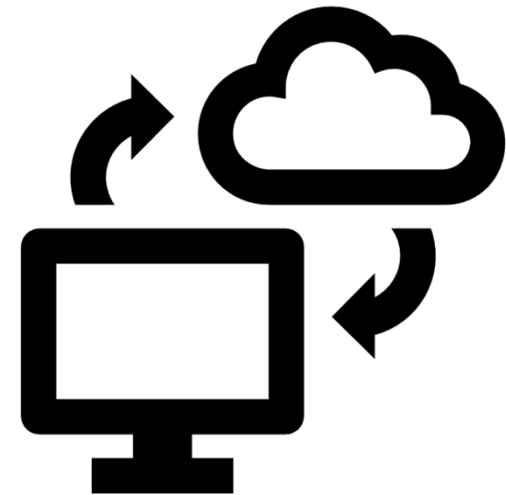


A web-based toolkit for unsupervised experiment with audio-visual data collection

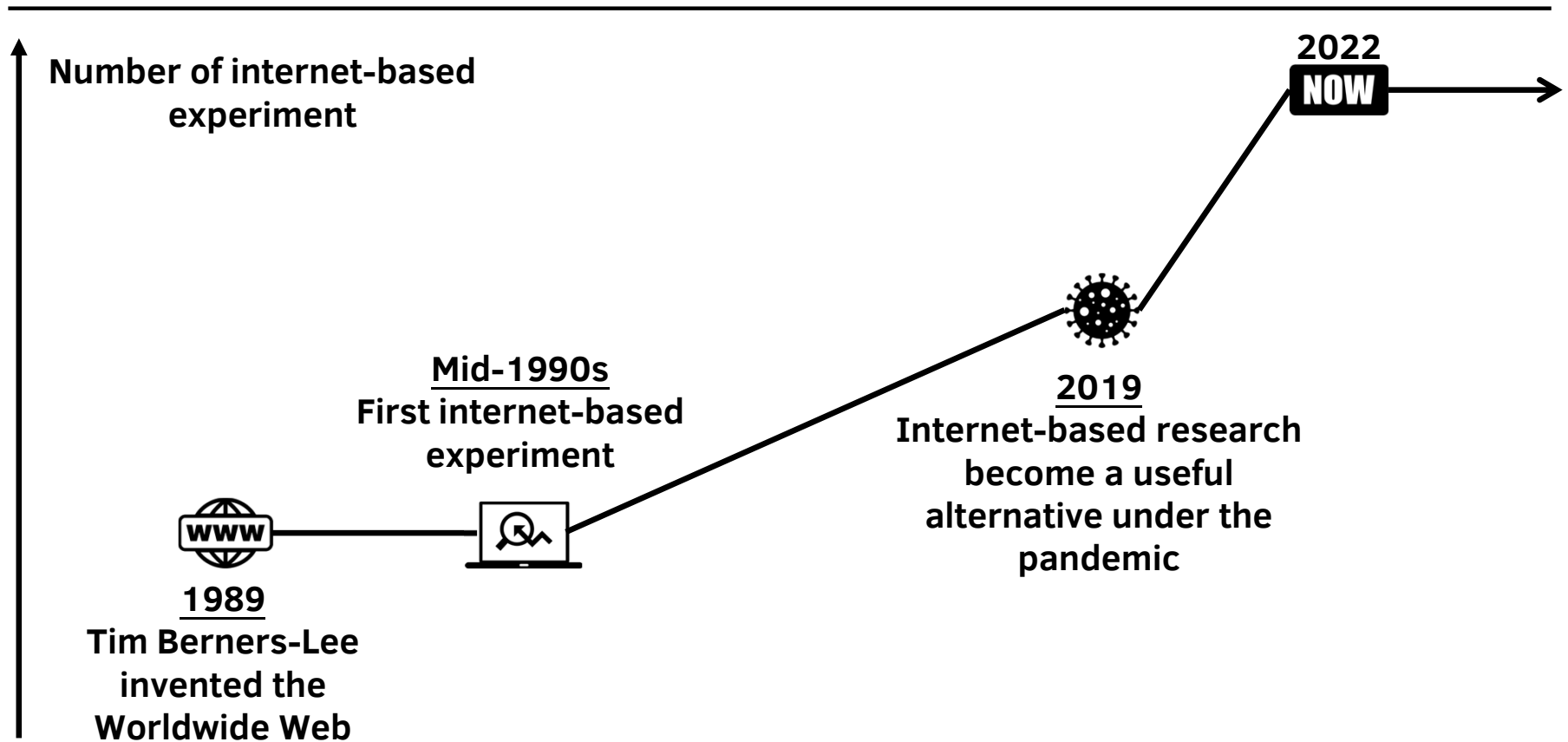
Jiahao Yang & Sotaro Kita

Department of Psychology, University of Warwick

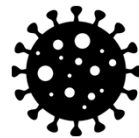
Jiahao.Yang@warwick.ac.uk



Online behavioral research



Online behavioral research - advantages



Impact of pandemic

Alternative for offline and face-to-face research.

Subject Population

Provide easy access to geographically unlimited subject population.



Internet-based experiment



Sample size

- High statistical power.
- Precise estimation of effect sizes.
- Detect small effects.
- Sufficient data to test computational models.

Cost reduction

Deliver experiments to participants. No laboratory rooms nor experimenters are needed.



