An Experimental Video Conference Platform to Bridge the Gap Between Digital and In-Person Communication

Chloe Eghtebas
Alexander Liebald
Maria Pospelova
Technical University of Munich
Germany

Ashika Manjunath
Julian Geheeb
Norma Puspitasari
Technical University of Munich
Germany

Jamie A Ward†
Gudrun Klinker‡
Goldsmiths, University of London, UK†
Technical University of Munich, Germany‡

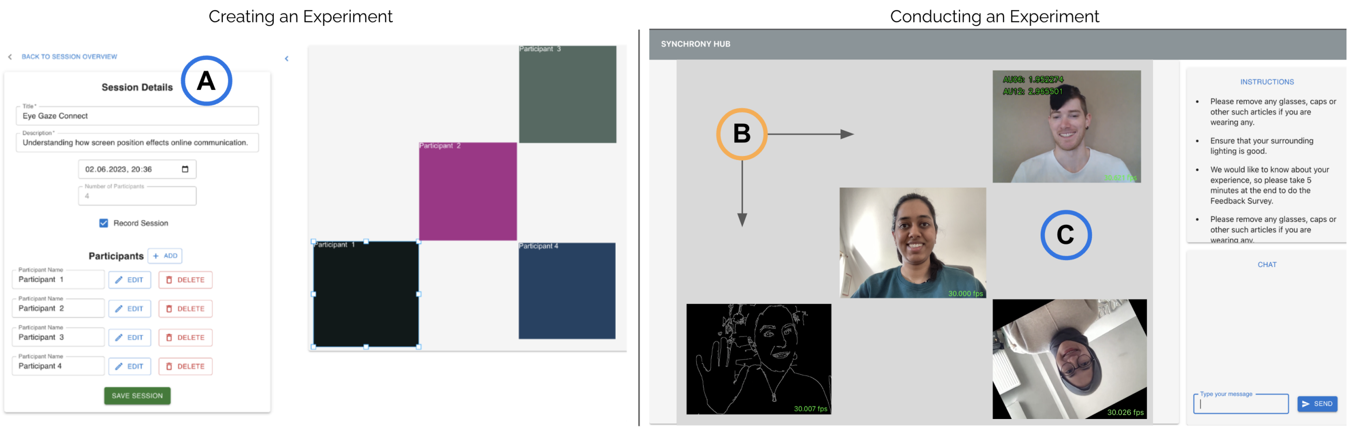

Figure 1: Impressions of the ExpHub during creating and conducting an experiment. A) Session Details form with Title, Date, Time, and Participant details, such as filters, can be added. Next to the form is the canvas where participants video streams can be resized, positioned, and ordered. B) The experiment room where the experimenter "oversees" the experiment. C) *Analysis* and *Manipulation* filters applied to video streams running OpenFace (top right), edge detection, and rotate (bottom row).

ABSTRACT

With many contemporary video conferencing platforms available, there is still a need for platforms that afford a researcher workflow to conduct controlled online experiments. We have developed an open source experimental video conferencing platform that enables researchers to design and conduct remote experiments. Our platform provides a high level of control over the user interface and video streams, which is essential for studying the differences between remote and in-person social interactions. We give an overview of our platform's usage and architecture and conduct a take-home study (N=9) to evaluate how accessible our system is to potential new contributors. We also follow up with an initial evaluation of technical performance bottlenecks for when our experimental platform is deployed, and show that the computational resources increases per each video stream as well as the type of filters applied to each participant. We end with a short discussion on next steps and the experimental hub's potential to be extended as a sandbox for testing browser based augmented reality (WebAR) filters to be adopted in interdisciplinary experimental procedures.

UbiComp/ISWC '23 Adjunct , October 8–12, 2023, Cancun, Quintana Roo, Mexico

ACM ISBN 979-8-4007-0200-6/23/10.
https://doi.org/10.1145/3594739.3610686

CCS CONCEPTS

• **Human-centered computing → User interface toolkits**.

