FaceSync: Open source framework for recording facial expressions with head-mounted cameras

Jin Hyun Cheong[1*], Sawyer Brooks[2], & Luke J. Chang[1]

[1]Department of Psychological and Brain Sciences
Dartmouth College
Hanover, NH

[2]Department of Neuroscience
Oberlin College
Oberlin, Ohio

[*]Corresponding Author (jin.hyun.cheong.gr@dartmouth.edu)

Word Count: 4,293

Abstract

Advances in computer vision and machine learning algorithms have enabled researchers to extract facial expression data from face video recordings with greater ease and speed than standard manual coding methods, which has led to a dramatic increase in the pace of facial expression research. However, there are many limitations in recording facial expressions in laboratory settings. Conventional video recording setups using webcams, tripod-mounted cameras, or pan-tilt-zoom cameras require making compromises between cost, reliability, and flexibility. As an alternative, we propose the use of a mobile head-mounted camera that can be easily constructed from our open-source instructions and blueprints at a fraction of the cost of conventional setups. The head-mounted camera framework is supported by the open source Python toolbox FaceSync, which provides an automated method for synchronizing videos. We provide four proof-of-concept studies demonstrating the benefits of this recording system in reliably measuring and analyzing facial expressions in diverse experimental setups including group interaction experiments.

Introduction

Facial expressions provide rich information about how a person is feeling, what they are thinking, and how they might act (Russell & Fernández-Dols, 1997). Facial expressions captured the interest of early theorists (Darwin, 1965; James, 1884) and remain a popular method for noninvasively studying behavioral displays of emotions. Pioneering work by Paul Ekman established the Facial Action Coding System (FACS; Ekman & Oster, 1979), which provided a reliable coding system of different facial muscles referred to as action units (AUs) and allowed facial expressions to be compared across people and cultures (Matsumoto & Ekman, 1989; Matsumoto, Willingham, & Olide, 2009).

Unfortunately, FACS coding is highly laborious and time-consuming when trying to annotate facial expression dynamics as each video frame must be manually coded. Becoming a certified FACS coder requires over 100 hours of training ("Paul Ekman Group," n.d.) and a well-trained coder can require over 100 minutes to code a single minute of video (Cohn, Ambadar, & Ekman, 2007). Automated approaches can substantially reduce the amount of time required to extract facial expression information. One technique is to continuously record measurements of subtle facial muscle movements using electromyographic (EMG) to measure muscle groups associated with facial AUs, such as the M. zygomaticus and M. corrugator supercilii. However, this technique requires a separate electrode for each facial muscle group, which means that in practice only a few muscle groups can be recorded simultaneously (Fridlund & Cacioppo, 1986; Wolf, 2015). Moreover, this technique does not scale well to recording multiple participants interacting in a social experiment as recordings can be sensitive to movement artifacts and having wires attached to one's face can be unnatural and obtrusive.

Automated extraction of facial expression information from face video recordings have emerged as a promising alternative that combines the benefits of continuous measurements of many action units and complex facial expressions without any manual coding. Advances in computer vision and machine-learning techniques (e.g., support vector machines, convolutional neural nets) and large-scale data collection have facilitated the development   of models that learn to transform pixels from videos into predicted facial AUs and emotional facial expressions (Amos, Ludwiczuk, & Satyanarayanan, 2016; Littlewort et al., 2011; Michel & Kaliouby, 2003; Susskind, Littlewort, Bartlett, Movellan, & Anderson, 2007). Consequently, this has facilitated an explosion in scientific articles related to facial expression analysis with a sixfold increase over the past decade.[1]

This automated approach has offered much insight into human behavior. Automated extraction of facial expressions has been used to predict a wide range of behaviors including task engagement (Whitehill, Serpell, Lin, Foster, & Movellan, 2014), automobile accidents (Ahn, Bailenson, & Fox, 2010), effectiveness of advertisements (McDuff, Kaliouby, Cohn, & Picard, 2015), and online purchase behaviors (Ahn, Jabon, & Bailenson, 2008). Cultural differences in facial behavior have also been examined at a larger scale spanning more than 31 countries

(McDuff, Girard, & Kaliouby, 2017) as well as sex differences in smiling (McDuff, Kodra, Kaliouby, & LaFrance, 2017). Facial expressions have also shown promise in clinical settings to quantify symptom severity in neuropsychiatric disorders such as Schizophrenia and Parkinson's disease (Bandini et al., 2017; Hamm, Kohler, Gur, & Verma, 2011) and depression (Girard, Cohn, Jeni, Sayette, & De la Torre, 2015), and also for detecting evidence of malingering pain symptoms (Bartlett, Littlewort, Frank, & Lee, 2014).

The *acquisition* of high temporal and spatial resolution of facial expressions in laboratory environments, however, has remained challenging. Popular solutions such as webcams, tripod-mounted cameras, and pan-tilt-zoom (PTZ) cameras (Figure 1A-C) require compromising between cost, flexibility, and reliability. In this article, we demonstrate the feasibility of head-mounted video cameras as an alternative to standard recording setups. We provide step-by-step instructions on how to build affordable head mounts using readily available materials and minimal technical expertise. We demonstrate how the head-mounted camera can provide a reliable recording that is invariant to head-rotation and can be flexibly used in a variety of experimental settings such as stimulus based tasks, natural viewing of videos, and social interactions. We also introduce the FaceSync toolbox which can be used in conjunction with head-mounted cameras to automatically synchronize videos in time based on audio. Overall, we provide a unique solution for recording facial expressions that is affordable, adaptable to different experimental setups, and reliable in recording an unobstructed view of the face.

