task and the roll orientation task was to determine how eyetrackers cope with eyes in nonoptimal head orientations.

Recovery task Each trial consisted of a 5-s period, during which a fixation dot was presented in the center of the screen. Participants were instructed to look at the fixation dot whenever they looked at the screen. After 1 s, a low-pitched beep sounded, and after another 2 s, a high-pitched beep sounded. Prior to starting the task, participants were instructed to turn their head to the left or the right at the low beep, and to turn back to their starting position at the high beep. After fixating for another 2 s, the next trial followed.

Yaw orientation task Each trial again consisted of a 5-s period, with a low-pitched beep after 1 s, and a high-pitched beep after another 2 s. Prior to starting the task, participants were instructed to turn their head to the left or to the right as far as possible while maintaining fixation on the fixation dot, and to turn back to their starting orientation at the high beep.

[2] www.tobii.com/en/eye-tracking-research/global/landingpages/analysis-sdk-30/

Table 2 Specifications of each eyetracking setup

	Tobii X2-60	Tobii T120	Tobii TX300: 120 Hz	Tobii TX300: 300 Hz	SMI REDm: 60 Hz	SMI REDm: 120 Hz	SMI RED 250	LC Technologies EyeFollower
Sampling rate	60 Hz	120 Hz	120 Hz	300 Hz	60 Hz	120 Hz	250 Hz	120 Hz interlaced
Accuracy	0.4°	0.4°	0.4°	0.4°	0.5°	0.5°	0.4°	0.4°
Precision	0.34°	0.16°	0.07°	0.07°	0.1°	0.1°	0.03°	*
Headbox	50 × 36 cm (at 70 cm)	30 × 22 cm (at 70 cm)	37 × 17 cm (at 65 cm)	37 × 17 cm (at 65 cm)	32 × 21 cm (at 60 cm)	32 × 21 cm (at 60 cm)	40 × 20 cm (at 70 cm)	76 × 50 cm
Tracking distance	40–90 cm	50–80 cm	50–80 cm	50–80 cm	50–75 cm	50–75 cm	60–80 cm	46–97 cm
Price**	€15,500	€25,900	€32,900	€32,900	€17,500	€17,500	€25,900	€24,400

Specifications are retrieved from the eyetracking manufacturers' product descriptions and are based on binocular data, not processed or filtered after recording. *Not specified by manufacturer. **Price is as quoted to us or our department. Excludes both VAT and any discounts

Figure 1 depicts the axis of head rotation along which the orientation took place.

Roll orientation task Each trial again consisted of a 5-s period, with a low-pitched beep after 1 s, and a high-pitched beep after another 2 s. Prior to starting the task, participants were instructed to tilt their head to the left or the right while maintaining fixation on the fixation dot at the low beep, and to turn back to their starting orientation at the high beep. Figure 1 depicts the axis of head rotation along which the orientation took place.

Participants were asked to remain fixated on a central point on screen so that we could compare offset of gaze data to this central point prior to movement and after movement, where we would expect offsets to be minimal at least prior to movement. Schematic overviews of the final head positions in all tasks, as well as the axes along which head orientations took place, are given in Fig. 1. For each task, participants started with ten trials with head movements to the left, followed by ten trials to the right (by instruction of the experimenter). To minimize the amount of movements in the wrong direction made by participants, the order of directions and the order of tasks were not counterbalanced. In addition, the order of movement directions was not counterbalanced across participants. This meant that the order of movement directions was identical for all participants in each eyetracking setup. No instructions were given to the participants with regard to blinking.

Although the head movements and orientations in the tasks presented here were far from ideal for the eyetrackers, we did measure with highly motivated, cooperating participants who understood the instructions and who were all familiar with psychological research using eyetrackers.

Results

This section is divided into four subsections: three describing point of regard (i.e., the gaze position reported by the eyetracker) from the three separate tasks, and one on data loss. We consider first, however, the results that we would expect if an eyetracker were to perform perfectly in all tasks. In the recovery task, we would expect point of regard to start on the fixation dot. Hereafter, we would expect to see the point of regard moving to the left or right as the gaze moved off screen. Finally, we would expect to see the reverse once the participant returned gaze back to the screen. For the yaw orientation task and the roll orientation task, we would expect point of regard to be on the fixation dot throughout the trial, since no eye movements away from the fixation dot were made. We, would, however, expect some minor changes in point of regard during the head movements themselves, since the eyes

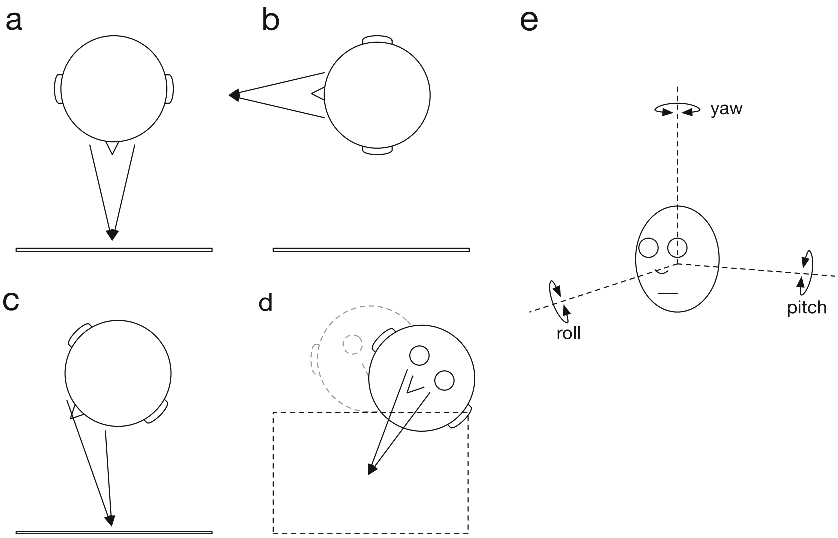


Fig. 1 Schematic overview of head positions. (**a**) Top view of starting and ending positions. (**b & c**) Top views of positions between beeps in (**b**) the recovery task and (**c**) the yaw orientation task. (**d**) Front view of a position between beeps in the roll orientation task. The dotted rectangle in this panel represents the screen and is moved down relative to its original position in the experiment for clarification purposes only. (**e**) Axes along which the head can rotate. Rotations along the yaw and roll axes are included in the present study in the yaw orientation and roll orientation tasks, respectively