Online behavioral research - concerns

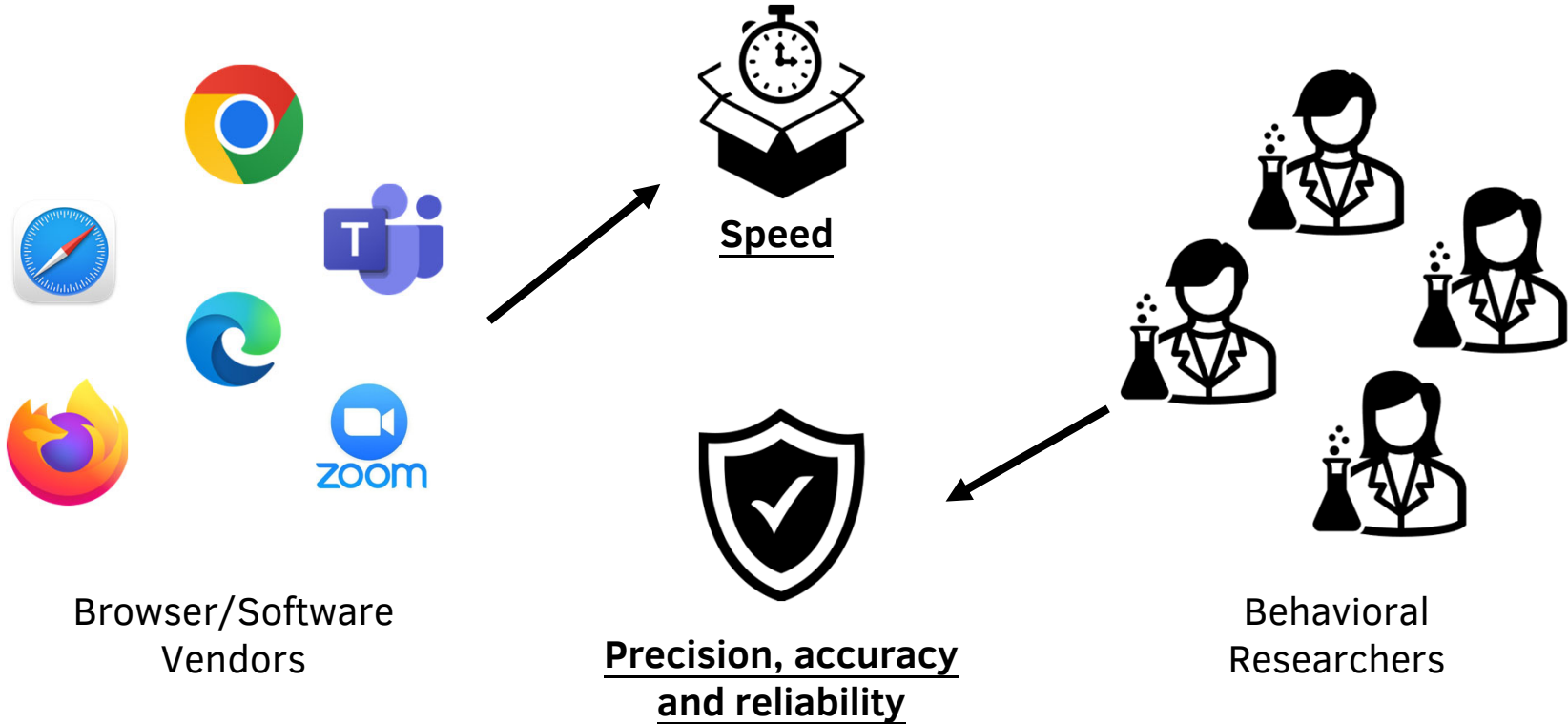


Temporal Precision

Whether browsers can present and record data accurately?

(e.g., Schmidt, 2007; Reips, 2007; Plant, 2016)

Online behavioral research - concerns



Online behavioral research – previous efforts



Efforts have been made to conduct accurate and reliably online behavioral research

Studies related to temporal precision concerns:

- Web technology benchmarks (van Steenbergen & Bocanegra, 2016).
- Methods/solutions for stimulus presentation/response time recording (Barnhoorn et al., 2015; de Leeuw & Motz, 2016; Garaizar & Reips, 2019).
- Evaluation of online platforms (Buhrmester et al., 2018; Crump et al., 2013; Garaizar & Vadillo, 2014).

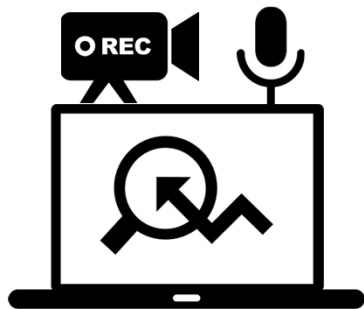


