A nonvisual eye tracker calibration method for video-based tracking

4 authors, including:

Vanessa Harrar
Université de Montréal
47 PUBLICATIONS   **1,170** CITATIONS

William Trung Le
University of Montreal Hospital Research Centre
13 PUBLICATIONS   **38** CITATIONS

A nonvisual eye tracker calibration method for video-based tracking

Vanessa Harrar

Vision, Attention, and Action Laboratory (VISATTAC), School of Optometry, University of Montreal, Montreal, Quebec, Canada ✉

William Le Trung

Vision, Attention, and Action Laboratory (VISATTAC), School of Optometry, University of Montreal, Montreal, Quebec, Canada ✉

Anton Malienko

Vision, Attention, and Action Laboratory (VISATTAC), School of Optometry, University of Montreal, Montreal, Quebec, Canada ✉

Aarlenne Zein Khan

Vision, Attention, and Action Laboratory (VISATTAC), School of Optometry, University of Montreal, Montreal, Quebec, Canada ✉

Video-based eye trackers have enabled major advancements in our understanding of eye movements through their ease of use and their non-invasiveness. One necessity to obtain accurate eye recordings using video-based trackers is calibration. The aim of the current study was to determine the feasibility and reliability of alternative calibration methods for scenarios in which the standard visual-calibration is not possible. Fourteen participants were tested using the EyeLink 1000 Plus video-based eye tracker, and each completed the following 5-point calibration methods: 1) *standard* visual-target calibration, 2) *described calibration* where participants were provided with verbal instructions about where to direct their eyes (without vision of the screen), 3) *proprioceptive* calibration where participants were asked to look at their hidden finger, 4) *replacement* calibration, where the visual calibration was performed by 3 different people; the calibrators were temporary substitutes for the participants. Following calibration, participants performed a simple visually-guided saccade task to 16 randomly presented targets on a grid. We found that precision errors were comparable across the alternative calibration methods. In terms of accuracy, compared to the standard calibration, non-visual calibration methods (described and proprioception) led to significantly larger errors, whilst the replacement calibration method had much smaller errors. In conditions where calibration is not possible, for example when testing blind or visually impaired people who are unable to foveate the calibration targets, we suggest that using a single stand-in to perform the calibration is a simple and easy alternative calibration method, which should only cause a minimal decrease in accuracy.

Introduction

Modern video-based eye trackers are one of the primary research tools used to measure eye movement properties such as fixation, saccades, and smooth pursuit under various experimental conditions. There are various eye detection and eye trackers available on the market, and gaze estimation models vary according to the photometric and geometric properties of the systems (e.g., the number of cameras used, the number of reference points, whether head translations or eye rotations are measured, 2D or 3D output, etc.), (Hansen & Ji, 2010). Video-based eye trackers typically use the relative position between the center of the pupil and the corneal reflection (the first Purkinje image) to compute gaze. One of the requirements to gathering high quality data using eye tracking systems is to determine the appropriate transformation so that eye movements can be mapped to targets in the real world (Hammoud, 2008; Kasprowski, Harezlak, & Stasch, 2014).

The method to achieve this mapping relies on having participants fixate visual targets in the real world with a known position, and then recording their eye position. This calibration procedure, generally done before an experiment begins, creates a coordinate transformation equation so that the position of the eyes throughout the experiment can be transformed into the relevant reference frame (Hammoud, 2008).

There are, however, several conditions when the visually based calibration procedure compromises the integrity of the experiment, or else is entirely not possible. Examples include during experimental conditions requiring complete darkness, when eye movements to nonvisual targets are being recorded, or when attempting to record the eye movements in blind participants. Some developers have designed eye-tracking systems that do not require a calibration phase, with varied levels of success. For example, Zhu and Ji (2004) reported that their calibration-less system has a resolution of 5° horizontally and 8° vertically, while more recently Chen and Ji (2015) reported much better resolution with <3° of error in the measurement. Instead of a system that skips calibration altogether (but appears to have considerable known error in the measurements), an alternative is to simplify or shorten the calibration process (Harezlak, Kasprowski, & Stasch, 2014), or to use nonvisual calibration methods.

In previous studies that have measured eye movements in the blind, calibration was either absent or unreliable because the blind subjects were unable to direct the fovea to the visual targets or maintain fixation (Leigh & Zee, 1980). Leigh and Zee reported limited success using electrooculography (EOG) to measure eye movements, because most of their blind subjects suffered from an ocular disease that attenuated the corneoretinal potential.

Several solutions have been developed to bypass the visual calibration step, with varied levels of success. The best way to get around the calibration problem in blind participants, and still measure eye movements precisely, is by using *scleral search coils* in a magnetic field (method described by Robinson, 1963, used with blind participants in Schneider et al., 2013). Though search coils are the most reliable measure of 3D eye rotations, they are avoided unless absolutely necessary because of their invasiveness and inconvenience.

Another solution, when visual calibration is not possible, is to generate a general calibration factor by calculating the average conversion factor from several control (sighted) participants. Although a *calibration factor* certainly has some errors associated with determining the absolute point of regard, it is a good solution for relative eye movement measurements such as determining the gain in vestibulo-ocular reflex in the blind, as demonstrated in Sherman and Keller (1986). Several papers reporting eye movements in blind participants have used a best approximate or mean calibration factor to convert raw eye positions to degrees (Hall, Gordon, Hainline, Abramov, & Engber, 2000; Kömpf & Piper, 1987; Sherman & Keller, 1986). Sherman and Keller applied a calibration factor to their data, which was obtained by averaging the calibration factors from eight sighted control participants; they noted that the average had a standard deviation of 18%, due to the variability between participants (Sherman & Keller, 1986). The variability between calibrating participants is caused by differences between people's eye physiology (e.g., pupil diameter, shape of the cornea), which affects the stability of the eye feature data, and therefore also the quality of the calibration factor (Nyström, Andersson, Holmqvist, & Van De Weijer, 2013). It remains unclear to what degree the use of this mean calibration factor increases measurement errors compared to individually calibrated recordings. It is also unknown if the increased error as a result of using an average calibration factor changes with the eccentricity of the targets, i.e., if the working area of a task should be restricted when using a calibration factor. More importantly, not all eye trackers will allow the user to input a calibration factor (e.g., EyeLink eye trackers), and attempting to apply a calibration factor to the data set offline is computationally intensive. There is, therefore, a need for a user-friendly alternative to the visually based calibration procedure. A simple alternative to manually inputting a calibration factor could be to have a replacement person (a stand-in) perform the calibration step immediately before testing. Although this is certainly a simpler alternative, we do not yet know if using calibration values from a single individual as a replacement calibrator, rather than an average across several individuals, provides data with minimal error.

