Video-based motion-resilient reconstruction of 3D position for fNIRS/EEG head mounted probes

Sagi Jaffe-Dax[1*@], Amit H. Bermano[2,3*] and Lauren L. Emberson[1]

[1]Psychology Department, Princeton University.

[2]Computer Sciences Department, Princeton University

[3]School of Computer Science, Tel-Aviv University.

*These authors contributed equally to this work.

[@] corresponding author: jaffedax@princeton.edu

Highlights

1. Automatic motion-resilient method for estimation of scalp probes position is proposed.
2. The new video-based method is validated compared to the state-of-the-art.
3. A demonstration of successful estimation of probes' positions on an infant scalp is provided.

Abstract

Measuring the exact placement of probes (e.g., electrodes, optodes) on a participant's head is a notoriously difficult step in acquiring neuroimaging data from methods which rely on scalp recordings (e.g., EEG and fNIRS) and particularly difficult for any clinical or developmental population. Existing methods of head measurements require the participant to remain still for a lengthy period of time, are laborious, or require extensive training. We propose an innovative video-based method for estimating the probes' positions relative to the participant's head, which is fast, motion-resilient, and automatic. These advantages allow, for the first time, the use of spatial co-registration methods in a normalized manner on developmental and clinical populations, where lengthy, motion-sensitive measurement methods routinely fail. We show that our method is both reliable and valid compared to existing state-of-the-art methods by estimating the probes positions in a single measurement, and by tracking their translation from session to session. Finally, we show that our automatic method is able to estimate the position of probes on an infant head without lengthy offline procedures, a task which is considered unachievable until now.

Introduction

Functional near infrared spectroscopy (fNIRS) and electroencephalography (EEG) require knowledge of the positioning of probes on subjects' scalp. Knowing the position of probes is an essential pre-requisite for key analytic steps: Aggregating results from a group of subjects, comparing between groups of participants (particularly from different developmental populations where head size systematically varies with group), and for estimating the source of the recorded signal (in fNIRS; Ferradal et al., 2014; Lloyd-Fox et al., 2014b). Desired positions of probes are determined *a priori* in accordance with either the standard positioning system (e.g., 10-20 in EEG; Jurcak et al., 2007), or a predefined array that target specific cortical regions (typically in fNIRS; Aslin et al., 2015). However, it is not possible to place the probes exactly on the desired scalp locations for every participant for every experimental session. This is due to various reasons, including variability in head size and shape between participants, different practices between researchers, and even head growth between sessions in longitudinal studies with infants. There would be substantial improvement in analytic methods if there was a reliable, quick, and robust method for calibrating the 3D positions of the probes relative to the subject's head.

Existing methods for 3D estimation of probes' location are not suitable for early developmental or clinical populations and reduce the portability of these methods (one of their major benefits over MRI). Using the most popular approach, the 3D digitizer (e.g., Polhemus' Fastrak or Patriot, Colchester, VT), to measure even a small number of points on the participant's scalp typically requires a few minutes while a participant remains completely motionless. Even with highly skilled developmental cognitive neuroscientists, early developmental populations (and many clinical populations) cannot meet this requirement and, thus, there are no published studies using a 3D digitizer for probe location estimation in early developmental populations. Furthermore, measurements with a 3D digitizer are often confounded by interference from metal objects in the participant's surroundings. Thus, even in populations that can comply with these methods, it is difficult to move this method between experimental contexts. This is problematic as one of the major benefits of scalp-based neuroimaging is their portability (e.g., recording in rural Africa, Lloyd-Fox et al., 2014a). The 3D digitizer its self can cause interference with sensitive medical devices, such as cochlear implants, which make it non suitable for this clinical population. New photogrammetry methods also rely on tightly restricted movements of the subject and allow only a specific movement trajectory of the camera around the subject (Clausner et al., 2017; Stopczynski et al., 2014). Other photography-based methods require expertise in manual annotation of images (e.g. in Emberson et al., 2015; Lloyd-Fox et al., 2014b) which are extremely time-intensive and are prone to between-researcher variability. In these methods, several photos of the infant wearing the fNIRS cap are taken from a few different angles. The photos are then manually scaled according to a known size marker on the cap. The researcher then needs to measure a few distances in the photos (e.g., nasion to inion, right ear to left ear) and to estimate the distances of the fNIRS channels from predefined fiducials. Our proposed video-based method takes the basic principles of this manual annotation methods and extends it to an automatic video-based method.

