What You Show is What You Get! Gestures for Microtask Crowdsourcing

Allen, Garrett; Hu, Andrea; Gadiraju, Ujwal

DOI
10.1145/3581754.3584175

Publication date
2023

Document Version
Final published version

Published in
IUI 2023 - Companion Proceedings of the 28th International Conference on Intelligent User Interfaces

What You Show is What You Get! Gestures for Microtask Crowdsourcing

Garrett Allen
Andrea Hu
Ujwal Gadiraju
G.M.Allen@tudelft.nl
hjc3299@gmail.com
U.K.Gadiraju@tudelft.nl
Delft University of Technology
Delft, Zuid-Holland, Netherlands

ABSTRACT

Crowdsourcing is a valuable tool to gather human input which enables the development of reliable artificial intelligence systems. Microtask platforms like Prolific and Amazon's Mechanical Turk have flourished by creating environments where crowd workers can provide such human input in a diverse and representative manner. Such marketplaces have evolved to support several hundreds of workers in earning their primary livelihood through crowd work. Crowd workers, however, often perform these tasks in sub-optimal work environments with poor ergonomics. Additionally, many of the various microtasks require input via the standard method of a mouse and keyboard and are repetitive in nature. As such, crowd workers who primarily earn their livelihoods in microtask marketplaces are at risk of injuries such as carpal tunnel syndrome. By changing the input modality from a mouse and keyboard to gesture-driven input, crowd workers can complete their work while simultaneously improving or safeguarding their physical health. Through three distinct microtasks, we constructed a dataset that enables the exploration of the physical and mental health of crowd workers while using gestures. In this work, we present the process of constructing this dataset, how we applied it, and the future applications we foresee.

KEYWORDS

crowdsourcing, microtasks, datasets, neural networks, pose detection

ACM Reference Format:
Garrett Allen, Andrea Hu, and Ujwal Gadiraju. 2023. What You Show is What You Get! Gestures for Microtask Crowdsourcing. In *28th International Conference on Intelligent User Interfaces (IUI '23 Companion), March 27–31, 2023, Sydney, NSW, Australia.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3581754.3584175

IUI '23 Companion, March 27–31, 2023, Sydney, NSW, Australia

ACM ISBN 979-8-4007-0107-8/23/03.
https://doi.org/10.1145/3581754.3584175

1 BACKGROUND AND INTRODUCTION

Research in the realm of Artificial Intelligence (AI) has seen rapid growth in recent years with widespread promises and real impact across various domains, on society and economies at large [9]. An empowering factor in this expansion is the availability of data and the potential of human input on demand via crowdsourcing [12]. The crowdsourcing paradigm gives researchers and practitioners access to reliable, high-quality, representative human input at scale, through marketplaces like Prolific or Amazon Mechanical Turk [18]. In turn, these marketplaces give individuals, known as crowd workers, a space to earn a livelihood through microtasks [3, 4]. While this paradigm benefits all involved, there are several factors that threaten the overall sustainability [8, 26, 29].

Crowd workers often perform microtasks in sub-optimal work environments with poor ergonomics [10, 14]. Over decades, research has been conducted regarding the ergonomics of desk work, particularly for sitting workers, and the physical discomforts that arise [21]. Yet, to date, there is a lack of investigation into the effects of crowd work specifically on the physical health of the workers. Past work has explored increasing worker engagement and satisfaction, as well as worker experience [5, 17, 24, 25]. Many of the various microtasks require input via the conventional method of using a mouse and keyboard and are repetitive in nature, resulting in crowd workers who primarily earn their livelihoods in microtask marketplaces being at risk of injuries such as carpal tunnel syndrome [23]. Mental wellbeing has been shown to be closely related to the physical wellbeing of people in a variety of contexts [22]. Considering that crowd workers deal with a variety of challenges that impede their mental wellbeing, ranging from unfair treatment to cognitive exhaustion [7, 19], or exposure to explicit or disturbing content [15, 28], it is important to consider the impact of repetitive work in monotonous task batches on their physical health. At this juncture we argue in the potential of considering the integration of a novel input modality that can serve the dual purpose of facilitating better physical health among workers while accommodating the primary need for input elicitation. By changing the input modality from a mouse and keyboard to gesture-driven input crowd workers can complete their work while potentially improving or safeguarding their physical health. With a similar vision, prior work has shown the potential of using sign language [27].