Endo, Fujikado, Kanda, Morimoto, and Nishida (2014) investigated the accuracy and precision of a calibration method based on *reaching movements*. In their protocol, participants attempted to fixate their occluded fingertip. Endo et al. (2014) reported a fairly consistent 20% underestimation of actual eye position when using this proprioception calibration method, i.e., saccades to the fingertips were shorter than saccades to the visual targets. Undershooting saccades, and the increased variability of saccades to proprioceptive targets, compared to control visual targets, has been robustly reported and confirmed (Ren et al., 2006; Van Beers, Baraduc, & Wolpert, 2002; Van Beers, Sittig, & van der Gon, 1998). Thus, eye tracking calibrated with proprioceptive targets is a reasonably simple alternative when testing blind participants; however, the error is large.

There is not yet a consensus on the best simple alternative calibration method, mostly because the alternative calibration methods outlined above are

either too complex or have not been directly compared against each other. Thus, we do not know if the working areas of the simple calibration methods are the same, or if some are better for large eccentricities. The aim of the current study is to compare a variety of practical calibration methods, for use when the usual visual calibration is not possible, and determine the most reliable alternative calibration method. We have identified three simple alternative calibration methods that we compared to the typically used visually guided five-point calibration method. First is the "Proprioceptive Method," where participants looked at their unseen fingertip. Second is the "Described Method," where the targets' position was verbally described to the participants. Third is the "Replacement Method," where calibration was performed by three separate participants, without combining their data into a single calibration factor, so that the validity of each calibrator could be determined. Although the proprioceptive method has been previously tested, the described method and the replacement method (as described here) are novel simple alternative calibration methods. We compared accuracy and precision of these different calibration methods with 14 sighted participants to determine the best simple alternative calibration method for future noninvasive eye-tracking experiments with blind or visually impaired participants.

Methods

Participants

Seventeen participants took part in this study (age range: 18 to 40 years, $M = 24.5$ years, $SD = 9.3$ years, eight male, 16 right-handed). All participants had normal or corrected-to-normal vision. Participants all signed a consent form prior to the study, preapproved by the Committee of Ethics on Health Research at the University of Montreal. Of the 17, 14 were testers (13 of which were right-handed) and three were calibrators (two male) for the replacement method. Two of the three calibrators are authors (WT & AM).

Apparatus

Participants sat in a dark room facing a VIEWPixx LCD monitor (VPixx Technologies, Montreal, Quebec, Canada), with specifications of 60 Hz, 52×29 cm or $1,920 \times 1,080$ pixels, on which the stimuli were presented at a distance of 32.5 cm from the participants' eyes. The participants' head was immobilized via a chin and forehead support placed at the edge of the table on which the monitor was located. Eye move-

ments were recorded using an infrared-emitting video-based eye tracker (EyeLink 1000 Plus, SR Research, Mississauga, ON, Canada). The tracker recorded the infrared reflection of the eye from a mirror placed between the participant's eyes and the display monitor (Figure 1A). In terms of EyeLink tracking settings, we used mono mode and pupil corneal reflection, at a 1K sample rate with an ellipse tracking. During experiments involving nonvisual calibration, an opaque black cloth was placed behind the infrared mirror, obscuring the display monitor from view.

Procedure

Four different calibration conditions were conducted in a single session. The four conditions differed only by the calibration method as described below. In all of the conditions, five calibration points were presented in a cross at 15° of visual angle (Figure 1B). The order of the presentation of the five points was randomized for every condition. Each location was tested once, and a validation procedure was not performed.

In the *standard calibration* condition, the screen was visible and the participant performed the customary visually-guided calibration (EyeLink 1000 Plus standard five-point calibration). The participant was asked to fixate each calibration point, and each one was accepted after a 1,000 ms fixation duration (threshold).

In the *described calibration condition*, the calibration points could not be seen by the participant (the screen was obscured by an opaque cloth). The points' positions were described to the participants by the experimenter via simple cardinal direction commands, always using center as a reference between each target. For example, "Please look straight ahead. Now look down. Go back to straight ahead. Now look to the left. Look straight ahead. Now look right. Look straight ahead. Now look up." The participant moved their gaze accordingly and fixated on imagined positions. Researchers had a reference diagram of the positions of the calibration dots, and in the case where participants did not move nearly far enough, they were told to, for example, "move a little more to the right."

In the *proprioception calibration condition*, once again the calibration points could not be seen by the participant. An experimenter placed the participants' right index finger on the fixation point, and the participant moved their gaze and fixated on the felt (sensed) position of their fingertip. Note that all participants but one were right-handed (as identified by self-report).