would have to correct their orientation for the change of head orientation. After the results from the recovery, yaw orientation, and roll orientation tasks are discussed, data loss in the three tasks will be presented: We do this to substantiate the qualitative results that we will outline in the first three sections and establish the robustness of the eyetrackers' gaze reporting in the two orientation tasks.

Recovery task

The main purpose for the recovery task was to determine what happens when eyetrackers lose track of and regain tracking of the eyes. Figure 2 depicts the horizontal screen coordinates reported by the eyetracker for all trials from the recovery task, separated for all eyetracker setups. Only samples of which the coordinates were on screen are depicted in Fig. 2; the vertical axis depicts the entire screen width. Between 0 and 1 s after trial onset, participants remained fixated on the middle of the screen, indicated by the coordinates being in the center of the vertical axis. When the first beep sounded (i.e., the first vertical bar in the graph), participants looked away from the screen. Between 2 and 3 s after trial onset, the maximum position of the head orientation was reached in almost all of the trials (i.e., the gaze was completely turned away from the screen, and no point of regard was reported by the eyetracker). In nearly all of the setups, the eyetracker was capable of following the gaze off-screen, as is visible from the eye coordinates moving from the center to the edge of the screen (which is also the edge of the graph). The eyetrackers with higher sample frequencies (120 Hz and higher) obviously collected more samples from onset to offset of movement, which means detection of looking away was easier in these systems: More information was available to do so, because there were more samples during the movement. When participants returned to fixation after 3 s (indicated by the second vertical bar), several remarkable differences can be seen between the setups. Only three eyetracker setups were able to track the gaze of the participant immediately upon gaze reentering the screen, as is visible from the horizontal coordinate moving from the edge of the graph back to the center: the Tobii TX300 (at both 120 and 300 Hz) and the LC Technologies EyeFollower. If mere sample frequency were the determining factor in being able to track the eyes immediately when the eyes returned to the screen, one would expect the SMI REDm at 120 Hz, the Tobii T120, and the SMI RED250 to be able to do this, as well. In the latter systems, however, it is impossible to detect the gaze returning back to the fixation point on screen *before* the point of regard hits the fixation dot. This led us to question what was the difference for the eyetracker prior to and after the gaze shift; the gaze shifts were identical, only in opposite directions.

Since the SMI RED250 and Tobii TX300 have similar sampling frequencies, 250 versus 300 Hz, we expected similar behavior from the eyetrackers in terms of tracking the eyes upon returning to the screen. Because the SMI RED250 and Tobii TX300 showed different behaviors, we decided to take a closer look at what these systems report when a participant looks away. Two typical trials in the recovery task from these two systems are depicted in Fig. 3. As is visible in the top graph, the SMI RED250 shows an increase in intersample interval shortly (~500 ms) after the gaze is shifted away from the screen completely. This means that during this period, the sampling frequency drops sharply from the manufacturer-specified 250 Hz and stabilizes at 20 Hz while the gaze