In this paper, we present a novel video-based method that is both easy to implement by novice experimenters and is robust to participant's head movements. Our method requires only ~20 seconds of video using widely-available photographic equipment, recorded around the participant's head, with the probes already mounted on the scalp. During acquisition, the participant can move his or her head freely without jeopardizing the accuracy of the measurement. We report both the validity and the reliability of our video-based method compared to the traditional 3D digitizer on a group of adult participants. We find a high level of agreement between methods (inter-method validity), and within the method in a test-retest repeated measurements configuration (intra-method reliability). Importantly, our video-based method exhibits similar intra-method reliability compared to the traditional 3D digitizer. Additionally, we estimate the session-to-session variability in cap placement using the video-based and the digitizer methods and find good agreement between the two. Finally, we demonstrate the successful measurement of infants' head mounted probe positions: An estimation which is not possible using currently available online methods (e.g., 3D digitizers).

Methods

Our automatic video-based method includes four steps: First, the participant's head, and mounted cap, are captured through a short video, using an off-the-shelf mid to high-end camera (e.g., GoPro, GoPro, San Mateo, CA). Then, for each frame in the video, we crop away everything but the cap. This cropping facilitates the handling of head motion, since it focuses the next steps solely on the cap and completely eliminates background distractions. Next, we reconstruct a 3D model of the head from the cropped images using computer vision techniques. Lastly, we extract the coordinates of specific fiducials - an intuitive identification task once we already have a reconstructed model. In a pre-processing step, we captured the cap in perfect conditions – well aligned on a plastic head – and manually marked the positions of all fiducials and probes, to create our cap base (model cap). The model cap is then used to interpolate the probe positions based on the fiducials. This pre-registration step is required for each new configuration of probe positions (i.e., for new experiments or populations).

Participant preparation

The basic requirement of the method is good tracking of the head with the cap in each video frame. To improve it, we placed multiple two-color patterned plastic sheets (red and blue in the example in Fig. 1B. A script for generating this specific pattern is in the supplementary, but any distinct two-color pattern is expected to suffice) on the fNIRS cap and nine solid color stickers (green in Fig. 1B) at specific fiducials on the participant's head: nasion (Nz), left tragus (AL), right tragus AR, and on the cap: around the vertex (Cz), around the inion (Iz), between them (Pz), on the left edge of the cap, on the right edge of the cap and on the front edge of the cap. See Figure 1B for cap preparation. The overall purpose of the method is to find the spatial relation between the three points on the head and the six points on the cap. Finding this relation will enable exact positioning of the cap on the participant's head. In order to ensure our cropping method is not confused by other objects in the scene, we cleared the surrounding of the participants from any similarly colored objects (e.g., no toys with similar colors, if the parent's shirt had similar colors, we covered it with a dark barber cape). The patterned two-colors plastic sheets are used to crop the subject's head out of each frame to reduce unnecessary information for the Visual SfM algorithm (see below). The stickers are used to identify the fiducial positions after the 3D model is built. The vast majority of these preparatory steps were completed before the participant's arrival with the exception of minimal sticker placement on the infant when the cap has been situated.