For the *replacement condition*, the calibration was performed by a stand-in instead of the participant. The participant remained seated but their chair was briefly wheeled away from the setup. The calibrator, standing,

A Experiment setup

B Calibration sequence

Center fixation
1000 ms
fixation threshold

4 calibration
points presented
in random order
1000 ms fixation
threshold
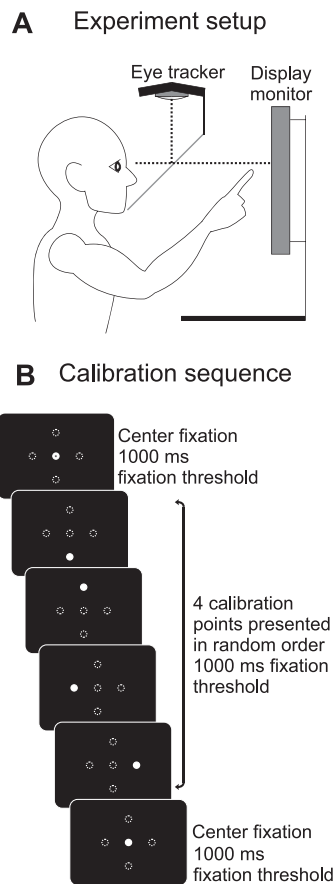
Center fixation
1000 ms
fixation threshold

Figure 1. Experimental setup and calibration sequence. (A) Experiment setup. Participants sat facing a vertical display monitor on which stimuli were displayed. An EyeLink 1000 Plus tower mount was used, placed above the setup. A transparent infrared reflecting mirror (gray diagonal line) was used to reflect infrared light from the eye tracker camera to the eyes, to enable eye position recordings (dotted line). During the proprioception and described conditions, a black cloth was placed behind the mirror, thus occluding vision of the display monitor and hand. In the proprioception condition as shown in the figure, the participant's fingertip was placed at the different calibration points (by the experimenter) and the participant was asked to look at their fingertip, even though it was occluded. (B) Calibration sequence. Each calibration began with the central fixation dot (filled white dot). If the eye position remained stationary for 1,000 ms, the fixation duration threshold, the eye tracker recorded the eye position. Once the eye position was recorded, the next calibration point was presented. The open circles show all possible target positions, for clarity here, but were not visible to the participant. The four noncentral calibration points were presented in random order. Participants were asked to follow the dots as they appeared on the screen. The central fixation point was then presented again at the end of calibration. Validation was not performed.

placed their head on the chin and forehead rest and completed the standard visual calibration procedure (as described already). After the calibrator performed the five-point calibration, the participants placed their head back on the chin rest and ran the experiment. The same three stand-ins were used for all participants, known as Replacement 1, Replacement 2, and Replacement 3. Their data were not combined, to see the difference across replacements. Therefore, the replacement calibration was run three times for each participant, each time followed by the experiment.

Immediately following each calibration procedure, the opaque cloth was removed (if necessary) and the participant performed the visually-guided experiment. The experiment required the participants to saccade to and fixate targets presented on the screen. Each target was a blue circle with a diameter of 1°, presented on a black background. The target randomly appeared on the screen at one of 16 possible locations (Figure 2A; corners on the gray grid) for a duration of 1,000 ms. Horizontal and vertical target positions were −20° (left of center or below center), −7°, 7°, and 20°. The order of target location was randomized for every experiment. The intertrial interval was randomized uniformly between 500 ms and 1,500 ms, at 100 ms intervals. Each target was presented twice at each location. Thus, the visually-guided experiment included 32 trials, taking a total of about 3 min to complete, and was completed by each participant following each calibration method.

Data analysis

A total of 2,688 trials were collected from the 14 participants (32 positions × 2 repetitions × 6 conditions). Saccade detection was performed offline in MATLAB (MathWorks, Natick, MA) using a velocity threshold of 50°/s. Fixation position was obtained by calculating the mean x and y position from 200 to 400 ms after the saccade. Saccade onset, offset, and the fixation interval were visually verified and, if need be, adjusted by the experimenter to ensure that only a stable fixation position was calculated. Trials were removed if there was no stable fixation due to eye tracker noise, no signal, or blinking during the fixation period. A total of 40 trials were removed (1.6% of total trials), leaving 2,644 trials for analysis.

We calculated accuracy and precision across all target positions, for each participant, for each condition, to enable statistical inferences. Accuracy was calculated as the absolute difference between each fixation location and the corresponding target location comprising both x and y dimensions, then averaged for each participant. As we were interested in the magnitude of error, absolute error was used, because otherwise signed errors may have cancelled out across
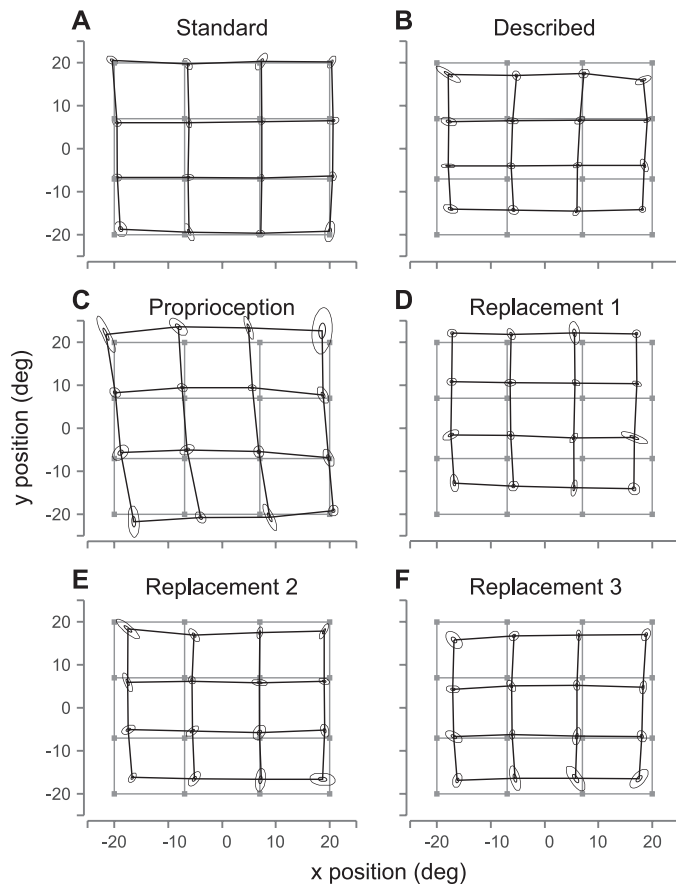
Figure 2. Mean fixation endpoints. Mean horizontal (x) and vertical (y) fixation endpoints across all participants are depicted for the different calibration conditions; standard (A), described (B), proprioception (C), replacement 1 (D), replacement 2 (E), and replacement 3 (F). The mean fixation point, at each target location, is represented by the center of the ellipses and the crossing of the black lines. The size and orientation of the inner thicker-lined black ellipses at each target location represents the dispersion error fitted across all participants for each target position (the ellipses fit 68% (one standard deviation) of the data). The outer thinner-lined ellipses are a three-fold expansion of the 68% ellipses, for visual purposes. Target positions are depicted by gray squares and grid.