KEYWORDS

Remote User Studies, Video Conferencing, WebAR

ACM Reference Format:
Chloe Eghtebas, Alexander Liebald, Maria Pospelova, Ashika Manjunath, Julian Geheeb, Norma Puspitasari, Jamie A Ward†, and Gudrun Klinker‡. 2023. An Experimental Video Conference Platform to Bridge the Gap Between Digital and In-Person Communication. In *Adjunct Proceedings of the 2023 ACM International Joint Conference on Pervasive and Ubiquitous Computing & the 2023 ACM International Symposium on Wearable Computing (UbiComp/ISWC '23 Adjunct), International Symposium on Wearable Computing (UbiComp/ISWC '23 Adjunct), October 8–12, 2023, Cancun, Quintana Roo, Mexico.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3594739.3610686

1 INTRODUCTION

There are noticeable differences in how we feel during communication on a 2D video conference than meeting in person. From zoom fatigue, latency, and jitter issues [6] to more complex visual manipulation filters that go beyond just 2D interfaces and interactions in emerging telepresence mixed reality (XR) applications who face similarly understudies complex problems (i.e. the effects of augmented reality (AR) filters, avatar appearance, or relative spatiality of holograms [13, 17]). These research issues are beneficial to preemptively explore, especially before they are deployed ubiquitously thus creating unintended societal consequences [3, 10]. Understanding what aspects of interpersonal, verbal, and nonverbal communication factors play a role is a highly multi-variate complex problem to map [18] and requires multiple research backgrounds and new workflows than what general purpose video conferencing tools accommodate [4]. Previous work explores the need of different workflows of video conferencing platforms by developing their own technical solutions for online music performances, conversational facilitation, and connecting long distance families [5, 8, 20]. This can be challenging because not every laboratory has equal access to technical expertise to create their own networked remote experiment. However, the communication, psychological, and user interface (UI) findings from such labs, as well as access to a more diverse remote participant pool from remote experiments, are critical contributions that can shape the future of how we virtually communicate. Finally, the sensitive nature of data collected from video experiments are also a potential limitation for data privacy obliging research.

Recently, there has been an increase in research-based tools for remote experiments in virtual reality (VR) [11, 12]. In this work, we present a self-hosted online video conferencing platform built for a *researcher workflow* with the intention of giving remote interpersonal coordination [15], 2D user interface design, and WebAR filter pipeline experiments a sense of laboratory control. Our open source platform has the goal of increasing *reproducibility* of experiments through experimental session templates, that can be created through our platform and shared to be included in other experimental procedures (See Figure 1 - Creating an Experiment). The experimenter can set WebAR filters, position, and order of participants joined to the call (See Figure 1 - Conducting and Experiment). Lastly, since our platform is self-hosted, it aims to increase the *data privacy* of sensitive video data and information that can be extracted from it, as they do not pass through any additional servers other than the researcher's. However, having our platform be self-hosted limits the accessibility for non-technical users to benefit from our experimental video conferencing platform.

To investigate how easily our experimental video conferencing platform can be adopted and deployed, we conducted a take-home study where participants received instructions and deployed our platform. We also investigated and report principles that would welcome new contributors, as that is a key factor to the ultimate success of sustaining a new open source project [16]. In this work, we also give a short overview of the experimental hub's usage and system implementation of the networking, frontend, and backend architecture and technical design decisions. Finally, we end with a preliminary performance evaluation and discussion on next steps for the experimental hub.

2 SYSTEM IMPLEMENTATION

Our platform backend is written in Python as it facilitates the inclusion of many great scientific libraries, some of which are our core dependencies including: aiortc, pyee, aiohttp, opencv[1]. Our frontend uses React.js with the Atomic Design pattern and Material UI library. Additional libraries utilized include React Konva, React Toastify, and Redux. The communication between front and backend is through a stateful JSON API. Following an event-driven model, our server is responsible for notifying clients if experiment details on the server-side change. Our repository also includes a wiki[2] outlining details about UML class diagrams, API design, contribution and usage guidelines.