Figure 1. Images of different recording setups.

A. Webcam B. Tripod-mounted camera C. PTZ camera

D. Head-mounted camera E. Head-mounted camera in experiment

Figure 1. (A). Webcam setup. (B). Tripod-mounted camera setup. (C). PTZ camera setup in a dedicated room. (D). Head-mounted camera. (E). Use of head-mounted camera in social interaction experiment.

Selecting a recording method

When choosing a framework for recording videos of facial behavior in lab settings, researchers must consider affordability, adaptability to different experimental settings, integration with other devices, and recording reliability. In this section we survey and summarize the strengths and shortcomings of popular setups including webcams, tripod-mounted cameras, and PTZ cameras (Table 1).

The most readily available and easy-to-implement option is to record from computer webcams. Webcams with good image resolution cost about $100 and are pre-installed on most modern laptops and computers. Integrated webcams may be limited in temporal resolution as they rely on shared computing resources and use variable frame rates to optimize use of computer resources. Webcams are effective for event based experiments in which facial expressions can be recorded using the same computer hosting the task. Webcams can be triggered and controlled via programming languages providing a scalable solution for recording social interactions through video call setups. Such interactions, however, do not provide the same experience as live face-to-face interactions. Webcams, therefore, are a less robust solution as they cannot be adjusted to account for head rotation or out-of-plane head motion.

Tripod-mounted cameras cost around $1,000 for production-quality camcorders and can provide high-resolution recordings at faster frame rates. Tripod-mounted cameras can be manually moved or adjusted by the experimenter to account for subject movements if the experimenter can be present during the experiment. Tripod-mounted cameras can be easily installed and removed to accommodate different experimental settings allowing for flexibility in experimental setups. They can moved to different experimental environments and adjusted to different heights and angles for best view of facial behaviors. Scalability is limited as adding additional cameras remains expensive and synchronizing across multiple cameras can be challenging as time-code or TTL (transistor-transistor-logic) pulse triggering capable camcorders can be more expensive.

The PTZ camera setup provides researchers centralized control over cameras that can be rotated or zoomed to account for subject movement. PTZ camera setups requires a dedicated experiment room with cameras installed and an adjacent console room where experimenters can monitor incoming video feeds. Central management of cameras can facilitate integration with other softwares or triggering the cameras to record simultaneously. The installation of cameras to corners of ceilings distant from the participant allows cameras to be less conspicuous. However this forces participants to stand or sit in particular locations in the room and renders the setup particularly susceptible to head rotations or occlusions of the face from body gestures. The PTZ camera setup is thus the least flexible yet most expensive option because additional camera installations require additional construction.

Overall, webcams, tripod-mounted cameras, and PTZ cameras do not provide an optimal solution for recording facial expressions. They commonly suffer data loss due to out-of-plane head motions and head rotations (Cohn & Sayette, 2010; Lucey et al., 2011; Werner et al., 2013). Another common challenge pertains to the temporal precision in the temporal alignment of simultaneous recordings. In the next section, we propose head-mounted cameras as an affordable, scalable, and flexible solution that can provide reliable recordings and can be easily synchronized with experimental stimuli and across recordings.

Table 1. List of different video recording method

	Adaptability to different experimental settings	Ease of integration with other recording devices	Reliability to record without obstruction	Affordability
Webcams	Low	High	Low	High
Tripod-mounted cameras	Medium	Low	Low	Medium
PTZ cameras	Low	Low	Low	Low
Head-mounted cameras	High	High	High	High

Head-mounted cameras

A head-mounted camera recording system (Figure 1D) provides a unique solution to the limitations of the surveyed methods. It is a highly adaptive system that can be used for different experimental setups ranging from computer-based tasks to multi-person social interaction experiments (Figure 1E). The head-mounted camera consists of a single camera attached to the head of the participant using lightweight head-gear. This removes the impact of head rotation or body movement in obstructing the view of the face and eliminates out-of-plane head motion. It is minimally cumbersome as it can be positioned below the line of sight of subjects, allowing them to view the monitor in computer based tasks or make eye contact and track others' facial expressions in social interaction tasks.

Commercial head-mounted cameras are often used in motion capture studios and remain expensive with costs ranging from $2,500 - $20,000 for a complete camera and mount setup.[2] However, assembling a head-mounted camera setup in the lab can be an affordable alternative option that requires minimal engineering expertise or expensive equipment. Action cameras such as the GoPro are well-suited for this purpose as they are inexpensive ($150 - $400), small in size, and lightweight. We provide step-by-step assembly instructions for building a head-mount for GoPro cameras in the Supplemental Materials along with a parts list and blueprint files to 3D print other parts. This allows researchers to easily construct their own head-mounted camera setup for less than $700 (including the camera).[3]

Synchronizing videos to stimuli using audio features

All video recording devices require a method to temporally align the video recordings to the experimental task. As mentioned earlier, some devices such as webcams can be automatically triggered to start and stop recording during the experimental paradigm. Other methods, including head-mounted devices, require an accurate and efficient method for aligning the videos to experimental events after the experiment has ended.

The traditional 'clap' method used by the film industry uses a sharp, loud sound at the beginning of the clip that allows multiple videos to be aligned to the resulting spike in the audio waveform. This audio-based synchronization method usually requires opening each video for manual inspection of the sound waveform and incrementally shifting the audio until the echo, which indicates phase misalignment, is eliminated. Humans are highly accurate in detecting and distinguishing audio offsets down to several milliseconds (ms), but manually synchronizing each video is labor intensive and can introduce unsystematic noise in the alignment (Litovsky, Colburn, Yost, & Guzman, 1999; Shrstha, Barbieri, & Weda, 2007).