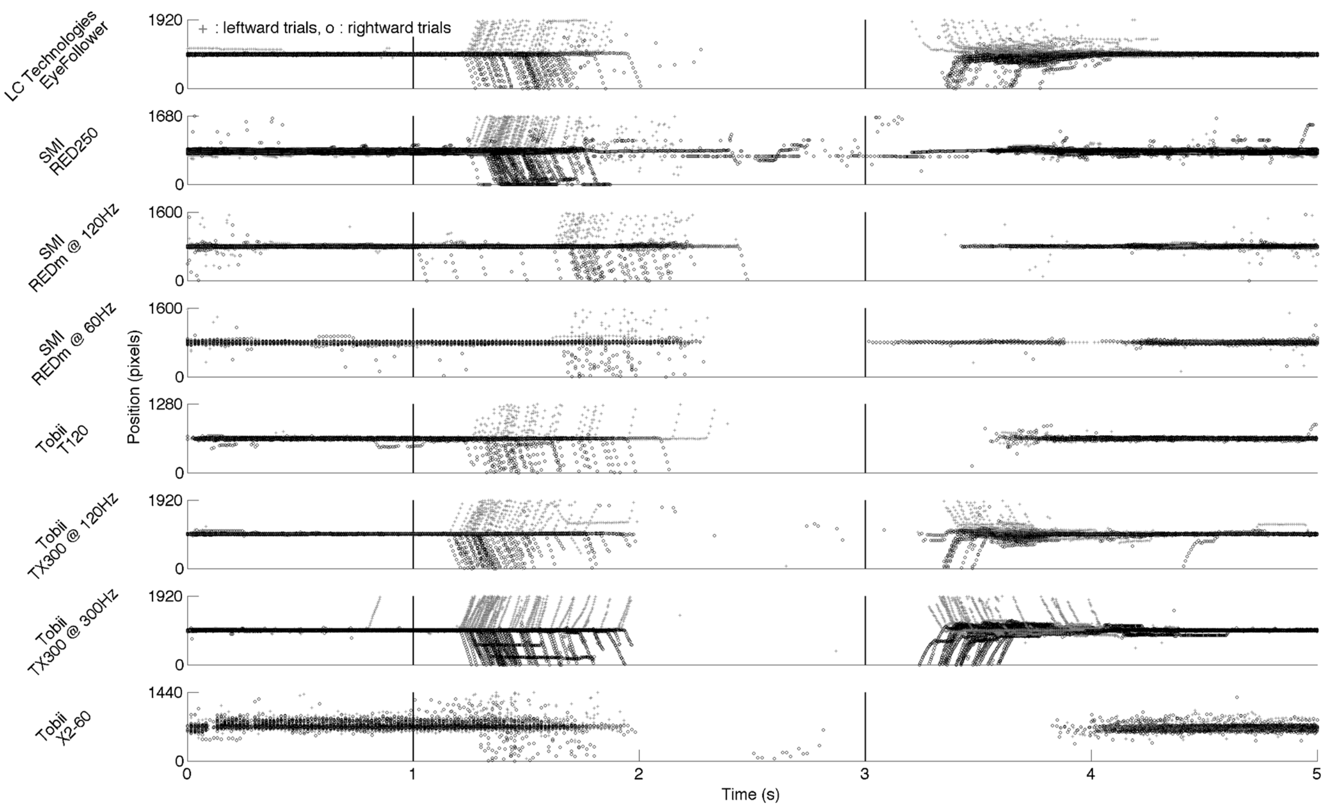
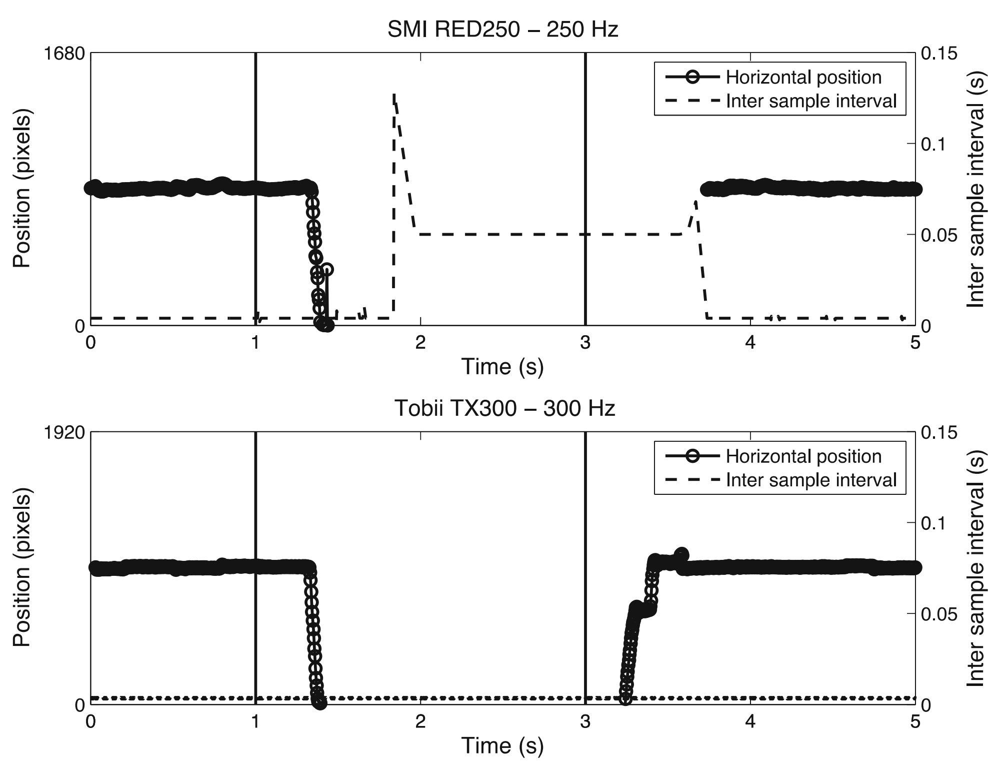


Fig. 2 Data from the recovery task for each eyetracker setup. Raw data from both eyes were averaged for all setups except the LC Technologies EyeFollower, which gives alternating coordinates from left and right eyes every other sample. All trials from five participants are overlaid. Since the movement in Task 1 was only in the horizontal direction, only horizontal coordinates are given. Black circles indicate trials with rightward movement, and gray crosses indicate trials with leftward movement. Black vertical bars indicate the beeps that signaled a movement to be made away from the screen at 1 s, and back to the screen at 3 s. Due to the audio latency in the SMI REDm/MATLAB setup, the trials here are shifted rightward slightly

Fig. 3 Horizontal positions and intersample intervals in two typical trials from the recovery task in the SMI RED250 and the Tobii TX300 at 300 Hz. In the SMI RED250 system, the intersample interval increases once the head is shifted completely away from the screen, whereas this does not occur in the Tobii TX300



remains turned away. When the gaze returns to screen, the intersample interval spikes, after which it returns to 250 Hz, and gaze data are again reported. In addition to the SMI RED250, the SMI REDm (at both 60 and 120 Hz) shows the same pattern. The Tobii TX300, on the other hand, continues to report empty samples at 300 Hz when the eyes are lost, and it reports gaze data shortly after the gaze is back on the screen. Although the reasons for this difference are technical in nature and beyond the scope of this article, whether or not an eyetracker drops in sampling frequency when movements are made might be important for the detection of fixations, saccades, and smooth pursuit (event detection). In addition, it might be important for detecting gaze shifts back to the screen.