Tools/research-oriented frameworks to simplify development and improve the timing accuracy

- JsPsych (de Leeuw, 2015).
- Lab.js (Henniger, Mertens, Shevechenko, & Hilbig, 2017).
- Lookit (Scott & Schulz, 2017).
- Psychopy (Peirce, 2007).
- JATOS (Lange et al., 2015).



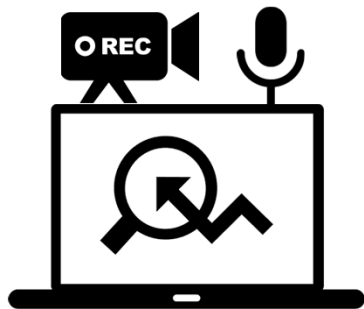
Online multi-modal research



Online multi-modal
research

Tools/Frameworks to support online multi-modal
research which requires collection of audio-visual data

Online multi-modal research

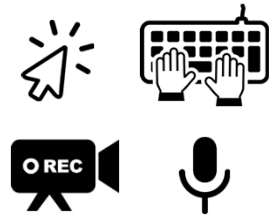


Online multi-modal research



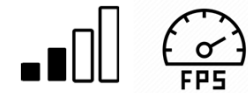
Data Collection

- Keyboard/pointing device
- Video/audio input device



Data Quality

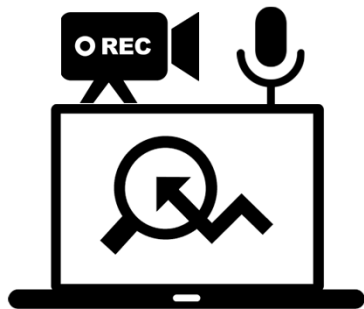
- Temporal precision.
- Framerate/Volume.
- FOV (Field of view): the area the webcam capture.
- Multiple angles.