participants. We used dispersion error as a measure of precision. We calculated mean corrected errors by calculating the difference of the two fixations (for each target location, for each participant, within each condition) from the mean, separately for the horizontal and vertical positions. Ellipses were fit to these mean corrected errors across all participants. The parameters of the ellipse were then calculated across all targets, for each condition. Dispersion error was defined as the area of the ellipse fitted to 68% of the data. Specifically, dispersion error was calculated as $A \times B \times \pi$, with A and B being the semi major and minor axes of the fitted ellipse, respectively. One-way repeated-measures analyses of variance (ANOVAs; corrected for sphericity when required) were used to compare accuracy and precision across conditions, with Bonferroni corrected post-hoc pairwise comparisons when justified.

Results

Figure 2 depicts the mean fixation points, averaged across participants, as a black grid overlaying the target grid (in gray), for each of the calibration conditions, the standard condition (A), the described condition (B), the proprioception condition (C), and the three replacement conditions (D–F). The crossings of the black grid depict the mean fixation position for each target and the inner black ellipses reflect the dispersion error, fit to all participants' mean corrected errors at each target position [the ellipses fit 68% (one standard deviation) of the data]. Because the ellipses were small, and thus their orientations were mostly hard to see, we also overlaid the same ellipses expanded by a factor of 3 (outer thinner ellipses).

Imperfect calibration can be simplified to two broad categories of errors, (1) a systematic shift and/or (2) a compression/expansion/rotation of the estimated gaze relative to actual gaze (assumed to be perfectly on target). In the standard calibration condition (Figure 2A), the recorded eye position matches the veridical position very well. In contrast, in the described condition (Figure 2B) there is a compression particularly for the outermost targets. In the proprioceptive condition (Figure 2C), one can observe a rotation and expansion, again particularly for the outermost target locations. Finally, in the replacement conditions (Figure 2D–F), the offsets are smaller than with the non-visual calibration methods, but here, too, there is a slight increase in error at the outer oblique targets. With replacement 1, there is a compression as well as a slight shift (e.g., Figure 2D), whereas with replacement 2 and 3 only compression can be observed (Figure 2E, F). In summary, fixation endpoints were closest to the target positions in the standard calibration condition (Figure 2A), with little shift or compression, whereas all other calibration conditions resulted in a poorer match. We also observed high overall precision, across all targets and all conditions, as can be seen by the very small 68% confidence ellipses (thinner outer ellipses are a threefold expansion), implying good reliability in the eye tracker recordings. We quantified the degree of error comparing accuracy and precision across the different conditions, to determine the best alternative calibration method to the standard calibration. In addition, we quantified the degree of error as a function of target eccentricity.
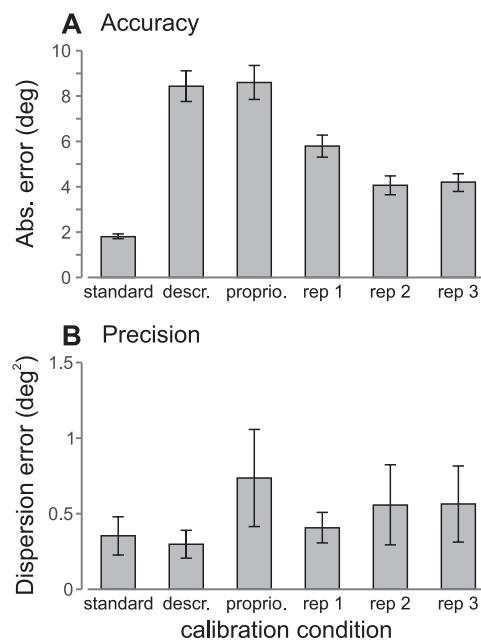
Figure 3. Accuracy and precision. (A) Accuracy. The mean absolute error from the target, across participants and target locations, is plotted for each of the different calibration conditions, in the following order from left to right; standard, described, proprioception, replacement 1, replacement 2, and replacement 3. (B) Precision. Dispersion errors (in deg^2) are plotted averaged across participants for each of the different calibration conditions, in the same order as (A). Error bars are standard error of the mean across participants.

Overall accuracy

Figure 3A depicts the mean absolute error of fixation position from the target, taking both the x and y dimensions into account. We performed one-way repeated measures ANOVA with calibration condition as a factor (six levels). There was a significant main effect of condition, $F(3.13, 40.65) = 30.25$, $p < 0.001$, $\eta^2_p = 0.7$.

Post-hoc pairwise comparisons with Bonferroni adjustments revealed significant differences between the standard calibration condition ($M = 1.81°$) and all other conditions (all $p < 0.002$). The described calibration condition ($M = 8.43°$) was not significantly different from replacement 1 ($M = 5.79°$, $p = 0.159$), but was significantly different from replacement 2 ($M = 4.03°$, $p < 0.001$) and replacement 3 ($M = 4.17°$, $p < 0.001$). The proprioception calibration condition ($M = 8.55°$) was not significantly different from replacement 1 ($p = 0.209$) but was significantly different from replacement 2 and replacement 3 (both $p < 0.002$). There was no difference between the described and the proprioceptive conditions ($p > 0.99$). Finally, there was no significant overall difference between any of the replacements ($p > 0.139$).

Taken together, the results suggest that accuracy is best when the standard visual calibration is used, and poorest when using the nonvisual calibration methods. Calibration accuracy was equally poor when calibration points were described verbally, or when calibration points were localized through proprioception. There were some differences in accuracy across the three replacement conditions; replacement 1 transferred more poorly to the participants' eyes than the other two replacements. Accuracy is generally better in the replacement calibration conditions compared to the nonvisual calibration methods, making the replacement calibration method a second best to the standard calibration.