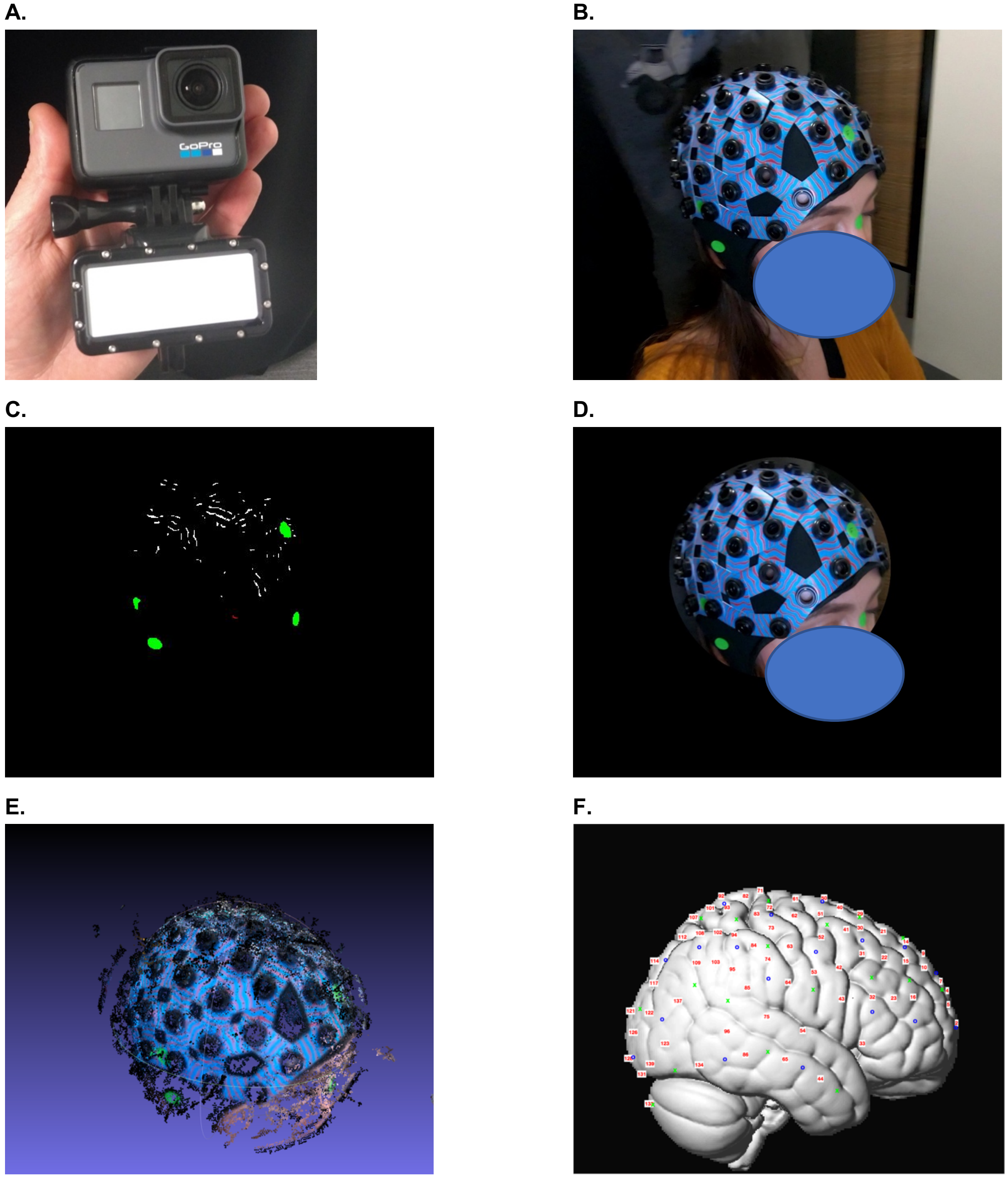


Figure 1. Participant preparation, video capturing and frame cropping. **A.** The short video was captured using an off-the-shelf camera with diffused flashlight. **B.** fNIRS cap was assembled with two-color pattern sheets and solid-color stickers were placed in the fiducial positions, which are to be estimated. **C.** The two-color patterns and the solid color stickers were automatically identified from each frame. **D.** A circle was cropped from each frame to include only the participant's head and cap. **E.** All cropped frames were combined to a 3D model. **F.** Fiducials were extracted from the model and channel positions were projected onto MNI space.

Video capture

We recorded the cap placement on the subject's head using a GoPro Hero 6 Black twice for each subject. We used the slow-motion (240 fps) feature to reduce motion blurring. Any camera with a similar slow-motion option will suffice for the requirements of our method (e.g., iPhone 8). We attached a diffused flashlight to the camera to compensate for uneven lighting conditions across all head sides (see Fig. 1A and Video 1 in supplementary). To allow us to connect each frame with its preceding and following frames without having edges effects and drifts in the Visual SfM algorithm (see below), we started and ended the video recording at approximately the same position relative to the subject's head. An example for a video recording which was used for this reconstruction is available at: https://youtu.be/QPpqoIxAPWs.

Cropping the video frames

In order to avoid passing irrelevant information to the 3D surface reconstruction stage, we cropped the head and cap from each video frame. First, the user manually picked, from a single frame, the three colors (red and blue for the cap, and green for the fiducial stickers) that characterized the cap. Then, the following cropping process was applied to each frame from the video automatically. The frames were white-balance corrected and all following image processing were conducted using OpenCV (Bradski, 2000) package for Matlab (https://github.com/kyamagu/mexopencv). The two-color patterned plastic sheets were automatically identified by dilating one color and finding the conjunction of it with the other color. This step identified the borders between the two colors as unique feature for identifying the two-color patterns (Fig. 1C). A circle was fitted to include all the identified two-color patterns (with a small additional margin) and the outside of the circle was cropped out of the frame (Fig. 1D). If solid colored stickers were identified in the frame, the circle was expanded to include these as well. Badly cropped frames, where there was more than 20% background in the cropped circle were automatically rejected from further processing.

3D surface reconstruction

In order to find the 3D coordinates of the fiducial points, a 3D model was automatically reconstructed. The cropped frames were passed to a Structure from Motion (Wu, 2013) software (Visual SfM; Wu, 2011). The software searched for shared image features between frames ("matches"; Lowe, 1999) which were less than two seconds apart. In addition, since the videos were taken such that they started and ended the same way, matches were also searched between the first and last frames of the video. This closed-loop helped in reducing accumulated drift, and in cases where matches tracking was lost mid-way through the video. The algorithm then placed the camera for each frame in the most probable location in space. A 3D model was then built using the matches and camera positions (Fig. 1E).

Fiducial extraction

Our basic assumption in this part is that the cap undergoes limited deformation when worn by the patient. Specifically, we assume the cap can only rotate globally with respect to the head, and scale along the anterior-to-posterior and left-to-right axes. Therefore, capturing five points on the cap are sufficient for complete cap-head registration. In practice, we used 6 points for redundancy, with two in the difficult-to-capture back region ("Pz" and "Iz"), because the infant was sitting on their parent's laps.

Later, in the reconstructed 3D model, we identified the position of the fiducials by grouping nearby reconstructed points which hold the solid sticker color and matching them with the expected relative position of the fiducials. We then used the head fiducials (Nz, AR, AL) to fit the cap points ("Cz", "Pz", "Iz", Front, Left, Right) into MNI space by rotation and scaling (Evans et al., 1992). The channel positions were then interpolated in MNI space based on the pre-registered cap model relative to the cap points using SPM for fNIRS (see first paragraph of the Methods, Tak et al., 2016; Fig. 1F).

3D digitizer measurement

We collected the 3D coordinates of the head points and cap points using a 3D digitizer (Fastrak, Polhemus, Colchester, VT) twice for each adult subject. We then projected the channel positions into MNI space using the same pre-defined base model and SPM for fNIRS.

Results

Video-based method, validated through the standard 3D digitizer

We measured the channel positions (139 channels; LABNIRS, Shimadzu inc., Kyoto, Japan) using our new video-based method and using the 3D digitizer (Fastrak, Polhemus, Colchester, VT) twice for each subject (N = 10) for the same session and compared the positions of the same channels both between the two methods (inter-methods validity) and within each one (intra-method reliability). We estimated the validity of our new video-based method by comparing the estimated positions of the same channels between the two methods of measurement. The intra-method comparison was used to estimate the error of each method independently (test-retest reliability).

We found that the validity of our new video-based method was comparable to the reliability of the field's standard method – the 3D digitizer (Fig 2A-B). Namely, the distances between the positions of the same channel as estimated by the two methods were similar to the distances between the position of the same channel as measured by the 3D digitizer twice (inter-methods validity of 4.6 $\pm$ 1.8 mm compared to intra-method reliability of 2.4 $\pm$ 0.5 mm; Mean $\pm$ STD). Additionally, the reliability of our video-based method was comparable to the reliability of the 3D digitizer. The distances between the estimated positions of the same channel as measured twice by our video-based method (4.1 $\pm$ 1.4 mm; Fig 2C) was similar to that of the 3D digitizer and also similar to the distances between the two methods. Overall, the magnitude of the errors that we found was smaller than half the size of the probes, whose positions were estimated (10 mm diameter).

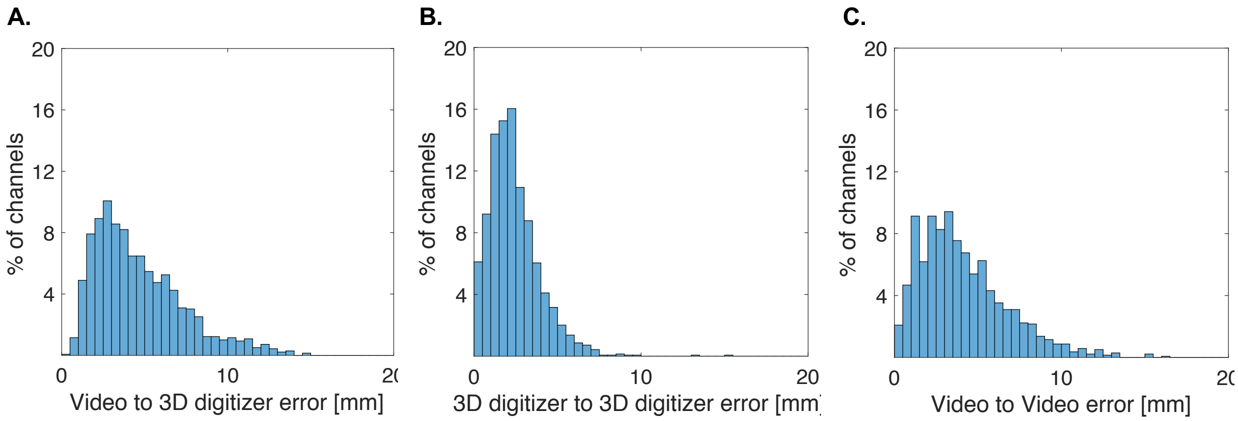
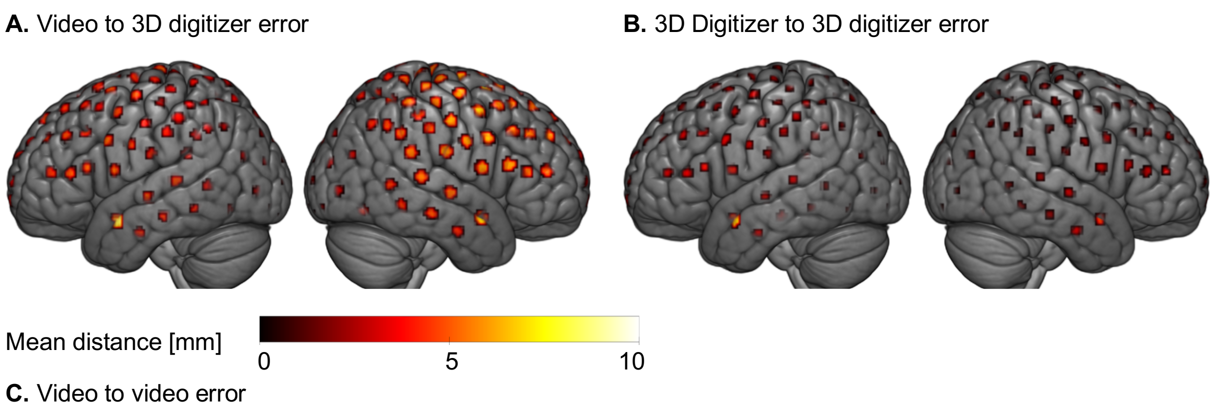


Figure 2. Individual distances between measurements of the same channels. **A.** 3D digitizer vs. Video (inter-method validation) **B.** 3D digitizer vs. 3D digitizer (intra-method reliability). **C.** Video vs. Video (intra-method reliability).

The discrepancies between the two position estimation methods were spread evenly through the scalp, with a tendency for larger distances in the anterior temporal channels (Fig. 3A). This tendency stems from smaller reliability of both methods in these areas (Fig. 3B-C).



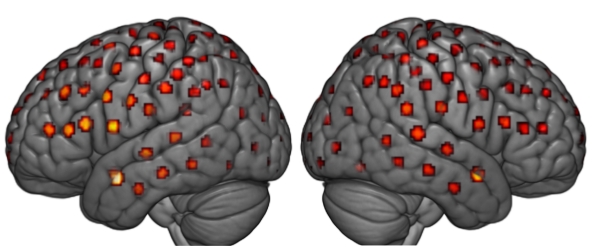


Figure 3. Spatial distribution of inter-method validity and intra-method reliability. **A.** Distance from video-based method to 3D digitizer method. **B.** Distance between two measurements of 3D digitizer. **C.** Distance between two measurements of video-based method. Distances are represented by the color and the diameter of the patches for each channel position.

Video-based method captures cap positioning variability

We measured the channel position for a separate group of adult subjects (N = 10) in two separate sessions to estimate whether the video-based method captured the shifts in cap positioning on the same subject and compared this measurement to the standard 3D digitizer. The same systems and methods as described above were used. Within each session, we found similar level of inter-methods validity as in the previous group of subjects (7.6 $\pm$ 4.1 mm), replicating the overall validity of our video-based method. We now considered how well these two localization methods identify any change in position of the cap (i.e., shift or displacement of each channel) across sessions. We found a significant regression coefficient for the size of the estimated shift in channel position between the two methods (Fig. 4A; $F_{1,1388} = 15.8$, $p < 10^{-5}$; Linear mixed effect model). Namely, the shift size as estimated by the video-base method predicts the shift size as estimated by the 3D digitizer method. We also found that the shifts in channel positions were estimated in similar directions using the two methods. The angle between the direction of the shifts, as estimated by the two methods, was small (closer to zero than a uniform distribution; Fig. 4B; $K = 0.11$, $p < 10^{-8}$; Kolmogorov-Smirnov test). The similarity in shift direction was found greater in the larger shifts, where this similarity matters most. Specifically, we examined the angle distribution when the sizes of the shifts from session to session were larger than the median in both estimation methods. Investigating these shifts in cap placement across sessions reveals how crucial it is for any co-registration to be accurate. Even if recordings take place same subject with the same cap mere days apart, these shifts in cap placement are large enough to put any given cortical region under different channels. We found that the angle between the estimated shifts when these shifts are large is even closer to zero (Fig. 4C; $K = 0.22$, $p < 10^{-9}$).