Research Question

- Attend to gesture, gaze, and posture produced spontaneously
- Create an experimental environment close to the natural communicative environment
- Free the hands from the mouse and the keyboard

Online multi-modal research



Online multi-modal
research



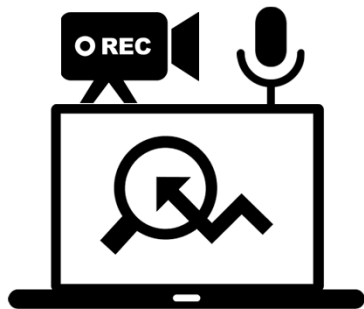
Video Chat Software

- Low frame rate (less than eight fps); Framerate will be influenced by internet connectivity and participants' geographical location.
- An experimenter is necessary.
- Low temporal precision for stimuli presentation and response collection

Online Platform

- Audio-visual data will be uploaded to a third-party platform.
- No mechanisms to determine the field of view of the camera.
- Uploading video files to the online platform may have data protection concerns.

Online multi-modal research



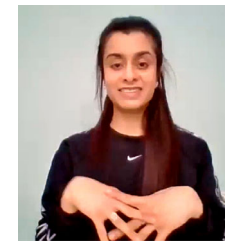
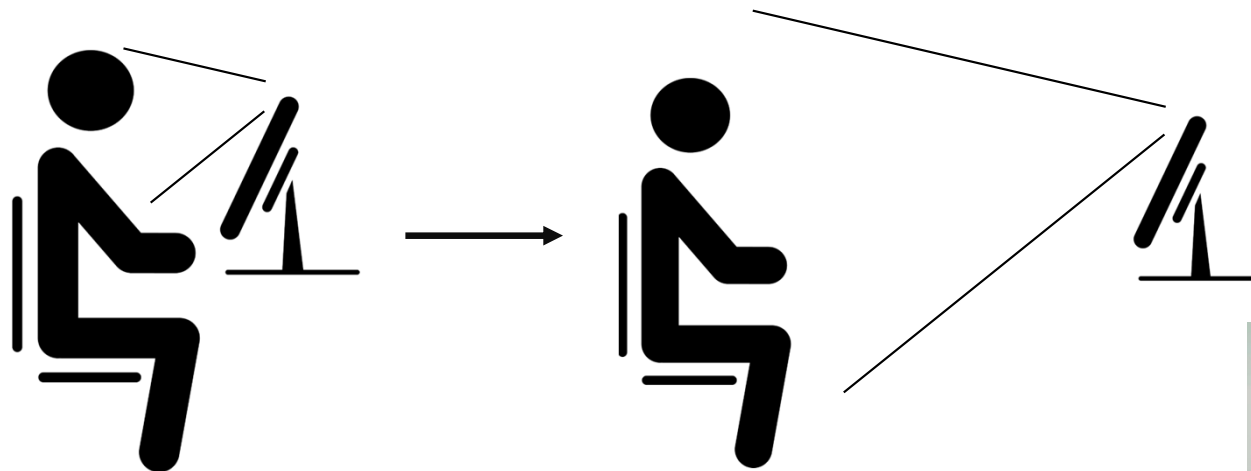
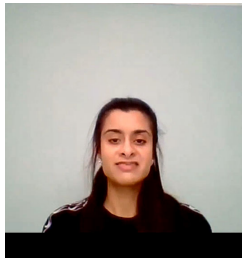
Online multi-modal
research

How to collect high-quality audio-visual data in an unsupervised way?

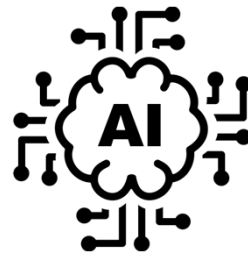
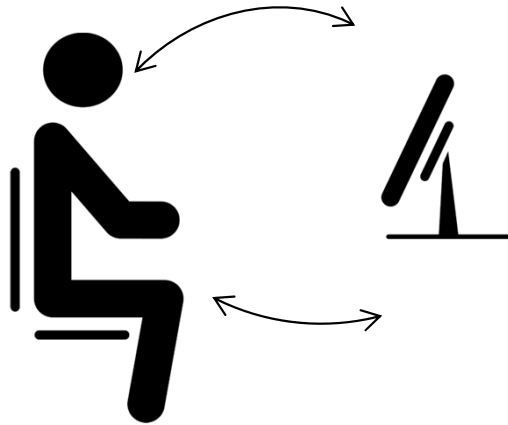
1. High framerate/appropriate volume
2. Appropriate field of view of webcam
3. Naturalistic communicative setting

Online multi-modal research

How to keep participants away from the computer but still be able to interact with the experiment in an unsupervised way?