Yaw orientation

The main purpose for the yaw orientation task was to determine how eyetrackers deal with nonoptimal head orientations; in this case, the head turned sideways, with the eyes still on the screen. In this orientation, one of the eyes moves (partially) behind the nasion. The primary questions were whether the eyetracker would be able to report gaze data during the nonoptimal orientation, and whether there would be any indication of (un)systematic errors when doing so. Figure 4 depicts the horizontal screen coordinates (vertical coordinates are provided in the supplementary materials) for all trials from the yaw orientation task for participants R.H. and T.C. in all eyetracker setups. Only the two observers who participated in all eight setups are pictured, because we wanted to determine whether the offset between the point of regard as reported by the eyetracker and the fixation dot were similar or different across participants. Although only the data from participants R.H. and T.C. are described, the patterns of eyetracker performance were comparable across all five participants in each setup: Separate data for all participants for a subset of systems (LC Technologies EyeFollower, Tobii TX300, and SMI RED 250) are provided in the supplementary materials. Furthermore, the data from all participants overlaid for every setup are also provided in the supplementary materials.

Particularly, the SMI RED250 and the Tobii T120 struggled during the yaw orientation task: As can be seen in Fig. 4, both trackers appear to report few data between 2 and 3 s, during which the difference in head orientation from the starting orientation was maximal. The LC Technologies EyeFollower, SMI REDm, Tobii TX300, and Tobii X2-60 were able to calculate point of regard, albeit with an offset from the fixation dot. The point-of-regard signal from the SMI REDm, at both 60 and 120 Hz, showed large offsets from the fixation dot during the shifted head orientation. Surprisingly, this shift seemed very persistent in some trials, causing large offsets in point of regard even when participants moved their head back to the starting orientation. This is visible from Fig. 4, where the SMI REDm at 120 Hz for participant T.C. shows a persistent offset in point of regard (i.e., the second black line shifted upward from the line in the center of the screen). This occurs even between 0 and 1 s, when the participant is not positioned in a nonoptimal head orientation. This occurred for participant T.C., but not for participant R.H., yet also for other participants (see the supplementary materials). These large offsets likely resulted from the SMI REDm system switching the position of the eyes. If the head is rotated left, the position of the right eye moves toward the left eye. If the left eye is then lost from the eye image, the right eye is mistaken for the left; hence, a large but systematic shift in point of regard occurs. This seemed to happen with the SMI REDm on several occasions, and the switch appeared quite persistent after the point of regard returned to the starting position.[3]

Roll orientation task

The main purpose for the roll orientation task was identical to that of the yaw orientation task: to determine how eyetrackers deal with nonoptimal head orientations. The primary questions were whether the eyetracker would be able to report gaze data during the nonoptimal orientation, and whether there would be any indication of (un)systematic errors when doing so. Figure 5 depicts the horizontal screen coordinates for all trials from the roll orientation task for participants R.H. and T.C. in all eyetracker setups. As in the yaw orientation task, only the two participants who participated in all eight setups are pictured, and only horizontal coordinates are given. Separate data for all participants in a subset of the systems and overlaid data from the remaining participants in all systems, as well as the vertical coordinates, are given in the supplementary materials.

As in the yaw orientation task, the Tobii T120 appeared to struggle with the head orientations in the roll orientation task, losing almost all data after movement in participants R.H. and T.C. The SMI RED250 also appeared to struggle, losing a lot of data during the nonoptimal head orientation, especially with participant T.C., though less than in the yaw orientation task. The eyetracker systems that did continue to report gaze data during the shifted head orientation appeared to report larger offsets from the fixation dot than in the yaw orientation task. In particular, the LC Technologies EyeFollower reported large offsets, although they seemed to be very systematic for participant R.H.: The offsets from the fixation dot are consistent in direction and amplitude over trials.

[3] SMI is aware of this issue and is working on it (Pötter 2014, personal communication).