To this end, we conducted a set of experiments to explore the performance and perception of gesture inputs for microtasks [1]. We

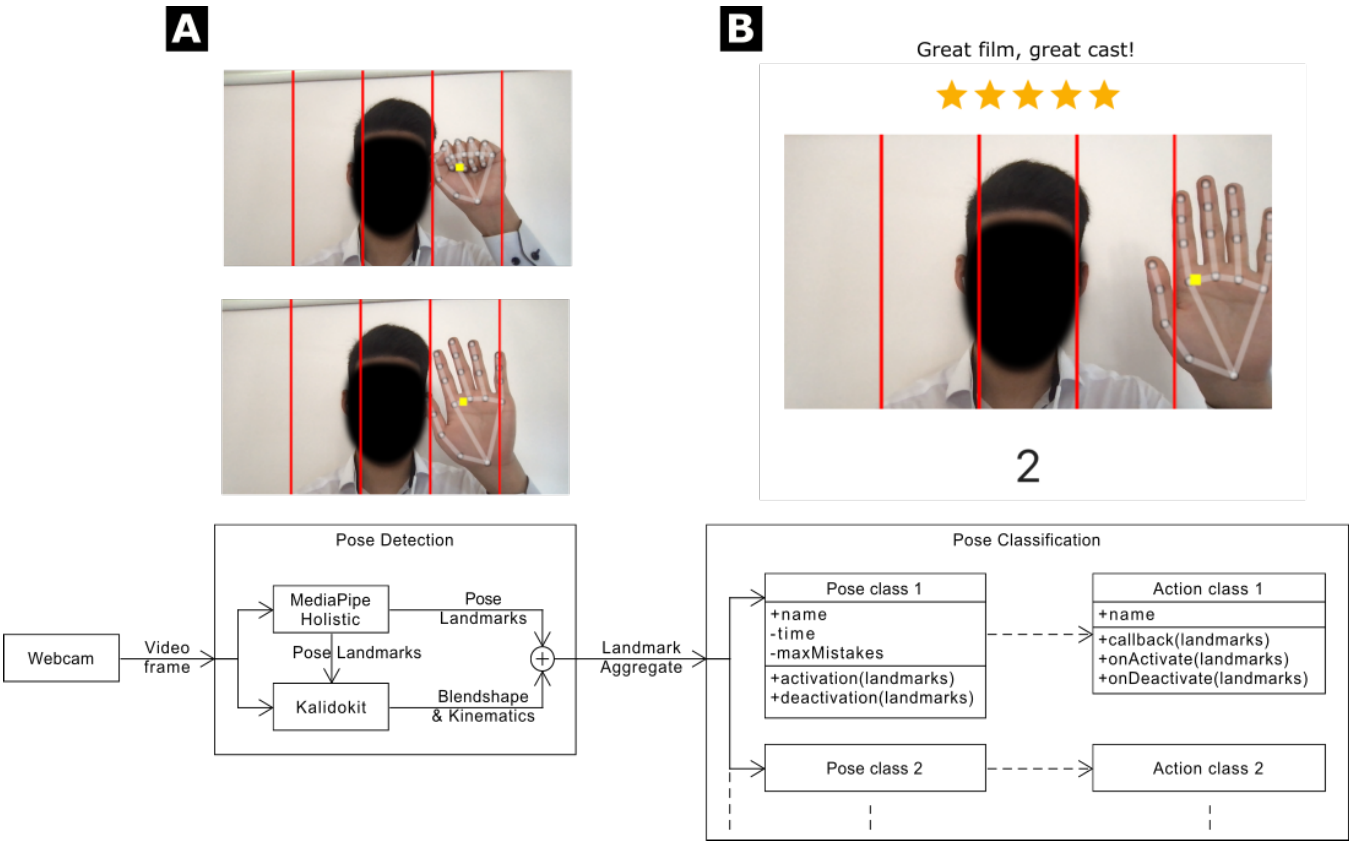

Figure 1: Structural diagram of the gesture capture pipeline. Images represent visual component of each stage in the pipeline: Part A for pose detection and Part B for pose classification.

used three distinct microtasks, informed by the taxonomy described by Gadiraju et al. [11]. Each task used a gesture input specifically designed for the task. An *information-finding* task made use of head and neck motions, with a tilted head highlighting an option and a nod forward submitting the selection. A *sentiment analysis* task made use of an open palm and a closed fist. The closed fist enables workers to change their selection freely, while the open palm submits the selection after two seconds. Finally, a *classification* task utilizes a number of digits displayed (*i.e.*, a show of fingers), excluding the thumb, to make a selection.