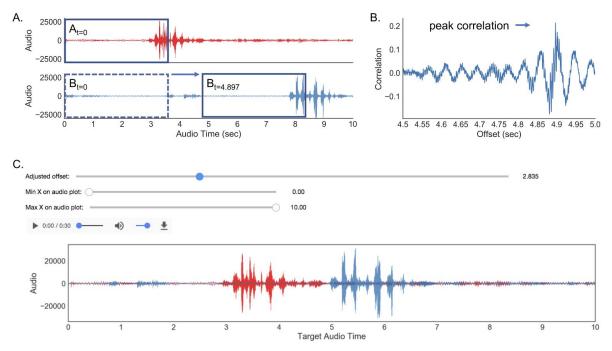## Figure 2. Schematic diagram of audio alignment

Figure 2. Schematic of automatic and manual audio alignment method. (A, B) Portion of audio is selected from target audio (red waveform top panel) which is compared to a portion of the sample audio selected in a sliding temporal window (blue waveform bottom panel). Correlation similarity is calculated at each window and the offset is determined to be the temporal window that maximizes similarity between the two audios. (C). Graphical interface in FaceSync toolbox to manually align two audios. One audio file can be shifted in time using the sliders until the two waveforms are aligned. Alignment can be inspected visually by examining the waveform plots and by listening to the combined audio.

To facilitate synchronization of videos, we developed FaceSync, an open-source Python toolbox, to automatically synchronize video recordings based on audio. In stimulus-based experiments, it requires a short audio segment to be played at the beginning of the experiment, which is also recorded by the camera. Based on this shared audio, the toolbox can align the video to the beginning of the experiment finding the optimal alignment with the original audio segment. In unstructured social interaction experiments, multiple videos can also be aligned to a target video. The toolbox offers a sliding window search to find the offset that maximizes the correlation between two audio signals (Figure 2) and a fourier transform based cross-correlation method. The FaceSync toolbox supports both Python versions 2 and 3 can be downloaded from our github repository (https://github.com/cosanlab/facesync).

Validation of head-mounted cameras

In the following sections, we demonstrate that the head-mounted camera recording system provides a robust way to record facial behavior in laboratory experiments. In Study 1, we verify that the head-mounted cameras can reliably record facial behavior invariant to head rotation. In Study 2-4 we demonstrate the flexibility of using head-mounted cameras in multiple experimental setups including an event-based paradigm (Study 2), naturalistic video watching paradigm (Study 3), and a social interaction paradigm (Study 4). In the three proof of concept experiments we also compare the performance of the FaceSync software in synchronizing recordings in comparison to the manual alignment method.

Reliability of face recordings using head-mounted cameras

Study 1: Face detection with head rotation.

Methods. To examine the impact of head rotation on face registration, we recorded the face of one male participant using a webcam and our head-mounted camera. In each recording session, the participant rotated the head 90 degrees left, returned to center, 90 degrees to the right, then returned to center. The head-mounted camera used a GoPro Hero4 camera to record 1280 x 720 resolution videos at 30 frames per second (fps). The webcam recording used the integrated camera on a Macbook Pro Retina laptop at 1080 x 720 resolution at approximately 30fps. Facial expressions were extracted using the iMotions Emotient FACET engine (*iMotions Biometric Research Platform 6.0*, 2016) which provides face registration success, landmark positions, AU predictions, and emotion predictions.

Results. Face recording using the head-mounted camera retained a continuous view of the entire face without face detection failure regardless of face rotation (Figure 3A top row). In contrast, the webcam face recording resulted in face detection failure when the head was rotated (Figure 3B right panel) which subsequently resulted in failure to extract facial expression predictions.

Face detection success was 100% of the video duration in the head-mounted camera recording compared to 75% in the webcam recording (Figure 3A, B middle row). Face expression prediction (e.g., neutral face, Figure 3A,B bottom row) in the webcam recording failed for 25% of the video while prediction from the head-mounted camera recording was uncompromised.

Figure 3. Face detection reliability between a head-mounted camera and webcam setup
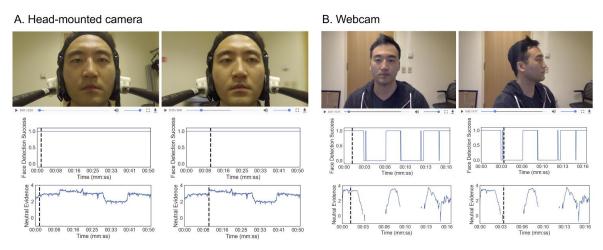

Figure 3. (A, B) Top row images show the face recordings when facing forward and the right panel image shows the face when facing left. Graphs in the middle row show face detection success at each frame. Graphs in the bottom row shows neutral face expression predictions. In (A), facial expressions are predicted in all frames in contrast to (B) where face detection and facial expressions predictions fail when the head is rotated away from the camera.

Discussion. Comparing face recordings from a webcam and a head-mounted camera, it is clear that the head-mounted camera provides a more reliable and continuous recording of the face invariant to head rotation. This is important as face expression software is unable to make predictions about facial expressions when it is unable to register a face. Though the head-mounted camera is invariant to head rotations, the head position is currently not tracked. Future work might add additional sensors to monitor head position dynamics. Overall, we demonstrate that the head-mounted camera can prevent data loss due to body and head rotations that can readily occur in most experimental settings without strict restriction of participants' natural movements.