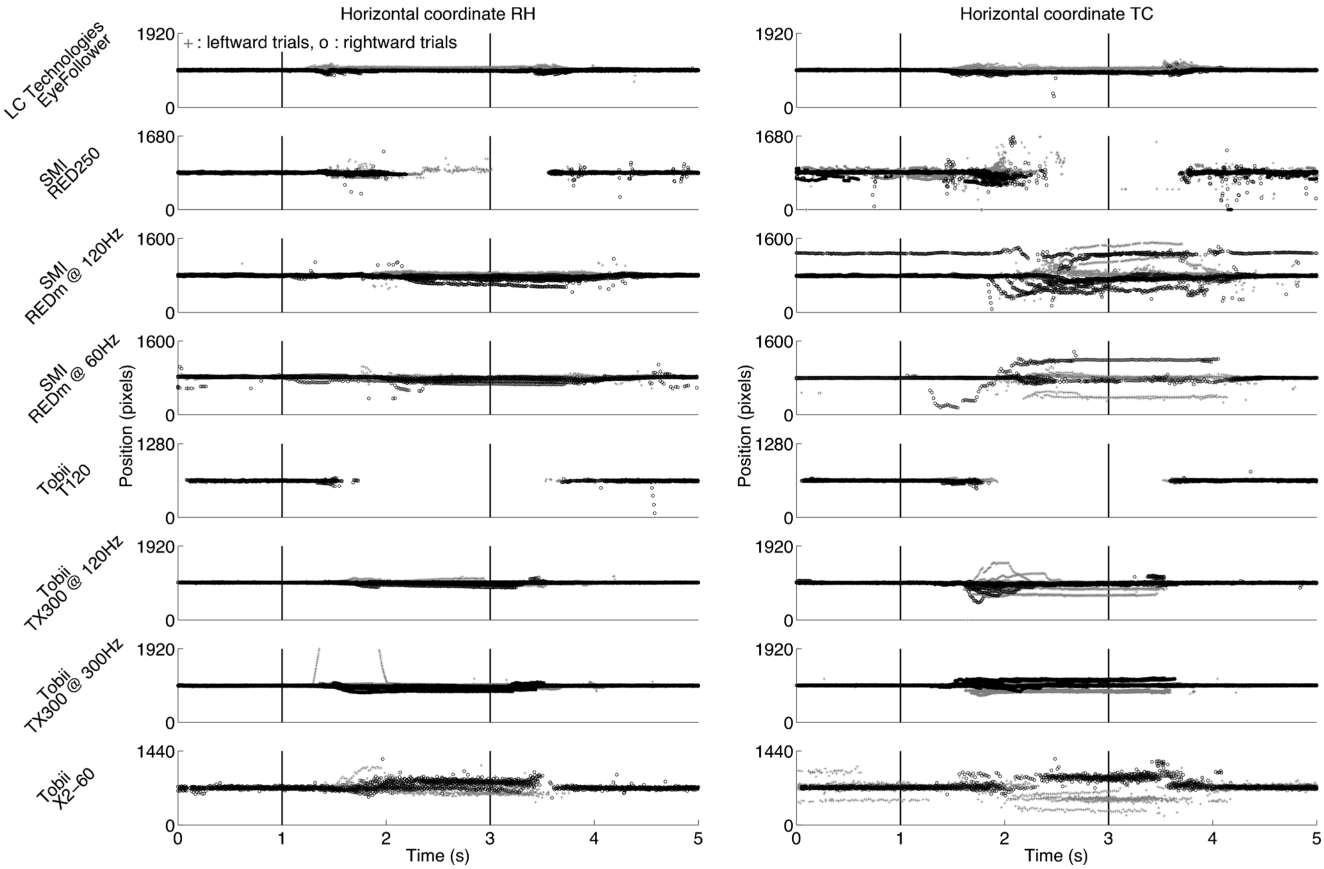


Fig. 4 Data from the yaw orientation task for each eyetracker setup. Raw data from both eyes were averaged for all setups except the LC Technologies EyeFollower, which gives alternating coordinates from the left and right eyes every other sample. All trials from observer R.H. (left) and observer T.C. (right) are overlaid. As in Task 1, only horizontal coordinates are given, because the horizontal coordinates showed the most interesting results. Vertical coordinates are given in the supplementary materials. Black circles indicate trials with rightward rotation, and gray crosses indicate trials with leftward rotation. Black vertical bars indicate the beeps that signaled a rotation to be made away from the screen at 1 s and back to the screen at 3 s. Due to the audio latency in the SMI REDm/MATLAB setup, the trials here are shifted rightward slightly

Data loss

In addition to the qualitative results described above, we attempted to give some quantification to the performance of the eyetrackers in our tasks, and calculated the proportions of data loss over all participants in two time periods. Data loss was defined as the number of samples in a time period in which point of regard (i.e., the gaze position reported by the eyetracker) was not reported for either eye, divided by the theoretical number of samples in that time period. This number would, for example, be 250 samples for the RED250 in 1 s, although this might not be the actual number of samples that were recorded, due to a drop in sampling rate. The first time period was between 0 and 1 s after trial onset. In this period, participants fixated on the screen center, and no movement occurred. We would expect little to no data loss here. The second period was between 2 and 3 s after trial onset. In this period, head movement occurred, and the difference from the first time period was largest. We would expect the data loss to be highest here if an eyetracker could not deal with a specific movement, and to be 0 if an eyetracker could deal with the specific movement. There was no gaze on screen in the recovery task between 2 and 3 s, and we included data loss in this task as a comparison: We expected little to no data loss in the period of 0–1 s, and near maximum data loss in the period 2–3 s. Data loss for all three tasks is given in Fig. 6.

In the recovery task, the proportion of data loss from 2 to 3 s was near 1 for all eyetracker setups, although it was slightly lower for the SMI REDm. However, this could be explained by the SMI REDm/MATLAB setup including a slightly longer audio latency, meaning that gaze shifts might have started later. From 0 to 1 s, the proportion of data loss was between 0 and .2 for all eyetracker setups. In addition, our qualitative assessment in the yaw orientation task, that the SMI RED250 and Tobii T120 were struggling most, was confirmed by the high proportions of data loss (between .6 and .85) reported in Fig. 6. The LC Technologies EyeFollower reported a high proportion of data loss, despite reporting fairly stable gaze