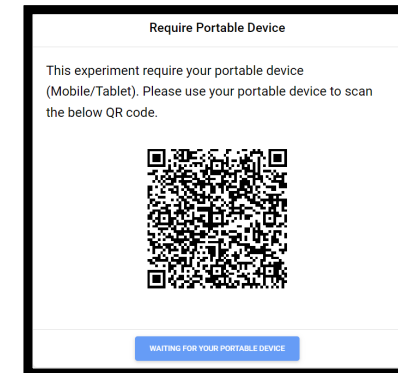
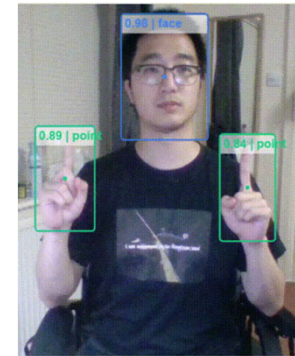
Online multi-modal research – our approach



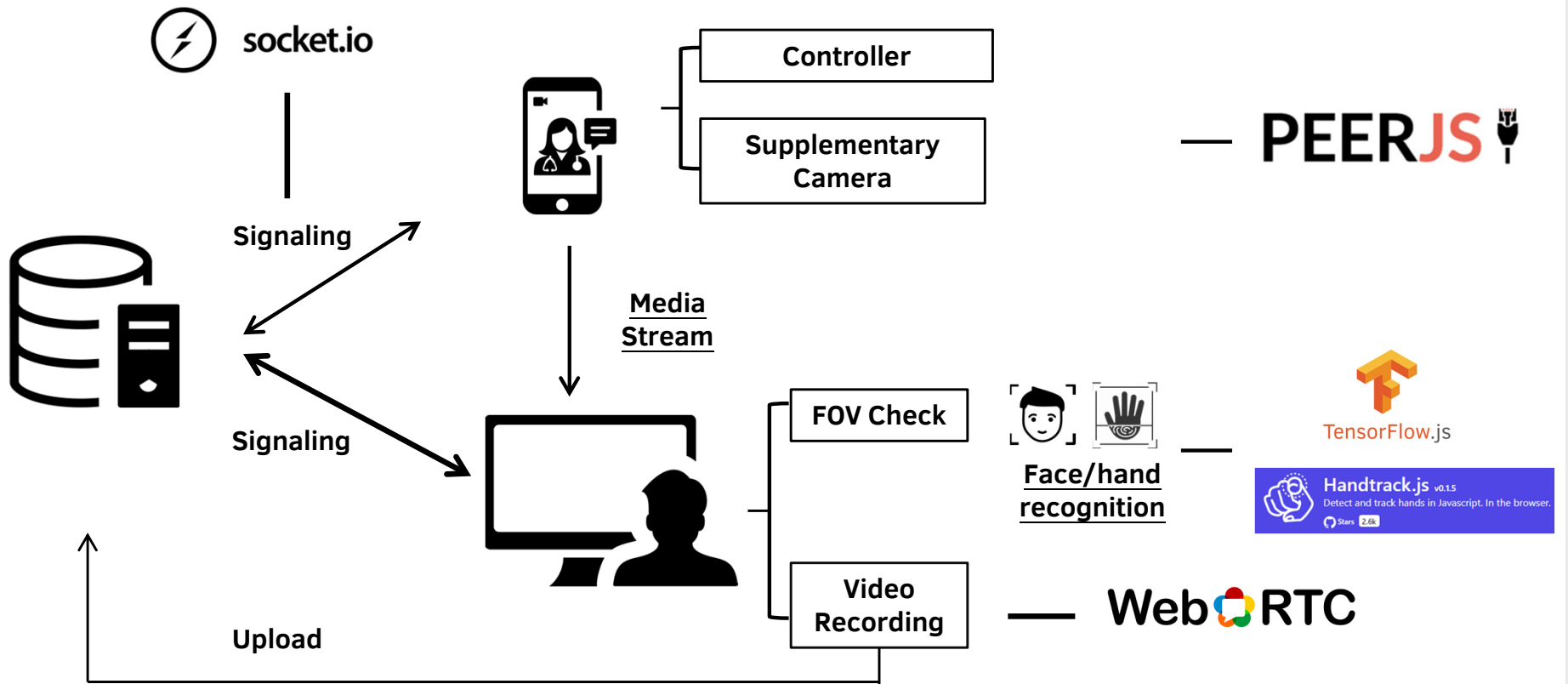
Browser-based Machine Learning Model



Cross-platform Integration



Unsupervised online multi-modal research - structure



Unsupervised online multi-modal research – demo

A demo for cross-platform integration

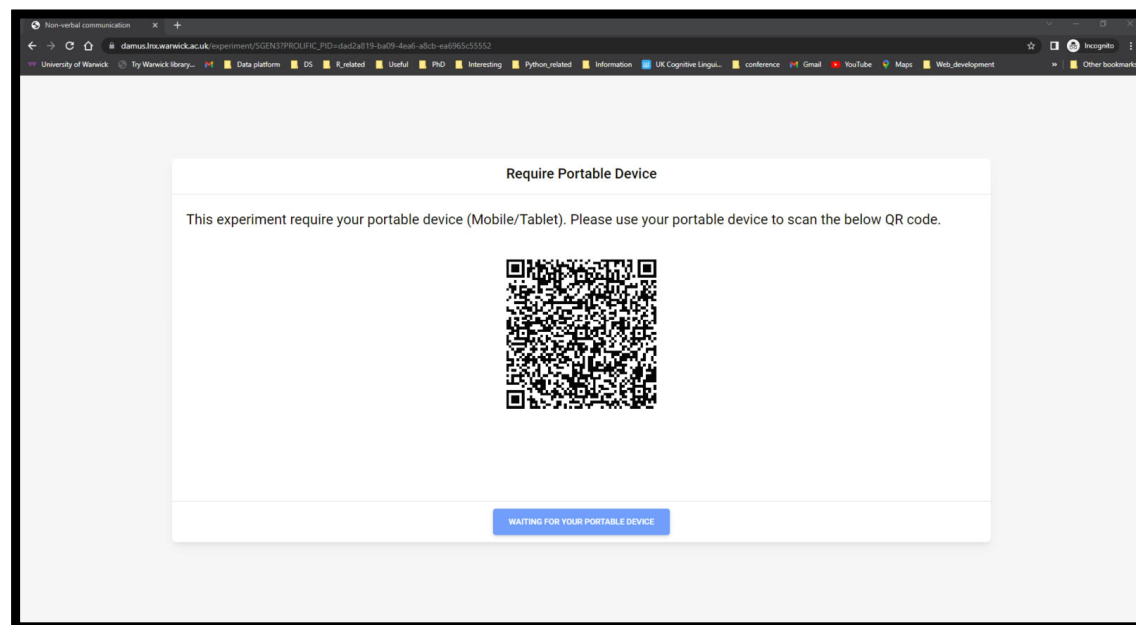
- Elicited pantomime study – participants are asked to produce gestures for the target word.
- A portable device is used as a supplementary camera.
- The FOV is set to be the upper body.



<https://damus.lnx.warwick.ac.uk/experiment/SGEN3>

Unsupervised online multi-modal research – demo

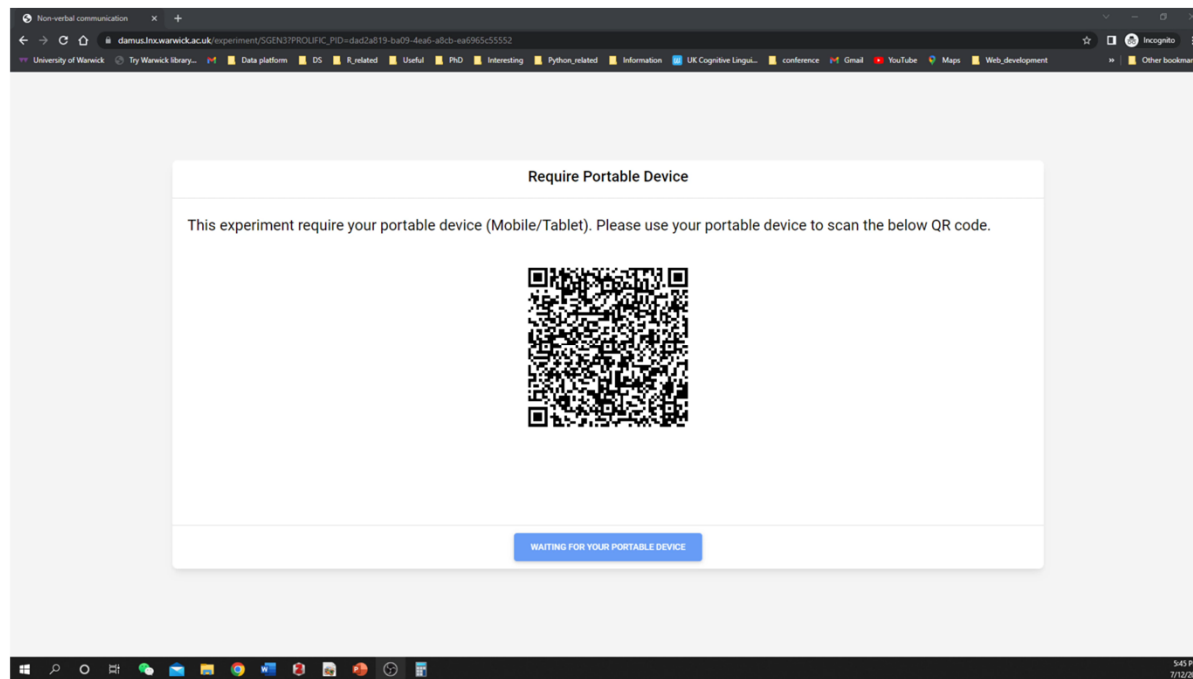
A demo for cross-platform integration



<https://damus.lnx.warwick.ac.uk/experiment/SGEN3>

Unsupervised online multi-modal research – demo

A demo for cross-platform integration



<https://damus.lnx.warwick.ac.uk/experiment/SGEN3>

Unsupervised online multi-modal research – demo

A demo for cross-platform integration

Now adjust your position so that your face and your hands are in the correct position

FPS: 36

0.92 / Face

Instruction

We will record your gestures with the webcam during the experiment. You will be told when you are being recorded.

To make sure that your gestures can be captured by the webcam. Please adjust the angle of your webcam and the distance between you and the laptop. The optimal position for your face and hands (when you hold up your hands in front of your chest) in the video recording is showed in the below picture.

If you are not in the optimal position, please do the following to adjust your position:

5:46 PM 17/12/2022



<https://damus.lnx.warwick.ac.uk/experiment/SGEN3>

Unsupervised online multi-modal research – demo

A demo for cross-platform integration



<https://damus.lnx.warwick.ac.uk/experiment/SGEN3>

Unsupervised online multi-modal research – case study

A case study - framerate analysis

- The web application is implemented on a Linux server (a two cores CPU, 4GiB Memory , equivalent to Amazon t2.medium server, 0.052 USD per hour).
- Elicited pantomime study – participants are asked to produce gestures for a target word.
- No portable device is required.
- 137 participants completed the online experiments. (121 from Research Experience – undergraduate participant recruitment system at the University of Warwick, 16 from Prolific)
- Data from 1 participant was excluded (he/she completed the experiment on a mobile phone, and video data uploaded to the server was corrupted)
- The framerate analysis is based on the 9347 depictions (11 or 16 seconds long)

Unsupervised online multi-modal research – case study

A case study - framerate analysis

Research experience

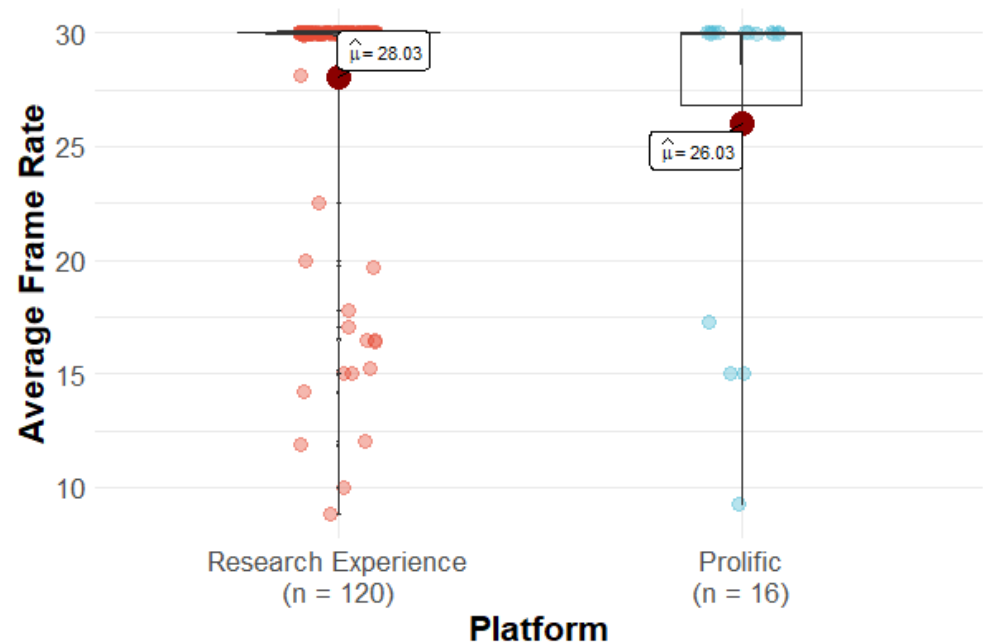
$\hat{\mu} = 28.03$, 95% CI [18.05, 38.02]

Prolific:

$\hat{\mu} = 26.03$, 95% CI [40.25, 11.81]

No bug report during the data collection

80% of the Data collection is completed in 5 days for Research experience and within 3 hours using Prolific.

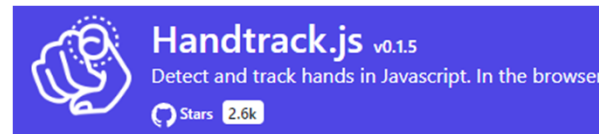
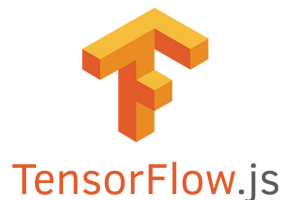


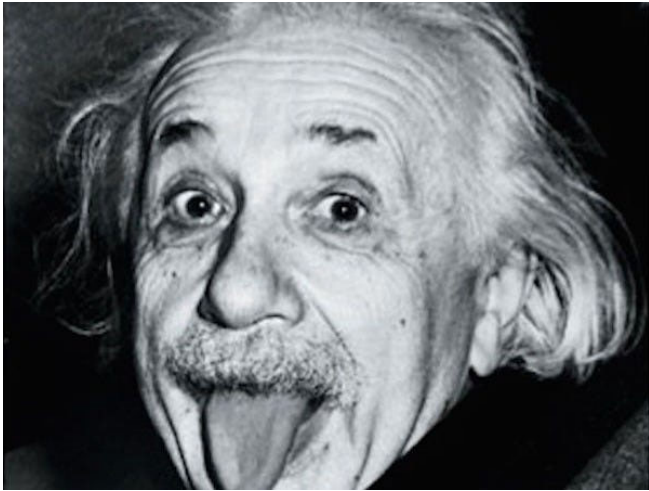
Unsupervised online multi-modal research – demo



Unsupervised online multi-modal research – next step

- The code for the web application will be open access in short via <https://github.com/JH-Eric-Yang/Toolkit-for-multimodal-expeirment>
- Simplify the code and make it more accessible for code migration
- Our framework can also manipulate the video chat (e.g., manipulate the communicative channel/swap communicative partner in each trial)





Take home message

With the support of the recent web technology, we can collect audio-visual data online in a reliable and unsupervised way.

Thanks!

Q & A

More questions?
Jiahao.Yang@warwick.ac.uk