Flexibility of head-mounted cameras across multiple experimental settings

Study 2: Recording facial expressions to event-based stimuli

This experiment demonstrates the use of head mounted cameras in recording facial expressions to time-locked stimulus presentations. Performance of automatic video alignment using FaceSync in comparison to manual adjustment is also provided.

Methods. One male participant viewed 10 positive images (e.g., kittens, puppies) and 10 negative images (e.g., injured bodies and faces) presented in MATLAB using Psychtoolbox

(Brainard, 1997; Pelli, 1997). Each image was presented for two seconds with jittered inter-trial intervals (ITI) of 4, 6, and 8 seconds (mean ITI = 5.4 seconds). The 20 images were selected from the IAPS picture database (Lang, Bradley, & Cuthbert, 2008). Facial behavior was recorded using a head-mounted GoPro Hero4 camera in 1,280 x 720 resolution at 30fps.

Audio offsets were determined both by manually synchronizing the recordings using the FaceSync AudioAlign graphical interface and the FaceSync automatic alignment function. The audio sample used for synchronization (synctune.wav) was a two-second harmonic tune constructed with sine waves at different frequencies. The audio extracted from the video was incrementally shifted by 1ms to the target audio using AudioAlign while listening to the shifted sound to minimize echo artifacts. For automatic alignment, we used the FaceSync sliding window correlation function (find_offset_corr) to detect the offset that maximizes the correlation similarity between the two audios. Videos were trimmed according to the calculated offsets and facial expressions were extracted using the iMotion Emotient FACET engine (*iMotions Biometric Research Platform 6.0*, 2016).

Results. An independent-samples t-test was conducted to compare positive facial expressions while viewing positive and negative images. Evidence of positive facial expressions while viewing positive images (M = 8.55, SD = .52) was significantly greater than the evidence while viewing negative images (M = 0.59, SD =1.53; t(18) = 15.58, p < 0.001; Figure 2C). Evidence for disgust facial expression while viewing negative images (M = 3.02, SD = 1.41) was significantly greater than the evidence while viewing positive images (M = .25, SD = .77, t(18) = 5.45, p < 0.001).

Audio alignment result is reported in Table 2. The difference between manual audio alignment and the automatic alignment was 3ms.

Figure 4. Different facial expressions for positive and negative images
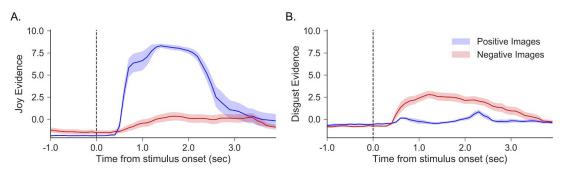

Figure 4. (A) Evidence of smiling to positive (blue) and negative (red) images. (B) Evidence of disgust facial expression to positive (blue) and negative (red) images.

Discussion. In Study 2, we demonstrate using head-mounted cameras to record facial expressions in response to time-locked stimuli. A participant displayed positive facial expressions (i.e. smiling) to positive images and disgust facial expressions to negative images

which was accurately retrieved from the analysis. Facial expressions were successfully linked to the stimuli that elicited the response by accurate alignment of the face recording to stimulus timing. Only a small difference was observed between the offsets determined automatically and manually. This study demonstrates the feasibility of using head-mounted camera setup for standard computer-based experiments and that facial expressions can be linked with the stimuli that elicited the response with high temporal precision.

Table 2. Audio alignment results for study 2 - 4

Sample Audio	Target Audio	Offset measured by manual adjustment	Offset measured by FaceSync algorithm	Difference between manual and algorithm alignment (sec)
Study 2				
Study2.wav	synctune.wav	7.913	7.916	-0.003
Study 3				
s01_w.wav	bigbunny.wav	5.816	5.807	-0.009
s02_w.wav	bigbunny.wav	5.399	5.387	-0.012
s03_w.wav	bigbunny.wav	6.363	6.372	0.009
s04_w.wav	bigbunny.wav	5.500	5.516	0.016
s05_w.wav	bigbunny.wav	11.833	11.833	0.000
Mean difference				0.001
Study 4				
s02_d.wav	s01_d.wav	4.900	4.897	-0.003
s03_d.wav	s01_d.wav	5.983	5.978	-0.005
s04_d.wav	s01_d.wav	5.658	5.660	0.002
s05_d.wav	s01_d.wav	4.508	4.504	-0.004
Mean difference				-0.003
Overall mean difference				-0.003

Study 3 and 4: Watching and discussing naturalistic stimuli together

In Study 3, we recorded individual facial expressions of a group while they watched a video together. In Study 4, we recorded facial expressions while the group discussed the contents of the video. We compare the facial expression behavior of each participant to one another with the hypothesis that participants would show synchronization of facial expressions in both conditions but to a greater extent in the movie watching experiment. Audio alignment offsets determined by manual and automatic alignment are compared.

Method. In Study 3, we measured the facial expressions of a group (N=5, 20% Female) watching a video, Big Buck Bunny ("Big Buck Bunny," n.d.). In Study 4, the group freely discussed the content of the video. Each person's facial behavior was recorded by their head-mounted GoPro Hero4 camera at 120fps and at 1,920 x 1,080 resolution.

In the Study 3, each face recording was aligned to the audio of the movie. In Study 4, each face recording was aligned to the audio of a single participant whose recording began the latest. In both cases, all cameras recorded the audio in the environment simultaneously, which allowed them to be aligned based on the shared audio. Each recording was aligned manually using the FaceSync AudioAlign graphical interface and automatically using the sliding window correlation alignment function (find_offset_corr). Differences in offset measured by the two methods were submitted to a one sample t-test to assess whether there was any difference between the two methods. After alignment, videos were trimmed using the FaceSync trim function.