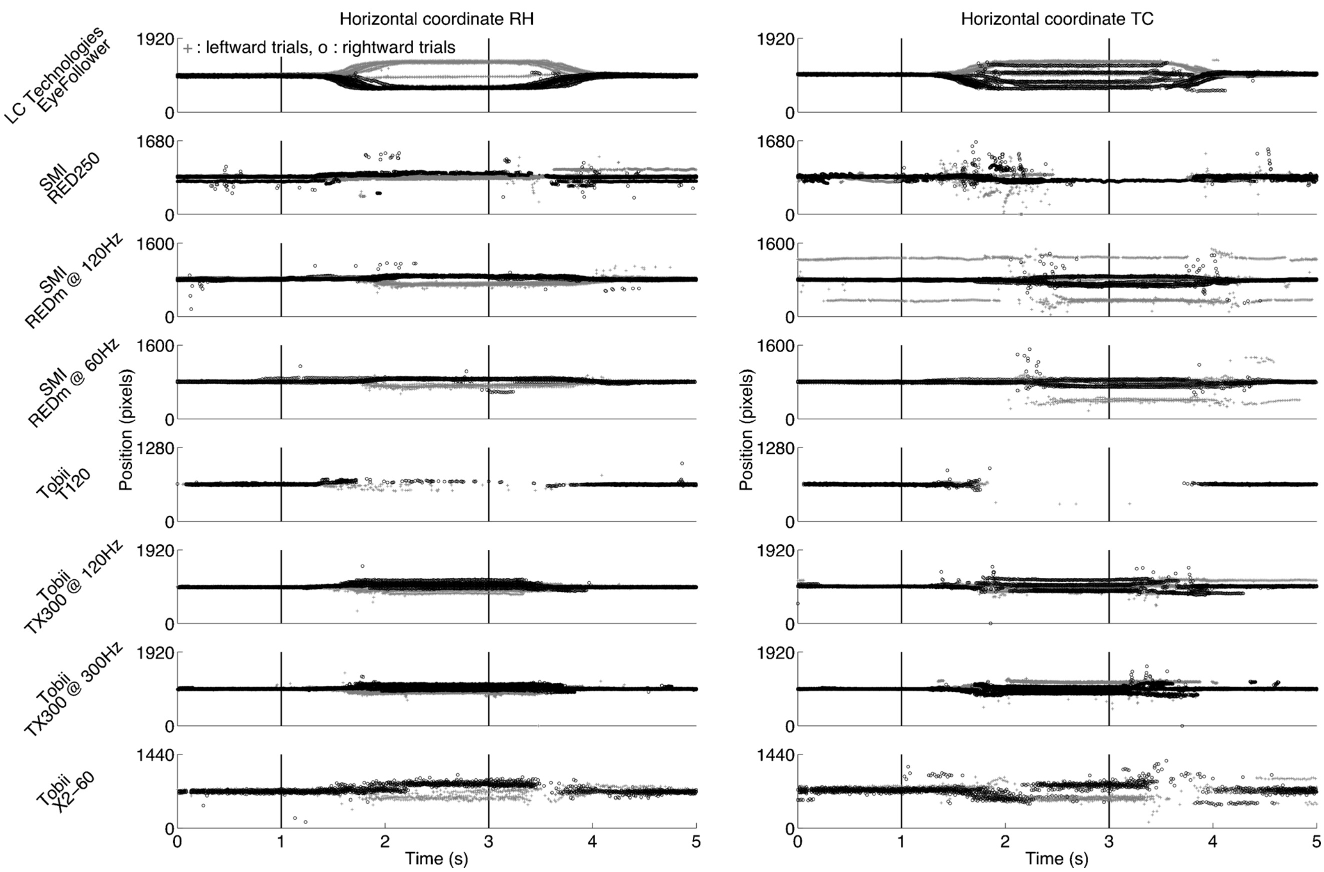


Fig. 5 Data from the roll orientation task for each eyetracker setup. Raw data from both eyes were averaged for all setups except the LC Technologies EyeFollower, which gives alternating coordinates from the left and right eyes every other sample. All trials from observer R.H. (left) and observer T.C. (right) are overlaid. As in Tasks 1 and 2, only horizontal coordinates are given, because the horizontal coordinates showed the most interesting results. Vertical coordinates are given in the supplementary materials. Black circles indicate trials with rightward rotation, and gray crosses indicate trials with leftward rotation. Black vertical bars indicate the beeps that signaled a rotation to be made away from the screen at 1 s and back to the screen at 3 s. Due to the audio latency in the SMI REDm/MATLAB setup, the trials here are shifted rightward slightly

data. As we reported before, in the yaw orientation task one of the eyes moves (partially) behind the nasion. The LC Technologies EyeFollower measures at 120 Hz interlaced (i.e., samples from each eye every other sample), and as one eye moved behind the nasion, we only obtained valid samples every other sample (i.e., from the eye that was still visible). We therefore expected that system to obtain a proportion of data loss nearing .5, which is confirmed in Fig. 6. In the roll orientation task, again the SMI RED250 and the Tobii T120 reported the highest proportions of data loss (between .5 and .9). The proportion of data loss for the LC Technologies EyeFollower is low in the roll orientation task, as compared to the yaw orientation task. In the roll orientation task, both eyes remained directly visible to the eyetracker camera, provided that the trackers' headbox was large enough. It is therefore not surprising that, unlike in the yaw orientation task, in which one eye moves behind the nasion, the proportion of data loss for the LC Technologies EyeFollower was low. The proportions of data loss for the SMI REDm setups, the Tobii TX300 setups, and the Tobii X2-60 are similar across the two orientation tasks.

Discussion

The aim of the present study was to provide a set of qualitative tests for judging eyetracker performance when movements and rotations are made. These movements and orientations were inspired by infant eyetracking research: Infants tend to be distracted easily, and often make gaze shifts away from and back to the screen. Additionally, it is practically impossible to restrain infants' movements in front of the eyetracker, which results in nonoptimal head orientations (i.e., a rotation other than looking straight ahead at the eyetracker). We modeled movements often seen in infant research in two behaviors (1) looking away from, and back to, the screen, and (2) head rotations, in order to determine how eyetrackers cope with loss