Client-side pose detection and classification processes are used to capture and interpret the gestures. The pose detection stage (A in Figure 1) detects the body, face and hand landmarks of the worker on each frame. Multiple pre-trained models perform real-time pose landmark estimation of the body, face, and hands. The models are sourced from MediaPipe.[1] The landmarks are converted into more directly interpretable data via multiple methods using Kalidokit.[2] Details on the augmentation can be found in the library documentation.

Pose classification (B in Figure 1) is achieved by mapping the interaction between a set of pose and action classes. Each pose class has an entry and exit condition. Upon activation of the entry condition, the pose class enters an "active" state, where it will remain until the exit condition is met. The entry and exit conditions are functions that, when given landmarks of the body, face and hands, check whether some conditions of the pose are met. For example, if the coordinates of a landmark enter a certain range of the webcam frame or if an angle of a joint exceeds a certain degree. During the active state of the pose class, time spent in that state is measured. An action class monitors this elapsed duration and triggers a set of functions as needed. The *callback* function is used for changing the answer choice or to submit the answer, while the activate and deactivate functions are for handling the starting and resetting of the timer. A modular setup with the ability to dynamically assign gestures to actions is a result of separating the pose classes and their related actions.

As a part of the experiment, we constructed a dataset that enables the exploration of the physical and mental health of crowd workers, as well as their task performance, while using gestures. In this work, we present the process of collecting and constructing this dataset, how we applied it, and the future applications we foresee. The contributions of this work include potential future applications of the dataset (*e.g.*, as a benchmark to test the effectiveness of future

[1] https://google.github.io/mediapipe/solutions/holistic.html
[2] https://github.com/yeemachine/kalidokit

Table 1: A detailed listing of the data collected in our study across different tasks and through pre/post-task surveys.

Task & Survey	Fields
Information Finding	professionPersons, midnamePerson, inputModality, taskType, entrySurveyCompleted, exitSurveyCompleted, revoked, id, dateAdded, complete, questionNumber, state, answers, poses, uid
Classification	birds, inputModality, taskType, entrySurveyCompleted, exitSurveyCompleted, revoked, id, dateAdded, complete, questionNumber, state, answers, poses, uid
Sentiment Analysis	movieReviews, inputModality, taskType, entrySurveyCompleted, exitSurveyCompleted, revoked, id, dateAdded, complete, questionNumber, state, answers, poses, uid
Pre-Task Survey	gender, age, mood, yearly_income, weekly_hours, working_times, experience, attention_check_pre, work_env_healthy, work_env_comfort body_parts_comfort_1, body_parts_comfort_2, body_parts_comfort_3, body_parts_comfort_4, body_parts_comfort_5, body_parts_comfort_6, taskType, inputModality, uid, Duration (in seconds)
Post-Task Survey	ues_perceived_usability, ues_reward, sf_36_emotional_well_being, sf_36_energy_and_fatigue, sf_36_health, task_load_index, copsoq_prod_emotion_reversed, copsoq_prod_speed_reversed, copsoq_prod_time, Duration (in seconds), reflection_on_error, performance_estimate_1, task, input_modality, uid

gestures in identical task settings), and transparency of our process through reflections on lessons learned during dataset creation.

2 DATASET DESCRIPTION AND LESSONS LEARNED FROM ITS CONSTRUCTION

The tasks share many data points, but each has unique data values that needed to be collected for the purpose of performance evaluation. The information-finding task required storing the professions and middle names, while the classification tasks needed the bird beak types, and the sentiment analysis required storing the ground-truth ratings. The pre-task and post-task survey values collected represent the perceptions of the workers regarding the different inputs. A complete overview of data collected can be seen in Table 1. The full description of each task and the corresponding data that was gathered is available on the Open Science Framework repository. [3]

2.1 Critical Reflection on What Went Well and Aspects That Can Improve

For a microtask crowdsourcing input modality to be effective and to facilitate adoption at scale, it is pivotal to consider the extent to which workers' privacy is safeguarded. To protect the workers' privacy, no personally identifiable information was collected. The images from workers' webcams that were used for gesture recognition were shown and processed on the participants' devices and were never sent to an external back-end server, or stored anywhere. Pose landmarks were collected on some actions, but such data are not personally identifiable information. The development of the gesture input modality followed a high degree of modularity, which supports and facilitates future extensions that can encompass an inventory of gestures.

[3]https://osf.io/7x526/

The gestures we considered in our first exploration do not fully reflect the tight coupling with ergonomics we envision. Future work should explore the creation of a well-informed mapping between potential gestures, the concomitant ergonomical benefits, and the suitable task types. The gestures we considered only relate to the use of hands and the head. These can be extended, especially to cover ergonomically well-informed gestures that relate to movements corresponding to other parts of the body.