Network Architecture is at the core of a web-conferencing platform. Using the Python implementation of WebRTC, we opted for a Selective Forwarding Unit (SFU) media server architecture which allows for less computational resources needed by the server CPU since the data streams are not merged between clients and servers [9]. Each new client establishes their own, authenticated main-connection with their session and participant IDs, and receives a session description response from the server. Although not foolproof in security, this limits the situations of malicious activity or if a participant accidentally falls outside of the intended workflow, thus disrupting an ongoing experimental session. In the example of multiple participants joining a session, the server notifies the client who initiated a sub-connection, which are distinct from the main connection as they solely receive streams from other participants, to view other participants video and audio streams. Similarly, multiple experimenters can join an instance of the self-hosted hub. SSH certificates for securely hosting and a basic password authentication are configurable in the backend to further deter malicious access to the experimenter's hosted instance of the platform.

Filters. Choosing a less intensive CPU performance architecture was important because of the otherwise resource consuming WebAR filters we wanted run in perceptual manipulation experiments on audio and or video streams. We further distinguish filters as something that can manipulate a stream, analyze a stream by only extracting information from it, or a combination of the two. Figure 1, shows an example of an edge detection and rotation manipulation filter (bottom left) and also an instance of OpenFace's [1] action unit extraction as an analysis filter (top right). In our platform, filters can also be chained, where the order of the filters applied determines the result of the analysis or video/audio stream displayed. It is possible to contribute filters from other existing research projects by extending our filter framework [3].

3 USAGE WORKFLOWS

In this section we walk through the EXPERIMENTER and PARTICIPANT workflows that Figure 2 summarises, delineated by the steps through the various experimental platform pages and states before, during, and after running an experiment.

[1]https://github.com/TUMFARSynchrony/experimental-hub/blob/main/backend/requirements.txt
[2]https://github.com/TUMFARSynchrony/experimental-hub/wiki
[3]https://github.com/TUMFARSynchrony/experimental-hub/wiki/Filters

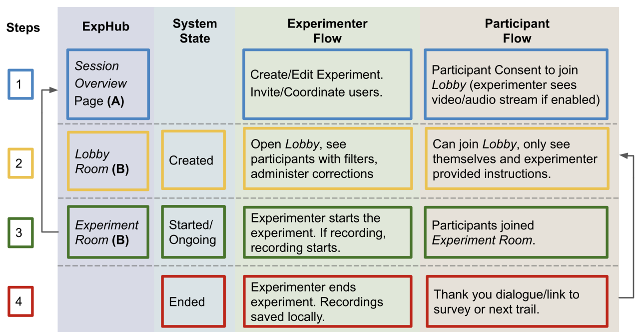

Figure 2: Experimenter and participant user flows along with the various pages and states of the experimental platform.

Experimenters start off on the *Session Overview* page by creating an experiment, either through duplicating an example experiment or by creating a new one through filling out the fields in the *Session Details* page (Figure 1 Creating an Experiment). Upon saving the experimental session data, they can go back to the *Session Overview* page and share the individual invite link to participants by expanding the drop-down next to "Participants". Next, experimenters can then click "Join" to open up the Lobby room such that it starts the session for participants to join. This step was designed to minimize social compounding effects, as participants cannot see each other yet until the experiment is started by the experimenter.

At this stage, the experimenter can check the functionality of applied filters and administer corrections or instructions to participants (e.g. remove glasses, change camera angle, fix lighting, etc.) to ensure a functioning technical setup and produce good recordings for analysis. The experimenter can then start once ready, but still retains the ability to ban, mute, and taking notes during the experiment through the interface. Once the experiment is complete, the experimenter ends the session and is navigated back to the "Session Overview" page. The audio and video streams are recorded to the experimenter's hosting device. A follow-up repeated measure session can be done by repeating the process above, and sharing the link to the next session.