Facial expressions, including AU activations and emotion predictions, were extracted using the iMotions Emotient FACET engine (*iMotions Biometric Research Platform 6.0*, 2016) then downsampled to 1hz. To assess the similarity of the affective experience, we calculated intersubject synchrony (ISC). This technique has been used in fMRI analysis to identify signals that are common across participants when watching naturalistic stimuli such as movies and listening to stories (Hasson, Nir, Levy, Fuhrmann, & Malach, 2004; Honey et al., 2012; Nummenmaa et al., 2012; Stephens, Silbert, & Hasson, 2010). For each experiment, ISC for joy facial expressions was calculated using pairwise correlation similarity of participants' predicted joy time-series. To determine whether the group was synchronizing in their smiling greater than chance, we calculated a one-sample t-test over all pairwise correlations to test whether the synchronization was significantly different from zero. In addition, we used a paired-sample t-test to assess whether the ISC was significantly different between the viewing (Study 3) and the discussion (Study 4). All t-tests were conducted on Fisher r-to-z transformed correlations.

Results. Overall, we found evidence that participants were having a similar affective experience while watching the video (Figure 5). Average ISC was r = .40 (SD=0.08), t(9) = 13.89, p<0.001. Participants also displayed synchronized facial expressions while discussing the video, r = .20, (SD=.16), t(9) = 3.75, p = 0.004. However, there appeared to be greater ISC of the joy facial expression viewing the show compared to discussing it afterwards, t(18) = 3.53, p = 0.002. This is likely because the joy facial expression was noisier while participants were talking and participants were not always in agreement with each other.

The average difference between the automated offset detection and the manual offset search was 1ms (SD = 11.82ms; absolute average offset M = 9ms, SD = 5.27ms; Table 2) for the movie watching videos and -3ms (SD = 3.11ms; absolute offset M = 4ms, SD = 1.12; Table 2) for the movie discussion session. Across all three experiments, differences in offsets calculated manually and automatically were not significantly different (t(9) = .58, p = .57).

Discussion. In these two studies, we observed a relatively high level of synchronization in affective experiences across participants while viewing and subsequently discussing a video. Speaking appears to decrease the sensitivity of ISC, likely as a result of the added noise from the mouth movements while speaking. The FaceSync toolbox can accurately align videos together even when the audio recordings are non uniform due to location of the camera position

and multiple people talking at the same time. Overall, these studies demonstrate the flexibility of using head-mounted cameras to record facial behavior in naturalistic experimental paradigms such as watching a movie and also in social experiments, in which participants interact with each other.

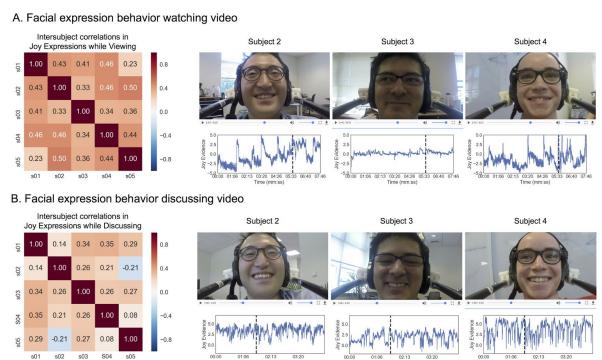**Figure 5. Facial expression behavior while watching and discussing video**

A. Facial expression behavior watching video

B. Facial expression behavior discussing video

Figure 5. (A) Pairwise intersubject similarity matrix for joy while watching the video is shown on the left panel. Three subject videos are shown on the right with corresponding joy evidence predictions. (B) Pairwise intersubject similarity matrix for joy while discussing the video is shown on the left panel.

General Discussion

In this paper, we provide evidence that head-mounted cameras offer a robust, flexible, and affordable solution to recording facial expressions in the laboratory. First, we demonstrate that the head-mounted camera yields *reliable* face recordings that allow facial expression analysis irrespective of head motion. Second, we demonstrate the flexibility of using head-mounted cameras across *different experimental settings* from traditional stimulus based experiments to group social interaction experiments. By using the FaceSync toolbox to align recordings to stimulus onsets, facial expressions were successfully linked to the events that triggered facial behavior, such as increased positive facial expressions in response to viewing positive images, and increased disgust facial expressions in response to viewing negative images. In addition, facial expressions were more synchronized when watching a video compared to when discussing the video. Most importantly, we demonstrate that the FaceSync toolbox can accurately and automatically align video recordings with comparable accuracy with manual

alignment. Together, these results demonstrate that the head-mounted camera setup offers a reliable and robust method for recording facial expressions in a variety of experimental settings and can scale to n-person social experiments.

The head-mounted camera setup can still be improved in several ways. For example, lighting conditions are important in face detection - poor luminance of the face or extreme backlights can lead to face detection failures. Researchers should be aware of this issue and should avoid situations where ceiling lights or window sunlights in the background decrease face luminance. LED lights can be attached to the head mount to control for these issues by providing equal luminance of the face.

Another potential improvement is the weight and size of the camera. The unique discomfort of the head mounted camera is the weight of the camera that pulls the headgear downward and the size of the camera that obstructs the view. At the time of construction, the camera (i.e., GoPro Hero4 Black with camera, memory card, and lens cover) weighed 92 grams but now more recent models (i.e., GoPro Hero5 Session) weigh only 74 grams. Even small reductions in weight can lead to greater comfort as the tugging force is significantly reduced based on the length of the headset. Researchers can take advantage of newer and lighter cameras as they become available as the 3D printable camera mount provided with the blueprints is compatible with other cameras.