2.2 Cognitive Biases and Measures for Bias Mitigation

Performing research involving crowd workers has the potential to introduce numerous cognitive biases depending on task design and workflow. Analyzing our study design using the Cognitive Biases Checklist introduced by Draws et al. [6], we identified the potential presence of *self-interest bias*, *familiarity bias*, *optimism bias*, *sunk cost fallacy*, and *disaster neglect*.

Self-interest bias could have potentially manifested due to the monetary compensation of crowd workers we recruited from the Prolific crowdsourcing platform. One worker demonstrated obviously low effort by giving the same response for every question and completing the task in a very short space of time. This behaviour was attributed to the self-interest bias, with the submission being rejected as a measure for mitigation. The familiarity bias is possible due to the comparative nature of the study. Intuitively, workers will have more familiarity with the standard inputs than the gestures. As an effort to offset this bias, time for workers to perform a tutorial was provided when using the gesture inputs. Through clear instructions and a detailed task description, we attempted to address the presence of *optimism bias* by ensuring the workers were well-informed before opting to complete our tasks. The *sunk cost fallacy* also potentially shaped the data we gathered. We conducted a small pilot study to get an informed estimate of how long each task would take in order to minimize this effect by pricing the task fairly [2]. Finally, there is the possibility of *disaster neglect*, or workers being improperly informed of the consequences of the task. We asked workers to complete an informed consent form to address this bias.

3 USING THE GESTURES DATASET

The dataset we thus produced, along with the task particulars, can help serve as a benchmark dataset for novel input modalities that are developed for microtask crowdsourcing in the future. For instance, we aim to identify a new family of ergonomically informed gestures to elicit input from crowd workers. We can then use the existing task and gesture data to serve as a comparative benchmark, with an aim to draw insights into the efficiency and effectiveness of the said gestures. Increasing transparency in the data collection process and examining the potential biases that could have played a role in shaping the elicited input can help in the better use of datasets in downstream tasks. This is largely in line with the spirit of prior works that have suggested the use of *datasheets* for datasets [13], *model cards* for models [20], dataset *nutrition labels* [16], and the *cognitive bias checklist* to identify, mitigate, and reflect on cognitive biases present as an artefact of datasets [6].

ACKNOWLEDGMENTS

This work was partially supported by the Delft Design@Scale AI Lab and the 4TU.CEE (grant number: EWI-ICT-#21UNCAGE). We thank the anonymous crowd workers from Prolific who participated in this study.

REFERENCES

[1] Garrett Allen, Andrea Hu, and Ujwal Gadiraju. 2022. Gesticulate for Health's Sake! Understanding the Use of Gestures as an Input Modality for Microtask Crowdsourcing. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 10. 14–26.

[2] Justin Cheng, Jaime Teevan, and Michael S Bernstein. 2015. Measuring crowdsourcing effort with error-time curves. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. 1365–1374.

[3] Djellel Difallah, Elena Filatova, and Panos Ipeirotis. 2018. Demographics and dynamics of mechanical Turk workers. In *Proceedings of the eleventh ACM international conference on web search and data mining*. 135–143.

[4] Djellel Eddine Difallah, Gianluca Demartini, and Philippe Cudré-Mauroux. 2012. Mechanical cheat: Spamming schemes and adversarial techniques on crowdsourcing platforms.. In *CrowdSearch*. 26–30.

[5] Mira Dontcheva, Robert R Morris, Joel R Brandt, and Elizabeth M Gerber. 2014. Combining crowdsourcing and learning to improve engagement and performance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 3379–3388.

[6] Tim Draws, Alisa Rieger, Oana Inel, Ujwal Gadiraju, and Nava Tintarev. 2021. A checklist to combat cognitive biases in crowdsourcing. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 9. 48–59.

[7] Tom Edixhoven, Sihang Qiu, Lucie Kuiper, Olivier Dikken, Gwennan Smitskamp, and Ujwal Gadiraju. 2021. Improving Reactions to Rejection in Crowdsourcing Through Self-Reflection. In *13th ACM Web Science Conference 2021*. 74–83.

[8] Shaoyang Fan, Ujwal Gadiraju, Alessandro Checco, and Gianluca Demartini. 2020. CrowdCO-OP: Sharing Risks and Rewards in Crowdsourcing. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW2 (2020), 1–24.

[9] Jason Furman and Robert Seamans. 2019. AI and the Economy. *Innovation policy and the economy* 19, 1 (2019), 161–191.