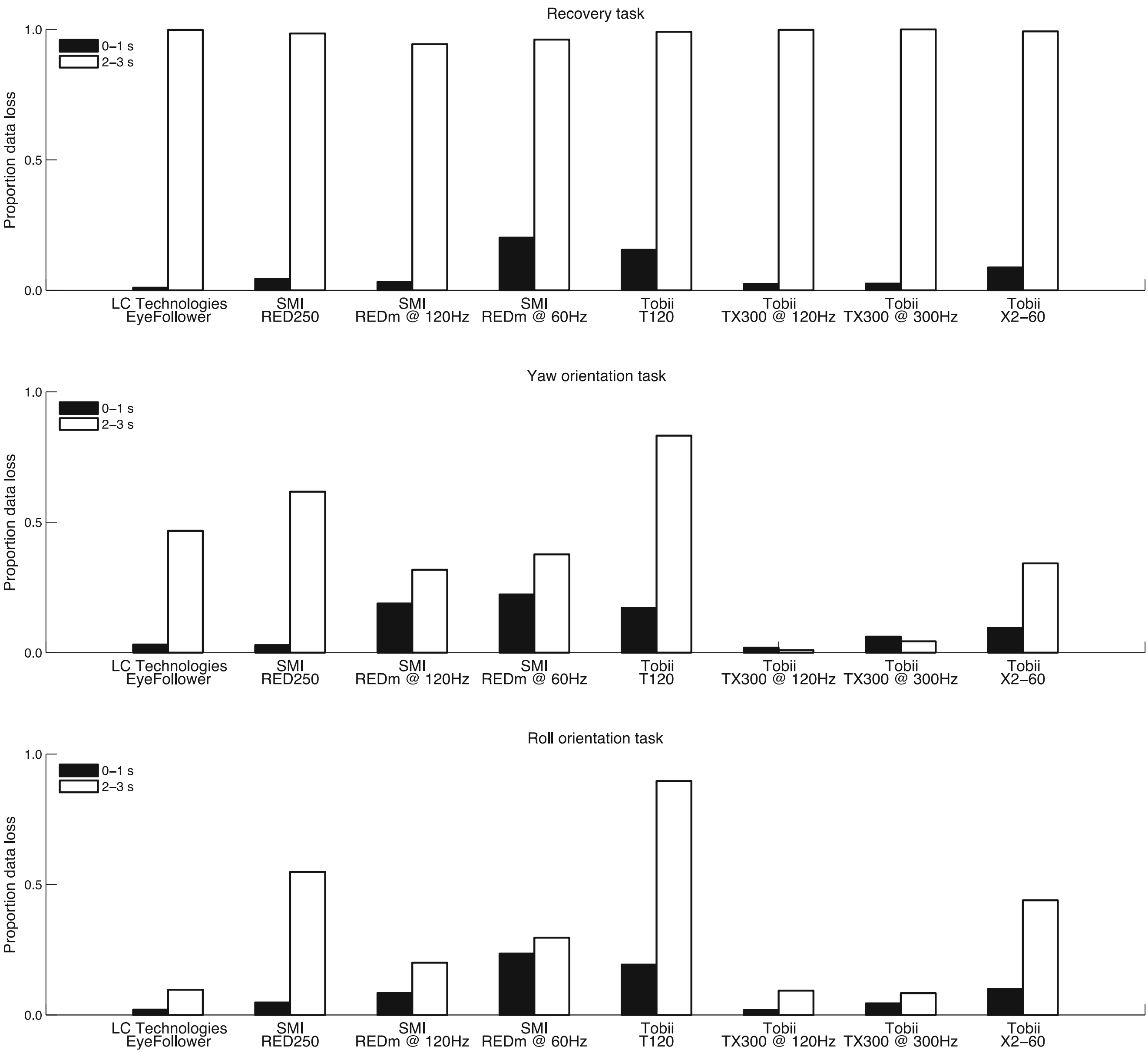


Fig. 6 Proportions of data loss for all eyetracker setups in the recovery, yaw orientation, and roll orientation tasks. Data loss was calculated in two periods: 0–1 s, when data loss was expected to be minimal, and 2–3 s, when data loss was expected to be maximal in a trial. Data loss was calculated by dividing the number of samples in which at least one eye was found by the theoretical number of samples we would expect in a second on the basis of the sampling frequency. For example, for the SMI RED250 we expected 250 samples in 1 s; see also the comparison of the SMI RED250 versus the Tobii TX300 in a trial in Fig. 3

of tracking the eyes in nonoptimal head positions. The head movements performed in the present tests are, however, not solely relevant to infant research. Any eyetracking research field in which participants cannot be fully instructed or positioned to the researchers' liking will encounter nonoptimal head orientations: For example when studying schoolchildren, patients with Down's syndrome, ADHD, or muscular disorders.

We report that it is possible with most eyetrackers to detect a gaze shift away from the screen, although it might be done more reliably when sampling frequency is high (i.e., 120 Hz or higher), as a result of more samples, and thus more information, being available during the shift. In our set of eyetracking setups, detecting a gaze shift back to the screen on the basis of horizontal gaze coordinates *during* the shift could be done with both the LC Technologies EyeFollower and Tobii TX300. This is potentially useful if one is interested in showing a dynamic stimulus for a fixed viewing time: If one wants to pause the eyetracker once an infant looks away and unpause it once the infant looks back, the gaze shift information has to be available. Implementing this could be done, for instance, by building a detector that is able to identify the gaze shift to the screen as it returns in the Tobii

TX300 trial in Fig. 3. This is opposed to the gaze coordinate after returning to the screen in the SMI RED250 trial in Fig. 3, in which no point of regard is reported *during* the gaze shift back to the screen. The latter could also occur as a result of technical difficulties during the measurement: that is, a participant is actually looking continuously at the screen, but the eyetracker is not able to report the point of regard.

We furthermore report that SMI eyetrackers drop in sampling frequency roughly 500 ms after the eyetracker cannot find the eyes and does not report gaze data anymore. The Tobii eyetrackers, on the other hand, do not drop in sampling frequency. We can speculate on two different situations in which the eyes would be lost. If the eyes were lost due to a blink or a hand moving in front of the eyes for less than 500 ms, we would expect the SMI eyetrackers to recover as quickly as the Tobii eyetrackers. If the eyes were lost due to a gaze shift away from the screen, which in our recent infant studies (unpublished data) are typically longer than 500 ms, we would expect the SMI eyetrackers to recover more slowly than the Tobii eyetrackers. Whether an eyetracker drops in sampling frequency might also be important to take into account for two other forms of data analysis: (1) when detecting periods of movements and stillness of the eyes (often referred to as event detection), *if* drops in sampling frequency occur also when gaze samples are reported, or (2) when interpolating over periods of data loss. Interpolating is often done in infant research (see, e.g., Frank, Vul, & Johnson, 2009; Saez de Urabain, Johnson, & Smith, 2014, for different methods of interpolating data loss), where short periods of data loss are more common than in adult studies. In event detection, one categorizes periods of stillness of the eyes as fixations, periods of ballistic movement as saccades, and periods of constant movement as smooth pursuit (e.g., when the eyes are following a moving object). Event detection is often done by calculating the velocity of the eye across samples (Holmqvist et al., 2011): that is, by taking the distance between samples on screen and dividing it by the intersample interval. If one assumes that the intersample interval is fixed (i.e., 4 ms, in the case of a 250-Hz SMI RED250), velocity values, and consequently the decision whether a sample belongs to a fixation or a saccade, will be different from when the actual intersample intervals reported in the eyetracker output are used. Changes in sampling frequency are thus vital for valid detection of events. As a consequence, when not taking into account drops in sampling frequency, all metrics based on the timing of the detected events might be invalid, too—for instance, when one calculates the total time spent looking at an area of interest.

Eyetrackers vary in how robust they are to nonoptimal head orientations (i.e., whether they report gaze data and whether it is accurate). Both the SMI RED250 and Tobii T120 lose a lot of data during nonoptimal head orientations, whereas the other eyetracker setups do report gaze data. In particular, the Tobii TX300 reports very little data loss in both the yaw orientation task and the roll orientation task. Even when an eyetracker is able to calculate point of regard in a nonoptimal head orientation, there appear to be large offsets relative to optimal head orientation. These offsets are quite systematic for some systems (i.e., in the roll orientation task for the LC Technologies EyeFollower), and less systematic for others. Since we have no reason to assume that participants were in fact looking somewhere else on the screen but the fixation dot, we therefore assume that this offset was a result of the reduced accuracy of the eyetracker. This would mean that although an eyetracker is robust to a certain movement and continues to report gaze data, accuracy might very well deteriorate.

If a participant is in a nonoptimal orientation for the eyetracker and accuracy drops, this should be taken into account when creating areas of interests for eyetracking data analysis. When doing area-of-interest analysis, even a change in accuracy of 0.5° can significantly alter the outcomes (Holmqvist, Nyström, & Mulvey, 2012). If one disregards the shifted orientations of one's participants, and assumes that the data are accurate, study outcomes might be invalid. Consider, for example, a study assessing the time that participants spend looking at the nose of a static face that is presented in the middle of the screen. If the participant is orientated nonoptimally, accuracy might deteriorate, and the time spent looking at the nose might be shifted up or down, toward the eyes or mouth. This would result in a lower total time spent looking at the nose than would have been observed when the participant was positioned optimally. One possible solution is to increase the size of areas of interest, although this would depend on the stimulus one was using (see, e.g., Holmqvist et al., 2011, for a more detailed discussion of creating areas of interest). One could furthermore discuss whether reporting highly inaccurate data is better than reporting no data at all when the participant is in a nonoptimal head orientation. Regardless of whether or not it is preferable to obtain inaccurate data during nonoptimal head orientations, researchers should be aware that the head orientation of their participant could be a factor in gaze accuracy. The present study should be helpful in determining whether or not this is something to be wary of in one's specific situation.

The present study provides a first assessment of several eyetrackers in nonoptimal conditions. However, several limitations should be noted. First, the movements in the three tasks were performed across three different axes of movement, but independent measures of head rotation across these axes were lacking. Future research might benefit from more controlled measurements of rotation in each direction, to better investigate eyetracker performance during nonoptimal head rotations. Second, no specific instructions with regard to blinking were given, nor were blinks removed from the analyses. Although blinks might have contributed slightly to the measures of data loss, we reason that the amount of data loss due to one or more blinks would be negligible compared to the periods of data loss that we observed.

Since research does not always take place in optimal conditions, eyetracker accuracy, precision, and sampling frequency might not be the best guides when deciding which eyetracker to use. We have proposed here a number of qualitative tests to investigate eyetracker recovery and robustness to nonoptimal head orientations, and have highlighted several eyetracker-specific issues. These tests, combined with an eyetracker's specifications, should provide a good overview of what an eyetracker can and cannot do, and could help researchers make a more informed choice in choosing an eyetracker. Furthermore, they provide an example of how to test one's own eyetracker for suitability to doing research in which the movements that we highlighted might occur. We conclude that although eyetracker specifications are certainly important, they are not necessarily decisive factors when one knows that research will not take place in optimal conditions. An eyetracker's performance in nonoptimal conditions can be just as important.

Author note We thank Kenneth Holmqvist, Fiona Mulvey, and the Eye-Tracking Group at the Lund University Humanities Lab for providing eyetracking setups in this study. We also thank Edwin Dalmaijer for help with the manuscript. This work was supported by a Netherlands Organization for Scientific Research (NWO) VICI Grant (45307004) and by an NWO Gravition Grant (024.001.003), both to C.K.

References

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10,* 433–436. doi:10.1163/156856897X00357

Frank, M. C., Vul, E., & Johnson, S. P. (2009). Development of infants' attention to faces during the first year. *Cognition, 110,* 160–170. doi:10.1016/j.cognition.2008.11.010

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford, UK: Oxford University Press.

Holmqvist, K., Nyström, M., & Mulvey, F. (2012). Eye tracker data quality: What it is and how to measure it. *Proceedings of the Symposium on Eye Tracking Research and Applications: ETRA ', 12,* 45–52. doi:10.1145/2168556.2168563

Saez de Urabain, I. R., Johnson, M. H., & Smith, T. J. (2014). GraFIX: A semiautomatic approach for parsing low- and high-quality eye-tracking data. *Behavior Research Methods*. Advance online publication. doi:10.3758/s13428-014-0456-0