Overall precision

Figure 3B shows the dispersion error for the different calibration conditions, calculated across all targets for each participant, then averaged across participants. Dispersion error is the area of distribution of mean corrected errors (area of the ellipse fitted to 68% of the data). As can be observed in Figure 3B, dispersion error was low and similar across all conditions. There appears to be slightly lower precision for the proprioception condition, but this was not significant [no main effect of condition: $F(1.82, 23.66) = 1.65$, $p = 0.21$, $\eta^2_p = 0.11$.

In summary, the pattern of results from the precision analysis shows that precision was similar across all conditions, showing overall good reliability in the eye tracker recordings.

Accuracy and precision as a function of target eccentricity

It is well known that accuracy is best within the central work space, directly in front, and generally tends to become worse with more eccentric eye positions. We investigated whether this decrease in accuracy and precision was consistent across the different calibration conditions. Figure 4A and 4B depict absolute error and dispersion error as a function of three target eccentricities; see inserts on x-axis. We performed a two-way repeated measures ANOVA for accuracy and precision separately with calibration condition (six levels) and target eccentricity (three levels) as factors.

For accuracy, we found a significant main effect of condition, as shown before, $F(3.1, 40.9) = 29.17$, $p < 0.001$, $\eta^2_p = 0.69$, as well as a significant main effect of target eccentricity, $F(1.1, 13.9) = 28.14$, $p < 0.001$, $\eta^2_p = 0.68$, and a significant interaction effect, $F(1.8, 23.3) = 3.98$, $p = 0.037$, $\eta^2_p = 0.23$. The significant interaction

A Accuracy

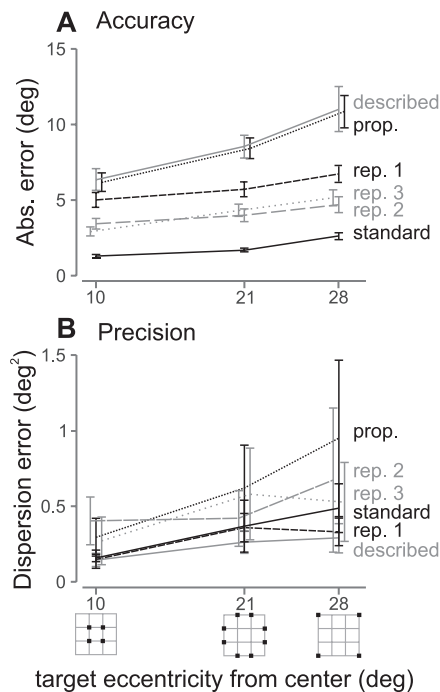**B** Precision

target eccentricity from center (deg)

Figure 4. Accuracy and precision with respect to target eccentricity. (A) Accuracy. The mean absolute errors are plotted for the three target eccentricities as described in the text and icons on the x-axis. The different conditions are coded as follows: standard (black solid lines), described (gray solid lines), proprioception (black dotted lines), replacement 1 (black dashed lines), replacement 2 (gray dashed lines), and replacement 3 (gray dotted lines). (B) Precision. The mean dispersion errors are plotted as a function of target eccentricity. Error bars are standard error of the mean across participants.

demonstrates that eccentricity affects accuracy more for some calibration methods than others. One-way ANOVAs conducted for the six calibration methods were all significant, $p < 0.014$, $\eta^2_p > 0.37$, demonstrating that eccentricity causes increased errors for all conditions (Figure 4A). In the standard calibration method, when the target eccentricity increased from 10° to 28°, the absolute error increased by 1.3° ($\pm SE = 0.28$). The increase in the standard calibration method was significantly less than the increase in the nonvisual calibration methods (error increase in the described method = 4.7° $\pm$ 1.08, $p = 0.003$; error increase in the proprioceptive method = 4.7° $\pm$ 1.65, $p = 0.043$), but not different from the replacement method (error increase in replacement 1 = 1.7° $\pm$ 0.40, $p = 0.486$; error increase in replacement 2 = 1.3° $\pm$ 0.40, $p = 0.886$; error increase in replacement 3 = 2.2° $\pm$ 0.42, $p = 0.130$).

For precision (Figure 4B), we found no significant main effect of condition ($p > 0.05$), a trend toward an effect of target eccentricity, $F(1.332, 17.3) = 3.99$, $p = 0.052$, $\eta^2_p = 0.235$, and no significant interaction effect ($p > 0.05$). These results show that precision is similar

for the different calibration methods, although it might decrease as a function of target eccentricity.

Discussion

We tested a variety of different alternatives to the traditional visual five-point calibration. Reliability in the eye movement measurements was no different following the standard versus the alternative calibration methods. In terms of accuracy, although the standard visual calibration had the least amount of error, the alternative calibration methods demonstrated some compression and shifts. Overall, for accuracy, the results showed that nonvisual calibration methods including the proprioceptive calibration method (pointing) and the described method (verbal instructions) were the poorest substitutes for standard calibration. In contrast, the three replacement calibrators demonstrated only small increases in absolute error, compared to the standard calibration method. We therefore propose that in conditions where it is not possible to perform the standard visual five-point calibration, for example, when testing blind participants, using a replacement calibrator (a stand-in) adds a relatively small degree of error to the eye measurements.

The described method, where participants were verbally given cardinal coordinates, had large intersubject errors and considerable compression. These errors are due to the fact that the eccentricity of the target (the distance from centre) was not directly available through vision or any other sense. People's eye movements tend to undershoot particularly for large centrifugal saccades (Kapoula & Robinson, 1986). This explains why the underestimation was more pronounced for the outer targets than the inner targets, in both the described and the proprioception calibration methods.

The calibration procedure that relies on proprioception, in which participants were asked to look at their sensed fingertip position, caused sizeable systematic errors, replicating previous reports (Endo et al., 2014; Ren et al., 2006; Van Beers et al., 1998; Van Beers et al., 2002). The large error is partially due to the fact that there are multiple sources of errors when making an eye movement to a proprioceptive target: knowledge of the arm position and knowledge of the eye position. Indeed, the saccade errors reported here were larger than with visual targets, and have a counterclockwise rotational skewed pattern, because the right arm was used for pointing (see errors in tilted coordinates in Ren et al., 2006). These types of elongated tilted errors, specific to the proprioceptive system (Van Beers et al., 2002; Van Beers et al., 1998), indicate that a

predominant source of error measured here in the proprioceptive condition originated from errors in localizing the proprioceptive target, rather than from the eye movements. Further, this error in localizing the hand increases at larger eccentricities.

Overall, we observed high precision in the eye tracker recordings. Dispersion errors were small and did not differ across calibration methods. These findings support the reliability of the video-based eye trackers; we can be confident that regardless of any offsets in the estimation of eye position (accuracy errors), the tracker will report the same offset for the same position in space.

Accuracy and precision were both degraded with more eccentric targets, across all calibration methods. The pattern of increased error with increasing target distance was observed in all of the calibrations tested, but caused particularly large errors in the nonvisual calibration methods. Although the visual calibration methods had a 1° increase from inner to outer targets, the nonvisual calibration methods had a 4° increase in absolute error. Thus, if other researchers intend to use the described or proprioception calibration method, they should keep the distance of the targets to a minimum. If, however, research questions require eccentric targets, and the standard calibration is not a possibility, then the replacement calibration method is a suitable alternative.

By using a replacement calibration method, the targets are visually mapped—with minimal error—and the errors reported instead correspond to the small differences between participants' eye anatomy. The three replacement calibrators tested did not have the same amount of error in eye recordings. Individual face and eye features that vary between replacement calibrators (such as the curvature of eyelashes, the size of the pupils, and cornea curvature) can account for the differences in the errors associated with the three replacement calibrators. In particular, replacement 1 had an upward shift in the data (see Figure 2), compared to the two other replacement calibrators (who generally showed compression), which may have been due to this calibrator's downward-pointing eyelashes (Nyström et al., 2013). In contrast, the general compression effect of all three calibrators, although not consistent across all participants, might be due to the fact that the calibrators were standing and had their head tilted downward in the apparatus, and may thus have had their eyes slightly closer to the screen than the test participants. In addition to the personal calibration parameters associated with the eye, the calibration procedure also calculates parameters associated with the intrinsic aspects of the camera used (camera calibration), the setup and positioning of the screens and cameras (geometric calibration), and determining the eye-gaze mapping function (gazing

mapping calibration) (Hansen & Ji, 2010). Thus, although some parameters are inevitably different between stand-in calibrators and the test participant (personal calibration parameters), they do not appear to vary much between adults. More importantly, the majority of the calibration parameters are the same between participants as they are related to the set-up itself. Based on results from Nyström et al. (2013), future research that uses the replacement method for calibration should also ensure that the calibrator is not wearing contact lenses, does not have blueish eyes, or downward-pointing eyelashes, all of which were shown to significantly interfere with the calibration measurements. Further, to ensure similar pupil size between the calibrator and the experimenter, ambient lighting should be maintained between calibration and testing; in dark adaptation procedures, the calibrator should also undergo dark adaptation before performing the calibration. Using these simple procedures, we can expect an addition of only about 2° of error in accuracy, and similar levels of precision, compared to the standard visual calibration methods. Similar overall errors have been reported in a no-calibration system that uses anthropomorphic averages for individual eye features, such as color, shape, and ambient lighting (Noureddin, Lawrence, & Man, 2005).

Systems have been developed to attempt to bypass the calibration procedure. Some eye trackers minimize the need for calibration by tracking the first and fourth Purkinje reflection (from the back of the lens), known as dual-Purkinje methods (Crane & Steele, 1985; Sigut & Sidha, 2011). However, since the fourth Purkinje image is weak and very difficult to detect, light conditions must be heavily controlled to use these systems. Other systems use multiple light sources (or multiple cameras) to decrease the sensitivity to head movements (see review in Hansen & Ji, 2010). In the area of deep learning, gaze estimation models have been developed that update their parameters online, i.e., learn incrementally, while participants are looking at highly salient images. These statistical approaches include nonlinear approximation (Betke & Kawai, 1999; Chen & Ji, 2011, 2015), and artificial neural networks (Ji & Zhu, 2003; Schneider, Schauerte, & Stiefelhagen, 2014; Stiefelhagen, Yang, & Waibel, 1997). When the head is constrained, statistical models are accurate to within 2° or better (Chen & Ji, 2011), whereas progress in the unconstrained gaze estimation is slower with measurement error down to 10° (Zhang, Sugano, Fritz, & Bulling, 2017). In the novel approach by Pfeuffer (Pfeuffer, Vidal, Turner, Bulling, & Gellersen, 2013), they used moving objects and tracked smooth pursuit eye movements, rather than saccades to a static image, and achieved errors of 1° or less in accuracy measurements.

Although these automatic gaze-detecting systems do not require per-user calibration, they generally still require extensive initial training (in the range of thousands of trials), and are computationally intensive to set up. There is a trade-off with ease of calibration and the error in the measurements (see review in Selvi & Lakshmi, 2015). More importantly, they still require participants to look at visual stimuli, and so could not be used to replace traditional eye tracking calibration for low vision experiments, or experiments performed completely in the dark. Another promising area for future research, in situations where visual calibration is not possible, is to track features from inside the eye, such as blood vessels and the macula, as with the simultaneous scanning laser ophthalmoscopy/optical coherence tomography (Pircher, Baumann, Götzinger, Sattmann, & Hitzenberger, 2007). However, these systems are expensive at the moment.

In research settings, models for eye and gaze have not replaced the standard individual-based calibration, because of the desire to keep the measurement error and cost at a minimum. It would be interesting to see if the simple replacement calibration method suggested here could be used with the high-precision cost-effective design from McGill (Farivar & Michaud-Landry, 2016), making eye-tracking studies dramatically more accessible in a variety of settings.

Using a replacement stand-in to perform the calibration for the blind or visually impaired person is the easiest way around the calibration problem at the moment. With the exception of Schneider et al. (2013) who used search coils, which are calibrated without the involvement of the participant, previous studies investigating eye movements in the blind have either not transformed eye movements into gaze coordinates (Hall et al., 2000), transformed with an average calibration factor determined from a group of sighted participants (Nau, Hertle, & Yang, 2012; Sherman & Keller, 1986), or transformed with an approximate calibration with an unknown degree of error (Kömpf & Piper, 1987; Leigh & Zee, 1980). Although determining an average calibration factor from a large group of people is an excellent alternative to visual calibration, not all eye trackers will accept an input for calibration (e.g., EyeLink 1000 Plus).

In the context of studies with populations suffering from partial loss of vision, such as in age-related macular degeneration (AMD), there is also an unknown degree of error in the calibration. Currently, calibration with AMD patients relies on the assumption that patients use a preferred retinal locus (PRL) (Guez, Le Gargasson, Rigaudiere, & O'Regan, 1993) as an alternative to central fixation during eye-tracker calibration (Seiple, Szlyk, McMahon, Pulido, & Fishman, 2005; Tarita-Nistor, Brent, Steinbach, & González, 2011; Thibaut, Delerue, Boucart, & Tran,

2016). Depending on the size of the scotoma, the use of a PRL will add an absolute error in the gaze estimate (on average around 20°, see table 2 in Schuchard, Naseer, de Castro, & Dev, 1999). Further, because participants often have multiple PRLs, which might be used interchangeably during calibration and testing (Crossland, Culham, Kabanarou, & Rubin, 2005; Thibaut et al., 2016), the precision in the gaze estimate would be severely compromised. In contrast, the replacement calibration method can be used to better approximate the central gaze position of patients with central visual impairment, maintaining precision in the measurement independent of the position or number of their PRLs.

Using the replacement method would not be useful in cases where the eye movements of participants are abnormal or nonexistent, e.g., visually impaired participants who have false eyes (since they do not move) or have uneven or nonfunctional pupils (Nyström et al., 2013). Many congenitally blind participants have considerable nystagmus, which limits the calibration, but not the ability to test the eye movements. Finally, in the case of visual deficits affecting gaze behavior such as cortical lesions, researchers need to keep in mind that the measurements would be centered on pupil gaze rather than "effective" gaze. However, these last problems are inherent to testing such populations rather than related to calibration issues per se.

The replacement calibration method could also be useful for measuring eye movements of sighted participants in circumstances where visual feedback is unwanted. In many cases, it would be important to avoid a visual calibration because it might interfere with the learning of an adapted movement (e.g., after using prism glasses or following saccadic adaptation). We suggest that the standard visual calibration may provide anchors and visual feedback that could partially reverse the learning, and therefore diminish the effect that is being studied (e.g., in Alahyane & Pélisson, 2005). Bypassing visual calibration would further allow for an entire experiment to be completed in the dark. Testing eye movements in the dark can provide rich information about the neural circuitry and underlying connections between the eyes and other modalities, for example those underpinning hand-eye coordination, or reflexive audio-visual eye movements.

In conclusion, we have measured three simple alternative calibration methods. We suggest that when using nonvisual calibration, the workspace should be restricted to within 10° eccentricity to minimize measurement errors. We report that the replacement method is a simple and practical calibration alternative for video-based eye tracking when standard calibration is not possible. The replacement method provides reliable eye position data to within 2°, and an average increase of only 1° from inner to outermost targets. In

addition to stability in the measurement, the replacement method is simple and does not require any additional knowledge of the physics required in calibration, and no additional calculations; it can therefore be applied by anyone using a video-based eye tracker. Although we concede that the error is slightly larger in the replacement calibration method compared to the standard visual calibration method, it is a far better alternative than the previously used approximate calibrations, or proprioceptive calibrations. We propose that the replacement calibration would be a good method for experiments intending to measure eye movements in the dark, or the eye movements of visually impaired participants.

Keywords: touch, auditory, multisensory, amodal, crossmodal

Acknowledgments

Commercial relationships: none.
Corresponding author: Aarlenne Zein Khan.
Email: aarlenne.khan@umontreal.ca.
Address: Vision, Attention, and Action Laboratory (VISATTAC), School of Optometry, University of Montreal, Montreal, Quebec, Canada.

References

Alahyane, N., & Pélisson, D. (2005). Long-lasting modifications of saccadic eye movements following adaptation induced in the double-step target paradigm. *Learning & Memory*, *12*(4), 433–443.

Betke, M., & Kawai, J. (1999). *Gaze detection via self-organizing gray-scale units*. Paper presented at the Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems, 1999. Proceedings. International Workshop.

Chen, J., & Ji, Q. (2011). *Probabilistic gaze estimation without active personal calibration*. Paper presented at the Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference.

Chen, J., & Ji, Q. (2015). A probabilistic approach to online eye gaze tracking without explicit personal calibration. *IEEE Transactions on Image Processing*, *24*(3), 1076–1086.

Crane, H. D., & Steele, C. M. (1985). Generation-V dual-Purkinje-image eyetracker. *Applied Optics*, *24*(4), 527–537.

Crossland, M. D., Culham, L. E., Kabanarou, S. A., & Rubin, G. S. (2005). Preferred retinal locus development in patients with macular disease. *Ophthalmology*, *112*(9), 1579–1585.

Endo, T., Fujikado, T., Kanda, H., Morimoto, T., & Nishida, K. (2014). Calibration of eye movements using reaching movements under simulated blindness conditions. *Investigative Ophthalmology & Visual Science*, *55*(13), 4162–4162.

Farivar, R., & Michaud-Landry, D. (2016). Construction and operation of a high-speed, high-precision eye tracker for tight stimulus synchronization and real-time gaze monitoring in human and animal subjects. *Frontiers in Systems Neuroscience*, *10*, 73.

Guez, J.-E., Le Gargasson, J.-F., Rigaudiere, F., & O'Regan, J. K. (1993). Is there a systematic location for the pseudo-fovea in patients with central scotoma? *Vision Research*, *33*(9), 1271–1279.

Hall, E. C., Gordon, J., Hainline, L., Abramov, I., & Engber, K. (2000). Childhood visual experience affects adult voluntary ocular motor control. *Optometry & Vision Science*, *77*(10), 511–523.

Hammoud, R. I. (2008). *Passive eye monitoring: Algorithms, applications and experiments*: Berlin, Germany: Springer Science & Business Media.

Hansen, D. W., & Ji, Q. (2010). In the eye of the beholder: A survey of models for eyes and gaze. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *32*(3), 478–500.

Harezlak, K., Kasprowski, P., & Stasch, M. (2014). Towards accurate eye tracker calibration–methods and procedures. *Procedia Computer Science*, *35*, 1073–1081.

Ji, Q., & Zhu, Z. (2003). Non-intrusive eye and gaze tracking for natural human computer interaction. *MMI-Interaktiv*, *6*, 1439–7854.

Kapoula, Z., & Robinson, D. (1986). Saccadic undershoot is not inevitable: Saccades can be accurate. *Vision Research*, *26*(5), 735–743.

Kasprowski, P., Harezlak, K., & Stasch, M. (2014). Guidelines for the eye tracker calibration using points of regard. *Information Technologies in Biomedicine*, *4*, 225–236.

Kömpf, D., & Piper, H.-F. (1987). Eye movements and vestibulo-ocular reflex in the blind. *Journal of Neurology*, *234*(5), 337–341.

Leigh, R., & Zee, D. S. (1980). Eye movements of the blind. *Investigative Ophthalmology & Visual Science*, *19*(3), 328–331.

Nau, A., Hertle, R. W., & Yang, D. (2012). Effect of tongue stimulation on nystagmus eye movements in blind patients. *Brain Structure and Function*, *217*(3), 761–765.

Noureddin, B., Lawrence, P. D., & Man, C. (2005). A non-contact device for tracking gaze in a human computer interface. *Computer Vision and Image Understanding*, *98*(1), 52–82.

Nyström, M., Andersson, R., Holmqvist, K., & Van De Weijer, J. (2013). The influence of calibration method and eye physiology on eyetracking data quality. *Behavior Research Methods*, *45*(1), 272–288.

Pfeuffer, K., Vidal, M., Turner, J., Bulling, A., & Gellersen, H. (2013). *Pursuit calibration: Making gaze calibration less tedious and more flexible*. Paper presented at the Proceedings of the 26th annual ACM symposium on User Interface Software and Technology.

Pircher, M., Baumann, B., Götzinger, E., Sattmann, H., & Hitzenberger, C. K. (2007). Simultaneous SLO/OCT imaging of the human retina with axial eye motion correction. *Optics Express*, *15*(25), 16922–16932.

Ren, L., Khan, A. Z., Blohm, G., Henriques, D. Y., Sergio, L. E., & Crawford, J. D. (2006). Proprioceptive guidance of saccades in eye-hand coordination. *Journal of Neurophysiology*, *96*(3), 1464–1477.

Robinson, D. A. (1963). A method of measuring eye movemnent using a scieral search coil in a magnetic field. *IEEE Transactions on Bio-medical Electronics*, *10*(4), 137–145.

Schneider, R. M., Thurtell, M. J., Eisele, S., Lincoff, N., Bala, E., & Leigh, R. J. (2013). Neurological basis for eye movements of the blind. *PLoS One*, *8*(2), e56556.

Schneider, T., Schauerte, B., & Stiefelhagen, R. (2014). *Manifold alignment for person independent appearance-based gaze estimation*. Paper presented at the Pattern Recognition (ICPR), 2014 22nd International Conference.

Schuchard, R. A., Naseer, S., de Castro, K., & Dev, J. R. R. (1999). Characteristics of AMD patients with low vision receiving visual rehabilitation. *Journal of Rehabilitation Research and Development*, *36*(4), 294–302.

Seiple, W., Szlyk, J. P., McMahon, T., Pulido, J., & Fishman, G. A. (2005). Eye-movement training for reading in patients with age-related macular degeneration. *Investigative Ophthalmology & Visual Science*, *46*(8), 2886–2896.

Selvi, S., & Lakshmi, C. (2015). A review: Towards quality improvement in real time eye-tracking and gaze detection. *International Journal of Applied Engineering Research*, *10*(6), 15731–15746.

Sherman, K. R., & Keller, E. L. (1986). Vestibulo-ocular reflexes of adventitiously and congenitally blind adults. *Investigative Ophthalmology & Visual Science*, *27*(7), 1154–1159.

Sigut, J., & Sidha, S.-A. (2011). Iris center corneal reflection method for gaze tracking using visible light. *IEEE Transactions on Biomedical Engineering*, *58*(2), 411–419.

Stiefelhagen, R., Yang, J., & Waibel, A. (1997). *Tracking eyes and monitoring eye gaze*. Paper presented at the Proceedings Workshop on Perceptual User Interfaces.

Tarita-Nistor, L., Brent, M. H., Steinbach, M. J., & González, E. G. (2011). Fixation stability during binocular viewing in patients with age-related macular degeneration. *Investigative Ophthalmology & Visual Science*, *52*(3), 1887–1893.

Thibaut, M., Delerue, C., Boucart, M., & Tran, T. (2016). Visual exploration of objects and scenes in patients with age-related macular degeneration. *Journal Francais d'Ophtalmologie*, *39*(1), 82–89.

Van Beers, R., Baraduc, P., & Wolpert, D. M. (2002). Role of uncertainty in sensorimotor control. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *357*(1424), 1137–1145.

Van Beers, R., Sittig, A. C., & van der Gon, J. J. D. (1998). The precision of proprioceptive position sense. *Experimental Brain Research*, *122*(4), 367–377.

Zhang, X., Sugano, Y., Fritz, M., & Bulling, A. (2017). MPIIGaze: Real-world dataset and deep appearance-based gaze estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Zhu, Z., & Ji, Q. (2004). Eye and gaze tracking for interactive graphic display. *Machine Vision and Applications*, *15*(3), 139–148.