In summary, we hope that these tools can benefit other researchers and further accelerate facial expression research. We anticipate that this framework will aid in improving our understanding of how emotions are experienced in naturalistic settings and how emotions surface and influence social interactions. In addition, we hope that these tools can aid researchers in developing new models of complex emotions such as regret, guilt, and gratitude. We look forward to a new era of facial expression research that examines facial behavior in both controlled and naturalistic settings to yield a robust and holistic understanding of how facial expressions reflect our thoughts, emotions, and future behavior.

Footnotes

1. Number of articles retrieved from Pubmed Central search with keyword 'facial expressions'.
2. Quote retrieved from http://facewaretech.com/pricing/ on July 2017.
3. Cost estimated at time when manuscript written on July 2017.

Acknowledgements

Author contributions

J.H.C., S.B., and L.J.C. wrote the manuscript. J.H.C., S.B., and L.J.C. designed the head-mounted camera and performed the experiments. J.H.C. developed the FaceSync toolbox.

Competing financial interests

The authors declare no competing financial interests.

References

Ahn, S. J., Bailenson, J., & Fox, J. (2010). 20 Using automated facial expression analysis for emotion and behavior prediction. *Emotions and Mass Media*. Retrieved from https://books.google.com/books?hl=en&lr=&id=jw16AgAAQBAJ&oi=fnd&pg=PA349&dq=Using+automated+facial+expression+analysis+emotion+behavior+prediction+Ahn+Bailenson&ots=cnbYkad9oP&sig=4AYiUN14l3kQM0Nm8L3SZuEZ6_g

Ahn, S. J., Jabon, M. E., & Bailenson, J. N. (2008). Facial expressions as predictors of online buying intention. In *Proceedings of the 58th Annual International Communication Association Conference* (pp. 22–26).

Amos, B., Ludwiczuk, B., & Satyanarayanan, M. (2016). Openface: A general-purpose face recognition library with mobile applications. *CMU School of Computer Science*. Retrieved from http://reports-archive.adm.cs.cmu.edu/anon/anon/usr0/ftp/2016/CMU-CS-16-118.pdf

Bandini, A., Orlandi, S., Escalante, H. J., Giovannelli, F., Cincotta, M., Reyes-Garcia, C. A., … Manfredi, C. (2017). Analysis of facial expressions in parkinson's disease through video-based automatic methods. *Journal of Neuroscience Methods*, *281*, 7–20.

Bartlett, M. S., Littlewort, G. C., Frank, M. G., & Lee, K. (2014). Automatic Decoding of Facial Movements Reveals Deceptive Pain Expressions. *CURBIO*, *24*(7), 738–743.

Big Buck Bunny. (n.d.). Retrieved October 27, 2017, from https://peach.blender.org/

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436.

Cohn, J. F., Ambadar, Z., & Ekman, P. (2007). Observer-based measurement of facial expression with the Facial Action Coding System. *The Handbook of Emotion Elicitation and Assessment*, 203–221.

Cohn, J. F., & Sayette, M. A. (2010). Spontaneous facial expression in a small group can be automatically measured: an initial demonstration. *Behavior Research Methods*, *42*(4), 1079–1086.

Darwin, C. (1965). *The Expression of the Emotions in Man and Animals*. University of Chicago Press.

Ekman, P., & Oster, H. (1979). Facial Expressions of Emotion. *Annual Review of Psychology*,

30(1), 527–554.

Fridlund, A. J., & Cacioppo, J. T. (1986). Guidelines for human electromyographic research. *Psychophysiology*, *23*(5), 567–589.

Girard, J. M., Cohn, J. F., Jeni, L. A., Sayette, M. A., & De la Torre, F. (2015). Spontaneous facial expression in unscripted social interactions can be measured automatically. *Behavior Research Methods*, *47*(4), 1136–1147.

Hamm, J., Kohler, C. G., Gur, R. C., & Verma, R. (2011). Automated Facial Action Coding System for dynamic analysis of facial expressions in neuropsychiatric disorders. *Journal of Neuroscience Methods*, *200*(2), 237–256.

Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., & Malach, R. (2004). Intersubject synchronization of cortical activity during natural vision. *Science*, *303*(5664), 1634–1640.

Honey, C. J., Thesen, T., Donner, T. H., Silbert, L. J., Carlson, C. E., Devinsky, O., … Hasson, U. (2012). Slow cortical dynamics and the accumulation of information over long timescales. *Neuron*, *76*(2), 423–434.

iMotions Biometric Research Platform 6.0. (2016). iMotions A/S, Copenhagen, Denmark.

James, W. (1884). What is an emotion? *Mind; a Quarterly Review of Psychology and Philosophy*, *9*(34), 188–205.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2008). *International affective picture system (IAPS): Affective ratings of pictures and instruction manual*. Retrieved from http://www.citeulike.org/group/13427/article/7208496

Litovsky, R. Y., Colburn, H. S., Yost, W. A., & Guzman, S. J. (1999). The precedence effect. *The Journal of the Acoustical Society of America*, *106*(4 Pt 1), 1633–1654.

Littlewort, G., Whitehill, J., Wu, T., Fasel, I., Frank, M., Movellan, J., & Bartlett, M. (2011). The computer expression recognition toolbox (CERT). In *Face and Gesture 2011* (pp. 298–305).