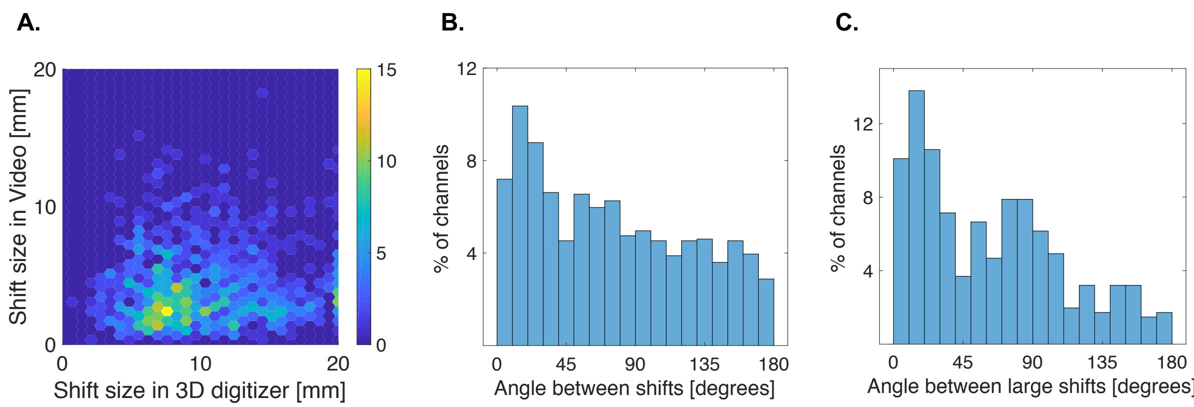


Figure 4. Comparison of cap shifting between session estimation through the video-based and the 3D digitizer methods. **A.** Estimation of shift size between the sessions of the video-based method vs the 3D digitizer one. The color of each hexagon represents the number of channels in its range. We found a high correlation between the shifts as estimated by the two methods. **B.** Angles between the shift as estimated by the video-based method and the shift as estimated by the 3D digitizer method. The distribution of angles in skewed towards zero. **C.** Angles between the shifts when the both estimated shifts are larger than the median. The distribution of angles for large shifts is skewed further towards zero than the distribution in B.

As Figure 4A clearly shows, there were a few cases where the two estimation methods did not agree on the shift size between the two sessions (e.g., a high shift size was estimated by the 3D digitizer and a small shift size was estimated by the video-based method for the points on the righthand side). We manually examined a few cases of large discrepancies between the shift estimation of the two methods. In most cases, the video-based method was actually closer to estimating the shift magnitude correctly. For example, in Figure 5, the shift between the sessions of channel 66 (circled in yellow) for participant 1 was estimated as 3.6 mm by the video-based method and 24.5 mm by the 3D digitizer. Our manual estimation, based on image-space measurements scaled by known distances between channels, suggests a shift of 7mm – a number much closer to the video-based estimation than to the 3D digitizer-based one (Fig. 5).

A. B.

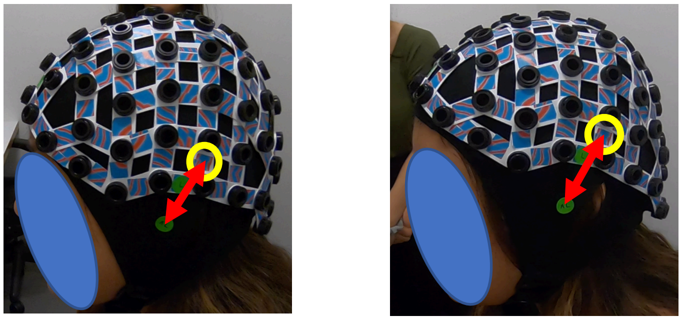


Figure 5. Manual estimation of a case of large discrepancy between the two estimation methods. **A.** Left view of participant 1 in session 1. The distance of channel 66 (circled in yellow) from the left tragus (lower green sticker) was 49.3 mm (the length of the red arrow). **B.** Left view of participant 1 in session 2. The distance between channel 66 and the left tragus was 53.3 mm. The shift of the channel position between the two sessions was closer to the video-based estimation (3.6 mm) than to the digitizer-based estimation (24.5 mm).

Automatic estimation of channel position in infants

We successfully reconstructed a 3D model from video recordings of infant participants and estimated the channel positions on their scalp through our automated video-based method (using methods described above, Fig. 6 for one sample infant). Thus, here we present a proof of concept that the same method can be used for early developmental populations. In practice, we have employed this method for many dozens of infants at different ages. Since there can be no comparison with a 3D digitizer, we focus this paper of the validity of the overall method in adult participants in comparison to established methods.

Discussion

We present an innovative approach for 3D localization of probes relative to participant's head, which is highly needed in scalp-based methods of neuroimaging such as EEG and fNIRS and particularly for developmental and clinical populations. The proposed method requires only a short video recording during the experimental session itself. Moreover, during this video, participants can move freely and interact with research staff and caregivers. These differences in methodology make this method suitable for developmental and clinical populations. Current methods have major disadvantages: They are either susceptible to participant's movements and thus are not useable for developmental populations (3D digitizer) or require laborious and bias-prone manual annotation (manual co-registration). By contrast, the video-based method we present here is both motion-resilient and automatic. Additionally, the video-based method does not require constant power supply and is not sensitive to metal in the environment and is thus applicable in out-of-lab settings and may be widely used in global neuroscience projects. The automatic analysis of the videos does not require the extensive manual annotation of images and thus less prone to bias and much less laborious.

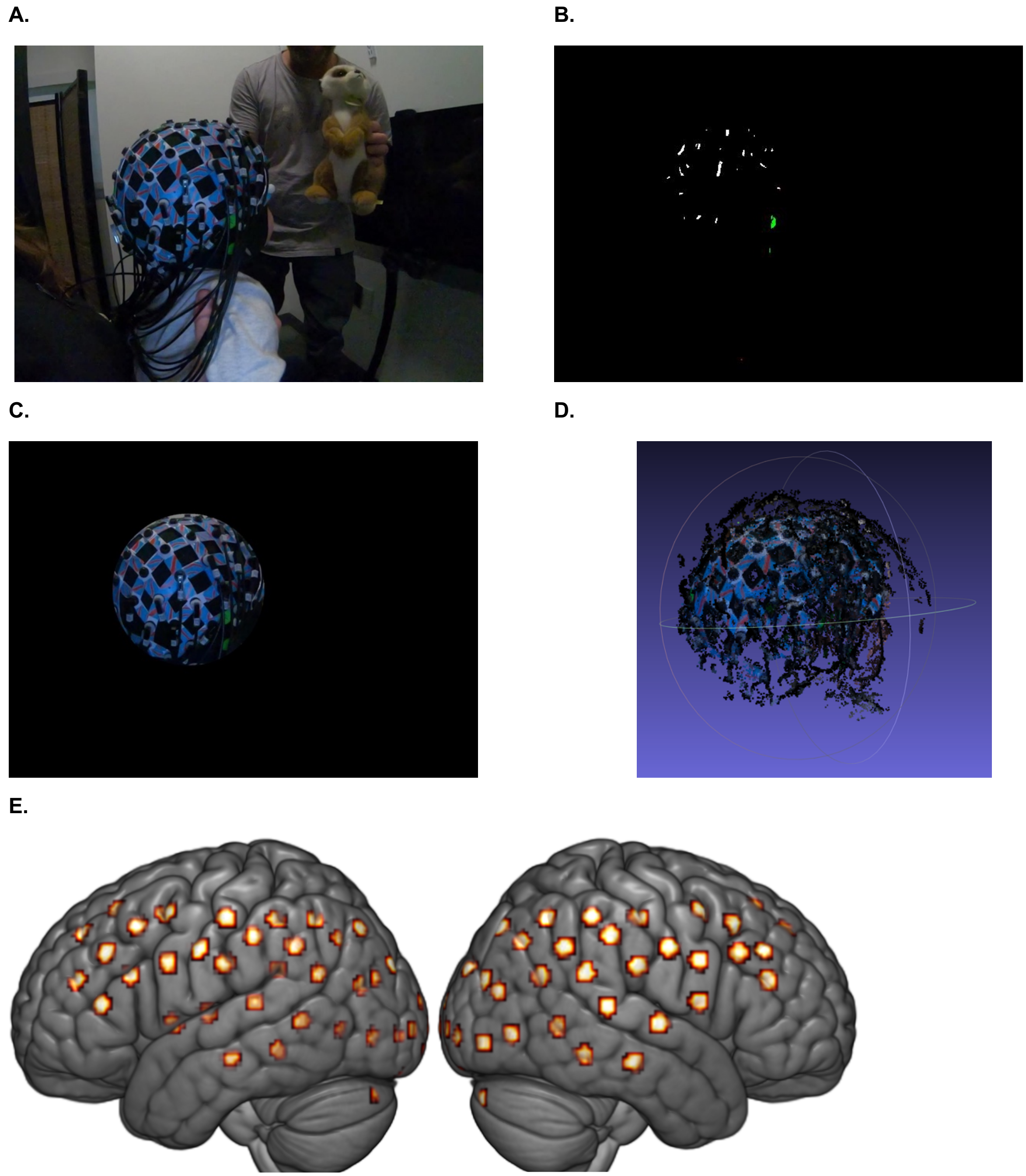


Figure 6. Infant channel position estimation using the video-based method. **A.** The captured video. **B.** Two-color patterns and fiducial stickers were isolated from the images. **C.** The images were cropped by a circle the fits all the valid information, to eliminate confounds of movements. **D.** A 3D model was reconstructed using Visual Structure from Motion. Coordinates of fiducial points were extracted from the model. **E.** Channel positions were interpolated from fiducial coordinates onto MNI space.

We present evidence that the proposed method is both valid and reliable compared to the standard method currently employed in the field, which uses a 3D digitizer. Overall, we find strong agreement between these methods of localization. Where these two methods had the lowest agreement, in the anterior temporal region, we suspect that it stemmed from difficulty in placing the digitizer in the correct angle, perpendicular to the scalp. We further tested the localization abilities of these methods by examining changes in localization of channels between sessions with the same subject. We find that these methods similarly estimate the displacement of the cap. Where there are disagreements between the methods, we have found that the video localization method can be more accurate in channel position estimation.

Future implementation of the proposed video-based method will include a user interface that will allow users to supply feedback on the quality of the spatial estimation of the channel positions. This feedback can be then uploaded to a repository, where a machine learning architecture will learn the parameters that would optimize the algorithm that was described in the methods.

Using a different model cap, as described at the beginning of the method section, researchers can easily adapt the outline that is proposed here to their own system and to any cap type. Our proposed co-registration method will allow combination of data from different locations and labs and will enable the creation of a unified anatomical framework for fNIRS analysis.

Acknowledgements

This research was supported by grants from the National Institutes of Health (R00 4R00HD076166-02), McDonnell Foundation (220020505).

References

Aslin, R.N., Shukla, M., Emberson, L.L., 2015. Hemodynamic Correlates of Cognition in Human Infants. Annu. Rev. Psychol. 66, 349–379. https://doi.org/10.1146/annurev-psych-010213-115108

Bradski, G., 2000. OpenCV.

Clausner, T., Dalal, S.S., Crespo-García, M., 2017. Photogrammetry-based head digitization for rapid and accurate localization of EEG electrodes and MEG fiducial markers using a single digital SLR camera. Front. Neurosci. 11, 1–12. https://doi.org/10.3389/fnins.2017.00264

Emberson, L.L., Richards, J.E., Aslin, R.N., 2015. Top-down modulation in the infant brain: Learning-induced expectations rapidly affect the sensory cortex at 6 months. Proc. Natl. Acad. Sci. 112, 9585–9590. https://doi.org/10.1073/pnas.1510343112

Evans, A.C., Marrett, S., Neelin, P., Collins, L., Worsley, K., Dai, W., Milot, S., Meyer, E., Bub, D., 1992. Anatomical mapping of functional activation in stereotactic coordinate space. Neuroimage 1, 43–53. https://doi.org/10.1016/1053-8119(92)90006-9

Ferradal, S.L., Eggebrecht, A.T., Hassanpour, M., Snyder, A.Z., Culver, J.P., 2014. Atlas-based head modeling and spatial normalization for high-density diffuse optical tomography: In vivo validation against fMRI. Neuroimage 85, 117–126. https://doi.org/10.1016/j.neuroimage.2013.03.069

Jurcak, V., Tsuzuki, D., Dan, I., 2007. 10/20, 10/10, and 10/5 systems revisited: Their validity as relative head-surface-based positioning systems. Neuroimage 34, 1600–1611. https://doi.org/10.1016/j.neuroimage.2006.09.024

Lloyd-Fox, S., Papademetriou, M., Darboe, M.K., Everdell, N.L., Wegmuller, R., Prentice, A.M., Moore, S.E., Elwell, C.E., 2014a. Functional near infrared spectroscopy (fNIRS) to assess cognitive function in infants in rural Africa. Sci. Rep. 4, 4740. https://doi.org/10.1038/srep04740

Lloyd-Fox, S., Richards, J.E., Blasi, A., Murphy, D.G.M., Elwell, C.E., Johnson, M.H., 2014b. Coregistering functional near-infrared spectroscopy with underlying cortical areas in infants. Neurophotonics 1, 025006. https://doi.org/10.1117/1.NPh.1.2.025006

Lowe, D.G., 1999. Object Recognition from Local Scale-Invariant Features, in: International Conferenceon Computer Vision. https://doi.org/10.1109/ICCV.1999.790410

Stopczynski, A., Stahlhut, C., Larsen, J.E., Petersen, M.K., Hansen, L.K., 2014. The smartphone brain scanner: A portable real-time neuroimaging system. PLoS One 9. https://doi.org/10.1371/journal.pone.0086733

Tak, S., Uga, M., Flandin, G., Dan, I., Penny, W.D., 2016. Sensor space group analysis for fNIRS data. J. Neurosci. Methods 264, 103–112. https://doi.org/10.1016/j.jneumeth.2016.03.003

Wu, C., 2013. Towards Linear-Time Incremental Structure from Motion, in: 2013 International Conference on 3D Vision - 3DV 2013. pp. 127–134. https://doi.org/10.1109/3DV.2013.25

Wu, C., 2011. VisualSFM : A Visual Structure from Motion System [WWW Document]. URL http://ccwu.me/vsfm/