Participants are sent an invite link to the session, and upon visiting the link they are prompted to provide consent to their video and audio stream being viewed and potentially recorded by the experimenter. If the experimenter has not "Joined" the session yet, Participants will not be allowed to join the *Lobby*, which is the next screen they are otherwise redirected to in Step 2 in Figure 2. The Lobby displays a video stream of the participant and experimental instructions provided (Similar to Figure 1 Conducting an Experiment, but without viewing other participants). Once the experimenter has started the experiment, participant streams begin recording.

4 METHOD

We conducted a take-home study to evaluate the on-boarding process and understand the factors that influence contributions to a new open source project like our experimental hub. Nine computer scientists (Male = 5, Female = 4; average age = 31 years, std = 15 years) were recruited through personal contacts who also had the ability (time and expertise) to complete the study. Seven were pursuing their masters degree, one their PhD, and one was an industry professional. Lastly, we recruited participants who were users across different operating systems (4 macOS, 3 Windows, and 2 Linux users).

	P1	P2	P3	P4	P5	P6	P7	P8	P9
1) Total Time taken in hrs	.67	3	2	1.5	.5	3	2	2.5	3.5
2) Success hosting	✓			✓	✓	✓	✓		✓
3) Joined video streams	10			10		10	10	9	2

Table 1: Table shows the task completion across participants for 1) Installing and Running the experimental platform, 2) Hosting it, and 3) Running a Trial Experiment.

Participants were emailed a consent, demographics, and an instructions document, outlining the steps to 1) set up their environment, install and run the experimental hub, 2) host the hub, and 3) run an experiment where they join up to 10 video streams in a single call. Participants were given a week to complete the instructions and encouraged to make edits in the take-home document that reflected the steps according to how their actual process went as well as upload screen shots along the way. We ended with a follow-up semi-structured interview, which lasted an average of 19.8 min (σ = 7.6 min) with 5 questions around open source best practices (2 questions), and experience running and using the experimental hub as compared to other video conferencing platforms (3 questions).

We additionally conducted a performance evaluation to investigate the take-home study findings by recording the CPU, memory, and network usage as number of video streams in a session increased. We conducted these tests on a mid-range laptop, 2.9 GHz 6-Core Intel Core i9, 32 GB RAM at 2400 MHz running macOS Monterey version 12.5.1, and compared the performance with and without the OpenFace Action Unit (AU) extraction filter activated on each individual video stream.

5 RESULTS

Take-home Study Task Completion All 9 participants completed the environmental setup and ran the experimental hub successfully through the take-home study instructions which took a self reported average of 2.1 hrs (σ =1.0 hrs) to complete. Participants made an average of 3.4 changes (std = 1.4) to the instruction document. Environmental setup section had the most unique new instructions added (8 changes), followed by hosting at (5 changes), Running the Hub and Running an Experiment at 3 changes, each. As also shown in Table 1, only six successfully hosted and completed running an experiment respectively. Of the six participants who ran an experiment, four participants managed to connect all 10 video streams in a single session, one managed 9 video streams, and one only could connect with 2 as shown in Table 1. Of the participants who ran the sample experiment, four did not report any noticeable lag or choppiness as more video streams joined the experimental