Lucey, P., Cohn, J. F., Matthews, I., Lucey, S., Sridharan, S., Howlett, J., & Prkachin, K. M. (2011). Automatically detecting pain in video through facial action units. *IEEE Transactions on Systems, Man, and Cybernetics. Part B, Cybernetics: A Publication of the IEEE Systems, Man, and Cybernetics Society*, *41*(3), 664–674.

Matsumoto, D., & Ekman, P. (1989). American-Japanese cultural differences in intensity ratings of facial expressions of emotion. *Motivation and Emotion*, *13*(2), 143–157.

Matsumoto, D., Willingham, B., & Olide, A. (2009). Sequential dynamics of culturally moderated facial expressions of emotion. *Psychological Science*, *20*(10), 1269–1275.

McDuff, D., Girard, J. M., & Kaliouby, R. E. (2017). Large-Scale Observational Evidence of Cross-Cultural Differences in Facial Behavior. *Journal of Nonverbal Behavior*, *41*(1), 1–19.

McDuff, D., Kaliouby, R. E., Cohn, J. F., & Picard, R. W. (2015). Predicting Ad Liking and Purchase Intent: Large-Scale Analysis of Facial Responses to Ads. *IEEE Transactions on Affective Computing*, *6*(3), 223–235.

McDuff, D., Kodra, E., Kaliouby, R. E., & LaFrance, M. (2017). A large-scale analysis of sex differences in facial expressions. *PloS One*, *12*(4), e0173942.

Michel, P., & Kaliouby, R. E. (2003). Real Time Facial Expression Recognition in Video Using Support Vector Machines. In *Proceedings of the 5th International Conference on Multimodal Interfaces* (pp. 258–264). New York, NY, USA: ACM.

Nummenmaa, L., Glerean, E., Viinikainen, M., Jääskeläinen, I. P., Hari, R., & Sams, M. (2012).

Emotions promote social interaction by synchronizing brain activity across individuals. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(24), 9599–9604.

Paul Ekman Group. (n.d.). Retrieved October 13, 2017, from https://www.paulekman.com/product-category/facs/

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442.

Russell, J. A., & Fernández-Dols, J. M. (1997). *The Psychology of Facial Expression*. Cambridge University Press.

Shrstha, P., Barbieri, M., & Weda, H. (2007). Synchronization of multi-camera video recordings based on audio. In *Proceedings of the 15th international conference on Multimedia - MULTIMEDIA '07*. https://doi.org/10.1145/1291233.1291367

Stephens, G. J., Silbert, L. J., & Hasson, U. (2010). Speaker–listener neural coupling underlies successful communication. *Proceedings of the National Academy of Sciences*, *107*(32), 14425–14430.

Susskind, J. M., Littlewort, G., Bartlett, M. S., Movellan, J., & Anderson, A. K. (2007). Human and computer recognition of facial expressions of emotion. *Neuropsychologia*, *45*(1), 152–162.

Werner, P., Al-Hamadi, A., Niese, R., Walter, S., Gruss, S., & Traue, H. C. (2013). Towards pain monitoring: Facial expression, head pose, a new database, an automatic system and remaining challenges. In *Proceedings of the British Machine Vision Conference* (pp. 119–111).

Whitehill, J., Serpell, Z., Lin, Y. C., Foster, A., & Movellan, J. R. (2014). The Faces of Engagement: Automatic Recognition of Student Engagement from Facial Expressions. *IEEE Transactions on Affective Computing*, *5*(1), 86–98.

Wolf, K. (2015). Measuring facial expression of emotion. *Dialogues in Clinical Neuroscience*, *17*(4), 457–462.

Supplementary Information

Here we present detailed instructions on how to build the FaceSync head-mounted recording system.

- Parts List
- Build Instructions

Parts list

Three primary components make up the construction: a helmet or headgear, supports to hold the camera, and a two-way sleeve joint connecting the supports to the helmet.

For the helmet, we found it easiest and most cost-effective to use a pre-existing helmet. Among many alternatives, this example uses wrestling headgear which is inexpensive, comfortable, easy to adjust, and lightweight. The selected models are detailed in Supplementary Table 1. We recommend having four and two-strap models available to accommodate subjects' different head shapes and preferences for comfort.

The support system should be lightweight but sturdy and sufficiently flexible to prevent excessive pressure on participants' heads. We used 100 Pounds per square inch (PSI) poly-pipe as it is rigid enough to support a camera while flexible enough to accommodate a range of head sizes without applying excessive pressure. Additionally, it is relatively easy to mount the supports for the camera and a counter weight on the back of the helmet. In our model, we used the GoPro Hero4 Black camera with its accompanying frame which was attached to the supporting pipes via a 3D printed mount.

The final component is a two-way sleeve joint used to connect the helmet to the support system. Using free online computer aided design (CAD) software (tinkercad.com) we created a two-way sleeve joint that connects the frame to the helmet. The joint is secured to the helmet using two bolts and a two-way coupling nut and to the frame via two nuts and two bolts that apply pressure to the piping holding the camera.

Full parts lists for the frame and the helmet, camera, frame, and sleeve joint are presented in Supplementary Tables 1 and 2. The assembled headset weighs approximately 760 grams excluding the camera.