[10] Ujwal Gadiraju, Alessandro Checco, Neha Gupta, and Gianluca Demartini. 2017. Modus operandi of crowd workers: The invisible role of microtask work environments. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 1–29.

[11] Ujwal Gadiraju, Ricardo Kawase, and Stefan Dietze. 2014. A taxonomy of microtasks on the web. In *Proceedings of the 25th ACM conference on Hypertext and social media*. 218–223.

[12] Ujwal Gadiraju and Jie Yang. 2020. What can crowd computing do for the next generation of AI systems?. In *2020 Crowd Science Workshop: Remoteness, Fairness, and Mechanisms as Challenges of Data Supply by Humans for Automation*. CEUR, 7–13.

[13] Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé Iii, and Kate Crawford. 2021. Datasheets for datasets. *Commun. ACM* 64, 12 (2021), 86–92.

[14] Neha Gupta, David Martin, Benjamin V Hanrahan, and Jacki O'Neill. 2014. Turk-life in India. In *Proceedings of the 18th International Conference on Supporting Group Work*. 1–11.

[15] Danula Hettiachchi and Jorge Goncalves. 2019. Towards effective crowd-powered online content moderation. In *Proceedings of the 31st Australian Conference on Human-Computer-Interaction*. 342–346.

[16] Sarah Holland, Ahmed Hosny, Sarah Newman, Joshua Joseph, and Kasia Chmielinski. 2020. The dataset nutrition label. *Data Protection and Privacy, Volume 12: Data Protection and Democracy* 12 (2020), 1.

[17] Ji-Youn Jung, Sihang Qiu, Alessandro Bozzon, and Ujwal Gadiraju. 2022. Great Chain of Agents: The Role of Metaphorical Representation of Agents in Conversational Crowdsourcing. In *CHI Conference on Human Factors in Computing Systems*. 1–22.

[18] Aniket Kittur, Jeffrey V Nickerson, Michael Bernstein, Elizabeth Gerber, Aaron Shaw, John Zimmerman, Matt Lease, and John Horton. 2013. The future of crowd work. In *Proceedings of the 2013 conference on Computer supported cooperative work*. 1301–1318.

[19] Brian McInnis, Dan Cosley, Chaebong Nam, and Gilly Leshed. 2016. Taking a HIT: Designing around rejection, mistrust, risk, and workers' experiences in Amazon Mechanical Turk. In *Proceedings of the 2016 CHI conference on human factors in computing systems*. 2271–2282.

[20] Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru. 2019. Model cards for model reporting. In *Proceedings of the conference on fairness, accountability, and transparency*. 220–229.

[21] K Murrell. 2012. *Ergonomics: Man in his working environment*. Springer Science & Business Media.

[22] Jodi Oakman, Natasha Kinsman, Rwth Stuckey, Melissa Graham, and Victoria Weale. 2020. A rapid review of mental and physical health effects of working at home: how do we optimise health? *BMC Public Health* 20, 1 (2020), 1–13.

[23] Vishal Patel, Austin Chesmore, Christopher M Legner, and Santosh Pandey. 2022. Trends in Workplace Wearable Technologies and Connected-Worker Solutions for Next-Generation Occupational Safety, Health, and Productivity. *Advanced Intelligent Systems* 4, 1 (2022), 2100099.

[24] Sihang Qiu, Alessandro Bozzon, Max V Birk, and Ujwal Gadiraju. 2021. Using worker avatars to improve microtask crowdsourcing. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–28.

[25] Jeffrey M Rzeszotarski, Ed Chi, Praveen Paritosh, and Peng Dai. 2013. Inserting micro-breaks into crowdsourcing workflows. In *First AAAI conference on human computation and crowdsourcing*.

[26] Shruti Sannon, Billie Sun, and Dan Cosley. 2022. Privacy, Surveillance, and Power in the Gig Economy. In *CHI Conference on Human Factors in Computing Systems*. 1–15.

[27] Aayush Singh, Sebastian Wehkamp, and Ujwal Gadiraju. 2022. SignUpCrowd: Using Sign-Language as an Input Modality for Microtask Crowdsourcing. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 10. 184–194.

[28] Miriah Steiger, Timir J Bharucha, Sukrit Venkatagiri, Martin J Riedl, and Matthew Lease. 2021. The psychological well-being of content moderators: the emotional labor of commercial moderation and avenues for improving support. In *Proceedings of the 2021 CHI conference on human factors in computing systems*. 1–14.

[29] Carlos Toxtli, Siddharth Suri, and Saiph Savage. 2021. Quantifying the Invisible Labor in Crowd Work. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–26.