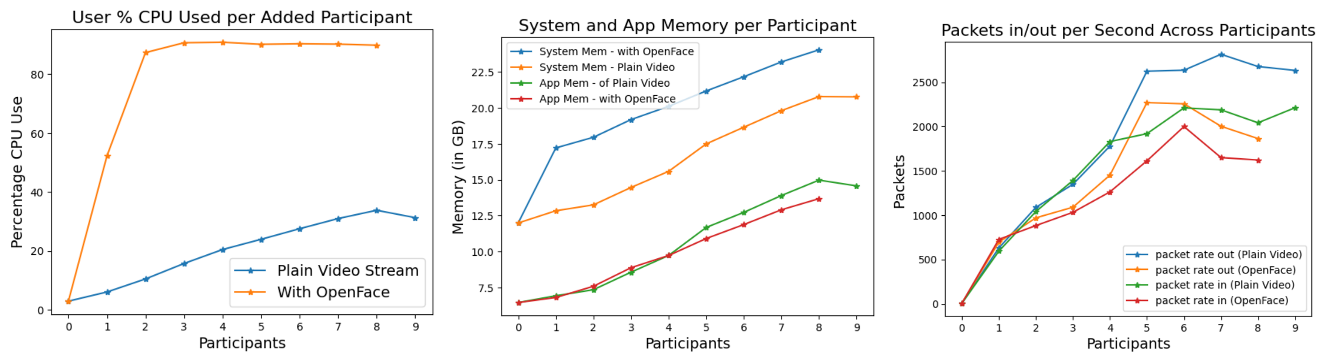

Figure 3: The CPU, Memory, and Network usage as number of participant video streams join an experimental session.

session. P9 experienced the video streams dropping after only two participants joined and P2 and P5 were able to connect the streams but had difficulty with accessing their web cameras either due to the browser they were using (e.g. Safari) or the fact that they had multiple cameras connected.

Performance Evaluation Figure 3 shows the CPU, Memory, and Network load of running a test experiment on our platform on a Chrome browser. The system we used for the performance evaluation could reliably join 8 video streams while hosting the experimental hub. The CPU of the test device maxes out after joining two participants with the OpenFace AU extraction filter enabled (each participant runs their own OpenFace AU extraction instance). Otherwise, as the number of participant streams increases, the amount of resources increases linearly across the tests. For packets in/out per second (3rd graph in Figure 3), the linear trend breaks after four participants, which is also when lag can be experienced across the video streams on the test system.

Take-home Study Interview Findings revolved around a couple of themes following a thematic analysis [2], the first being HEURISTICS FOR OPEN SOURCE PROJECTS AND CONTRIBUTIONS. Seven participants reported having used open source projects and only three have contributed to them, from committing bug reports to contributing code from tasks at work. All seven emphasized the documentation as being critical for the on-boarding of an open source project *"If the setup is fine, everything else works well"* (P3). Three participants mentioned preference to text over instructional videos and OS-specific instructions for the setup respectively. Participants also mentioned how they would be motivated to contribute to a open source project if there was a shared *"common goal"* (P8) or *"tools that I use that I would like to improve"* (P6). Other motivational factors for contributing included regular versioning and updates to instructions (P2) as well as the number of stars (P1) and papers linked (P1 + P3) to the GitHub repository as indicators of an active and inviting open source project.

The other theme in the interview was around the NOVELTY OF OUR EXPERIMENTAL PLATFORM AGAINST OTHER VIDEO CONFERENCING TOOLS. The background of participants in our take-home study was technical but they were not researchers who would benefit most from our platform. Users, like P4 and P6, compared our video conferencing tool to the ones they have most familiarity with, which are usually desktop applications, *"I've used professional tools like zoom, it [the experimental platform] is different from what we are used to ... its like a tool that normal technical people would use but would be advanced for general people."* (P6). Participants continued to note the difference between our user interface among other conferencing platforms and mentioned how they thought it was *'strange'* how the video streams in the trial experiment they ran were intentionally different sizes (N=5). Three participants also continued to explore the platform outside of the instructions document by creating a new test experiment (P7), attempting to take notes in the ongoing experiment page (P2), and even adding an 11th video stream to the call (P6).

6 DISCUSSION AND CONCLUSION

As seen from the interview results from the take-home study, it is critical next step to evaluate our workflows with our intended users – researchers conducting experiments, both through a series of expert interviews and performance evaluations on experiments being run through our platform. Interviews also showed the level of documentation to create an inviting and sustainable open source community is vital to first impressions when deciding to use the code base.

As of now, technical familiarity is required for setting up the platform, but this process could be automated with a bash script or containerized according to project version releases. Another limitation of our project is the hosting of our platform, although it affords data privacy, it also comes at the cost of needing access to network settings which could be an administrative hassle.

Our independent and take-home performance evaluation showed how even with a mid-range laptop, the experimental hub architecture is able to accommodate multiple remote participants joining a call, thus making our code base accessible to anyone with minimal hardware or computational resources. However, joining many participants or even enabling complex filters like OpenFace AU extraction requires significantly more resources, which in the future, such filters can be run post-experiment in situations where this limitation must be countered.

Our self-hosted remote experimental platform is the basis of a new open source, community-led effort for multidisciplinary researchers investigating verbal and nonverbal remote communication. We hope it garners interest from this research community, enough to use and standardize in the design of remote experimental procedures and as a sandbox for testing WebAR filters in social and psychological experiments. We welcome collaborations on extending our filters with existing computer vision research pipelines, such as determining facial similarity, rPPG, and detection of mental states like depression [7, 14, 19], and are excited to continue development and see how the platform is adopted in new research agendas.

7 ACKNOWLEDGEMENTS

This work is partly supported under the Leverhulme Trust, and the Academies APEX award (APX101093).

REFERENCES

[1] Tadas Baltrušaitis, Peter Robinson, and Louis-Philippe Morency. 2016. Openface: an open source facial behavior analysis toolkit. In *2016 IEEE winter conference on applications of computer vision (WACV)*. IEEE, 1–10.

[2] Virginia Braun and Victoria Clarke. 2022. *Thematic Analysis: a Practical Guide.* SAGE, Los Angeles London New Delhi Singapore Washington DC Melbourne.

[3] Chloe Eghtebas, Gudrun Klinker, Susanne Boll, and Marion Koelle. 2023. Co-Speculating on Dark Scenarios and Unintended Consequences of a Ubiquitous(Ly) Augmented Reality. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference* (Pittsburgh, PA, USA) *(DIS '23)*. Association for Computing Machinery, New York, NY, USA, 2392–2407. https://doi.org/10.1145/3563657.3596073

[4] Chris Elsden, David Chatting, Michael Duggan, Andrew Carl Dwyer, and Pip Thornton. 2022. Zoom Obscura: Counterfunctional Design for Video-Conferencing. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 143, 17 pages. https://doi.org/10.1145/3491102.3501973

[5] Sean Follmer, Hayes Raffle, Janet Go, and Hiroshi Ishii. 2010. Video Play: Playful Interactions in Video Conferencing for Long-Distance Families with Young Children. In *CHI '10 Extended Abstracts on Human Factors in Computing Systems* (Atlanta, Georgia, USA) *(CHI EA '10)*. Association for Computing Machinery, New York, NY, USA, 3397–3402. https://doi.org/10.1145/1753846.1753991

[6] Shrikant Garg, Ayushi Srivastava, Mashhuda Glencross, and Ojaswa Sharma. 2022. A Study of the Effects of Network Latency on Visual Task Performance in Video Conferencing. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI EA '22)*. Association for Computing Machinery, New York, NY, USA, Article 213, 7 pages. https://doi.org/10.1145/3491101.3519678

[7] Maximilian Henneberg, Chloe Eghtebas, Oliver De Candido, Kai Kunze, and Jamie A Ward. 2023. Detecting an Offset-Adjusted Similarity Score Based on Duchenne Smiles. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI EA '23)*. Association for Computing Machinery, New York, NY, USA, Article 83, 5 pages. https://doi.org/10.1145/3544549.3585709

[8] Margaret A Hughes and Deb Roy. 2021. Keeper: A synchronous online conversation environment informed by in-person facilitation practices. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–14.

[9] Bart Jansen, Timothy Goodwin, Varun Gupta, Fernando Kuipers, and Gil Zussman. 2018. Performance Evaluation of WebRTC-Based Video Conferencing. *SIGMETRICS Perform. Eval. Rev.* 45, 3 (mar 2018), 56–68. https://doi.org/10.1145/3199524.3199534

[10] Ross Johnstone, Neil McDonnell, and Julie R. Williamson. 2022. When Virtuality Surpasses Reality: Possible Futures of Ubiquitous XR. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI EA '22)*. Association for Computing Machinery, New York, NY, USA, Article 6, 8 pages. https://doi.org/10.1145/3491101.3516396

[11] Jaewook Lee, Raahul Natarrajan, Sebastian S. Rodriguez, Payod Panda, and Eyal Ofek. 2022. RemoteLab: A VR Remote Study Toolkit. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology* (Bend, OR, USA) *(UIST '22)*. Association for Computing Machinery, New York, NY, USA, Article 51, 9 pages. https://doi.org/10.1145/3526113.3545679

[12] Tobias Loetscher, Nadia Siena Jurkovic, Stefan Carlo Michalski, Mark Billinghurst, and Gun Lee. 2023. Online platforms for remote immersive Virtual Reality testing: an emerging tool for experimental behavioral research. *Multimodal Technologies and Interaction* 7, 3 (2023), 32.

[13] Mark Roman Miller, Hanseul Jun, Fernanda Herrera, Jacob Yu Villa, Greg Welch, and Jeremy N Bailenson. 2019. Social interaction in augmented reality. *PloS one* 14, 5 (2019), e0216290.

[14] Le Nguyen, Manuel Lage Cañellas, Constantino Álvarez Casado, Xiaoting Wu, and Miguel Bordallo López. 2023. Synchrony-Based Depression Score Aggregation from Single-Modality Models. In *Adjunct Proceedings of the 2022 ACM International Joint Conference on Pervasive and Ubiquitous Computing and the 2022 ACM International Symposium on Wearable Computers* (Cambridge, United Kingdom) *(UbiComp/ISWC '22 Adjunct)*. Association for Computing Machinery, New York, NY, USA, 198–201. https://doi.org/10.1145/3544793.3563410

[15] Alexandra Paxton, Rick Dale, and Daniel C Richardson. 2016. Social coordination of verbal and non-verbal behaviours. In *Interpersonal coordination and performance in social systems*. Routledge, 277–292.

[16] Alexandra Paxton, Nelle Varoquaux, Chris Holdgraf, and R Stuart Geiger. 2022. Community, Time, and (Con) text: A Dynamical Systems Analysis of Online Communication and Community Health among Open-Source Software Communities. *Cognitive Science* 46, 5 (2022), e13134.

[17] Susruthi Rajanala, Mayra BC Maymone, and Neelam A Vashi. 2018. Selfies—living in the era of filtered photographs. *JAMA facial plastic surgery* 20, 6 (2018), 443–444.

[18] Rabindra Ratan, Dave B Miller, and Jeremy N Bailenson. 2022. Facial appearance dissatisfaction explains differences in zoom fatigue. *Cyberpsychology, Behavior, and Social Networking* 25, 2 (2022), 124–129.

[19] Zhaodong Sun, Alexander Vedernikov, Virpi-Liisa Kykyri, Mikko Pohjola, Miriam Nokia, and Xiaobai Li. 2023. Estimating Stress in Online Meetings by Remote Physiological Signal and Behavioral Features. In *Adjunct Proceedings of the 2022 ACM International Joint Conference on Pervasive and Ubiquitous Computing and the 2022 ACM International Symposium on Wearable Computers* (Cambridge, United Kingdom) *(UbiComp/ISWC '22 Adjunct)*. Association for Computing Machinery, New York, NY, USA, 216–220. https://doi.org/10.1145/3544793.3563406

[20] Takuro Yonezawa and Hideyuki Tokuda. 2012. Enhancing Communication and Dramatic Impact of Online Live Performance with Cooperative Audience Control. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing* (Pittsburgh, Pennsylvania) *(UbiComp '12)*. Association for Computing Machinery, New York, NY, USA, 103–112. https://doi.org/10.1145/2370216.2370234

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009