Supplementary Table 1. Suggested Parts Purchase List - Camera and Helmet

Name	Manufacturer	Quantity	Price*
GoPro HERO4 Black Camera	GoPro	1	$449.00
64GB Memory Card	SanDisk	1	$47.28
GoPro-The Frame	GoPro	1	$38.95
Helmet	Cliff Keen (F5 Tornado, F3 Twister)	1	$32.95
Weighted Clamping Collar	Climax Metal	1	$6.80
Total Cost			$574.98

*All costs are estimated from Amazon listings at time of purchase

Supplementary Table 2. Suggested Parts Purchase List - Sleeve Joint and Frame

Name	Size	Quantity	Unit Price	Price*
PVC bars	61cm/2ft	2	$4.00	$8
	20cm/7.75in	2	$4.00	$8
PVC L joints		4	$1.50	$6
Hex-Head Screws	¼-20 0.75in/20mm	2	$.50	$1
	¼-20 1in/25mm	4	$.50	$2
	⅜-16 0.75in/20mm	2	$.50	$1
	M5-0.8 x 18mm Hex Screw	2	$.50	$1
Nuts	¼-20 Hex Nut	4	$.10	$.40
	M5-0.8 x 18mm Hex Nut	2	$.10	$.20
Coupling Nuts	¼-20 to ⅜-16	2	$2	$4
Hole Washers	0.75in/6.5mm	2	$.15	$.30
	0.33in/8.5mm	2	$.15	$.30
Total Price				$32.2

*All costs are estimated from Amazon listings at time of purchase

Supplementary Table 3. 3D Printing Parts List

Name	Quantity
Rotating Two-way Connector Sleeve	2
Cuff	2
GoPro Mount Clamp	1

*Total Price of 3D Printing : Approximately $90 https://www.shapeways.com/
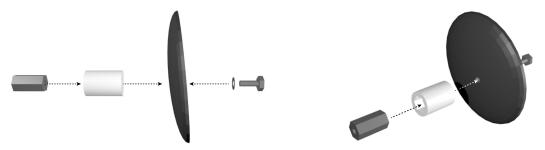*All parts are available for download at https://tinkercad.com/things/1z4V1HrdvFb

Build instructions

The ear portion of the helmet are represented as black disks and 3D printed materials are represented in white. Nuts and bolts are represented in gray.
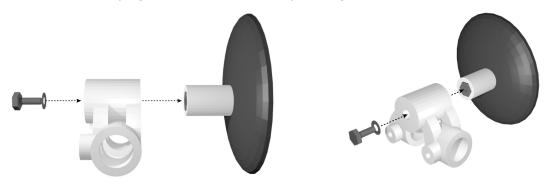
Recommended tools for assembly include a socket wrench with 11 and 14 mm heads, GoPro hex screwdriver, 20mm drill bit and drill.

Step 1: Drill a small hole in each of the helmet's earpieces. If using the F3 Twister headgear, drill directly into the logo; if using the F5 tornado, drill into the *second* air vent below the logo. The goal is to position the frame near the jawline, with the sleeve joint connecting to the helmet near the base of the ear.

Step 2: Place the coupling nut inside of the 3D printed cylindrical mold such that it will attach to the helmet with the *open end* facing away from the ear, with the wider ⅜ inch end facing inside towards the ear. Place the ⅜-16 20mm screw through the hole drilled in step one, separating the inside of the helmet and the screw with the 8.5mm washer; the screw should extend outside of the helmet. Use the socket wrench and the 14mm head to screw the coupling nut in the mold to the side of the helmet.
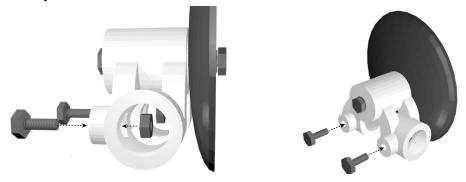

Step 3: Using the socket wrench and a ¼-20 20mm screw, secure the joint onto the the coupling nut with the 11mm head; use the 6.5 mm washer to separate the front of the joint and the screw. Wait to fully tighten so the helmet may be angled as appropriate.
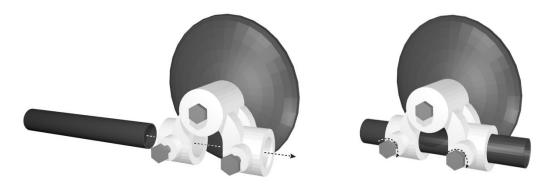

Step 4: Place one ¼-20 hex nut into each of the slots on the back of the two protruding screw holes. Using the 11mm head and the socket wrench, secure the two 1 inch ¼-20 screws into the

front of the protruding holes, so that they extend fully into the hex nuts without protruding into the open space beyond them.

Step 5: Thread one of the 61 cm PVC pipe through the two lateral holes of the sleeve joint. Roughly 18 cm of the pipe should extend out of the *back* of the sleeve joint. With the 11mm head, use the socket wrench to tighten the two screws inserted in step four to tighten the pipe in place. Repeat steps one through five for the other side of the helmet.

Step 6: Attach the 20cm PVC pipes to the front and back of the helmet using the four L-joints.
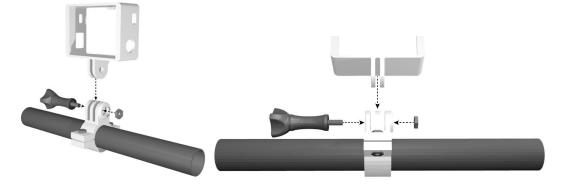

Step 7: Attach the joint for the GoPro frame to the front of the helmet by placing the two nuts into each slots and using the hex wrench to tighten the M5 hex screws into place. We

recommend lining the inside of the joint with electric tape to provide stability and traction for the mount. Repeat this step for the counterweight on the back of the helmet; only the hex screws are required.

Step 8: Fit the GoPro frame in between the three slots in the joint. Place the M5 nut into the slot on the side of the joint and attach the frame using the screw for the GoPro.

Final Result: