Gesticulate for Health's Sake! Understanding the Use of Gestures as an Input Modality for Microtask Crowdsourcing

Garrett Allen, Andrea Hu, Ujwal Gadiraju

Web Information Systems, Delft University of Technology, The Netherlands
G.M.Allen@tudelft.nl, hjc3299@gmail.com, U.K.Gadiraju@tudelft.nl

Abstract

Human input is pivotal in building reliable and robust artificial intelligence systems. By providing a means to gather diverse, high-quality, representative, and cost-effective human input on demand, microtask crowdsourcing marketplaces have thrived. Despite the unmistakable benefits available from online crowd work, the lack of health provisions and safeguards, along with existing work practices threatens the sustainability of this paradigm. Prior work has investigated worker engagement and mental health, yet no such investigations into the effects of crowd work on the physical health of workers have been undertaken. Crowd workers complete their work in various sub-optimal work environments, often using a conventional input modality of a mouse and keyboard. The repetitive nature of microtask crowdsourcing can lead to stress-related injuries, such as the well-documented carpal tunnel syndrome. It is known that stretching exercises can help reduce injuries and discomfort in office workers. *Gestures*, the act of using the body intentionally to affect the behavior of an intelligent system, can serve as both stretches and an alternative form of input for microtasks. To better understand the usefulness of the dual-purpose input modality of ergonomically-informed gestures across different crowdsourced microtasks, we carried out a controlled 2×3 between-subjects study (*N=294*). Considering the potential benefits of gestures as an input modality, our results suggest a real trade-off between worker accuracy in exchange for potential short to long-term health benefits.

Introduction

Artificial intelligence (AI) techniques and machine learning (ML) in particular, are drastically changing our lives through technological revolutions across several domains (Erlei et al. 2020, 2022). A primary stimulant that has led to these rapid advances in AI and ML in recent times, apart from the computational resources now available, is the design and development of crowdsourcing methods over the last decades to harness human intelligence at scale (Gray and Suri 2019). Human input is pivotal in building reliable and robust AI systems (Gadiraju and Yang 2020); it is required for data generation, evaluation, and debugging of ML models (Vaughan 2017), and plays a central role in building hybrid human-machine systems (Demartini et al. 2017). Microtask crowdsourcing marketplaces have thrived due to the growing demand for accessible, high-quality, representative, and cost-effective human input (Kittur et al. 2013). We have witnessed a growth and a continuous influx of people turning towards microtask crowdsourcing platforms around the globe, to earn a significant portion of their livelihoods (Difallah et al. 2015; Difallah, Filatova, and Ipeirotis 2018).

Despite the unmistakable benefits that can be reaped from online crowd work, issues emerging from the dynamics of platform work, the lack of health provisions or safeguards, and existing work practices threaten the sustainability of this paradigm (Sannon and Cosley 2019; Fan et al. 2020; Toxtli, Suri, and Savage 2021; Lascau et al. 2022). Prior work has tangentially addressed some of these issues by proposing different methods to increase worker engagement and improve the overall experience of crowd workers. To improve worker engagement, Rzeszotarski et al. (2013) proposed to use micro-breaks while Dontcheva et al. (2014) proposed to combine learning elements within tasks. More recently, authors proposed to use conversational interfaces, worker avatars, human and non-human metaphors to increase worker engagement and satisfaction (Qiu et al. 2021; Jung et al. 2022). Researchers have reported that both physical and mental fatigue can negatively impact crowd work (Mao, Kamar, and Horvitz 2013; Zhang, Ding, and Gu 2018), and creating enjoyable experiences can positively impact the mental health of crowd workers (Allan et al. 2018).

Prior work however, has not explored the broad implications of microtask crowdsourcing on the physical health of workers. This is despite the fact that physical discomforts and ergonomics of desk work, particularly for sitting workers, has been studied for several decades (Murrell 2012). Crowd workers are also known to have suboptimal work environments and are often embedded in multitasking contexts (Gupta et al. 2014; Gadiraju et al. 2017). There have also been reports of stress related injuries among crowd workers due to the repetitive nature of microtask crowdsourcing – for example, carpal tunnel syndrome results from repetitive tasks and specific wrist positions, leading to pressure on the wrist's median nerve and associated pain, numbness, and weakness in the hand and fingers (Patel et al. 2022). It is well-known that stretching exercises at the workplace can help in producing short-term effects in reducing musculoskeletal pain in office workers (Dubey et al. 2019). Others have shown that taking active mini-breaks for neck exercises at workstations creates a greater reduction in neck pain symptoms rather than modifying the workstation alone (Smith 2014). Several applications and health interventions have also been developed over the last decade to help users improve their workspace practices and health awareness. However, behavioral change is a well-documented challenge across several domains (Stawarz and Cox 2015). This is exacerbated by the whimsical dynamics of crowd work, leaving few opportunities for workers to prioritize their health. Workers who are striving to complete more tasks and maximize their monetary earnings in the immediate present may not be sufficiently incentivized to delve into practices that can bear health returns in the future.

We draw inspiration from working ergonomics research and envision the dual role of novel input modalities that can simultaneously serve as exercises to benefit crowd workers' health while acquiring input from workers on different tasks. An example of this can be seen in Figure 1. For the purposes of this work, we adopt the definition of a "gesture" as introduced by Carfì and Mastrogiovanni (2021): "*Gestures are body actions that humans intentionally perform to affect the behavior of an intelligent system*." As a first step towards that end, in this paper, we assess the trade-offs of using gestures as an input modality for crowdsourced microtasks. The choices of gestures in this work are informed by literature in the fields of workplace ergonomics, physical therapy, and medical sciences (Lee, Park, and Kim 2013; Vuletic et al. 2019; Carfì and Mastrogiovanni 2021). We thereby aim to address the following research questions:

> **RQ#1**: How effective and efficient are ergonomically informed gesture-based input modalities for microtask crowdsourcing?
> **RQ#2**: How do workers perceive ergonomically informed gesture-based input modalities in different types of crowdsourced microtasks?

We carried out a 2×3 factorial between-subjects study with 294 participants recruited from the crowdsourcing platform Prolific.[1] Participants completed one of three microtasks, while using one of two different input modalities depending on the experimental condition – one with gestures as an input modality and another using the conventional keyboard and mouse as input. We have participants complete a pre-task survey for the purpose of gathering demographic information. The pre-task survey also allows us to get a view of how each participant perceives the comfort and health levels of their work environment, and specific parts of the body. Further, we employed a post-task survey to assess the worker engagement, their perceived cognitive load, and determine whether workers perceived any health benefits while using gestures for providing their input. Performance measures such as accuracy and task completion time to complete the task were also collected. Results show comparable performance for both input modalities, yet workers find the standard input more usable while gestures are viewed as more rewarding. Our findings indicate that with potentially reasonable trade-offs gestures can serve as an alternative input modality for microtask crowdsourcing.

[1]https://prolific.co

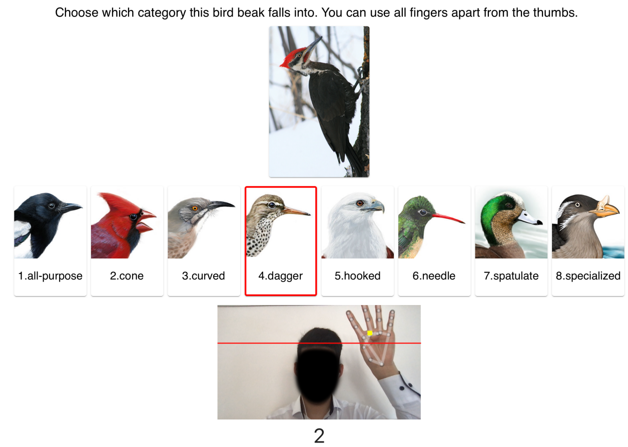


Figure 1: An example of a task with `Gesture` input that includes the question prompt, the possible answers, the chosen answer, the webcam view and a countdown timer.

Original contributions. Through our work, we created a gesture capture pipeline, introducing a novel, modular pose classification strategy that enables flexible, real-time gesture capture. Using this new pipeline, we found that crowd workers can execute microtasks accurately using a novel input modality of ergonomically-informed gestures (with a drop in accuracy of 6% in comparison to a conventional baseline). However, the task completion time of workers was not significantly different when using gesture-based input. We reveal a viable trade-off between the choice of the input modality and the promise of short to long-term health benefits for workers. In our study, workers did not view input modalities in terms of their health, or how such inputs may impact their health. A potential application of using gestures is the creation of health-driven interventions where crowd workers complete tasks using ergonomically informed gestures (Luttmann, Schmidt, and Jäger 2010; Mehrparvar et al. 2014) to reduce musculoskeletal discomforts. Our findings have important design implications for promoting crowd worker health and well-being, as a pivotal step towards ensuring the sustainability of the paradigm.

All data and code corresponding to our work, along with supplementary material can be found on the Open Science Framework companion page.[2]

Related Work

Ergonomics, Health, and Crowd Work

Ergonomics is the relationship between a worker and their environment (Murrell 2012; Woo, White, and Lai 2016). Investigations into the working environments of crowd workers discovered that workers work in a large variety of environments, with different access to hardware and software needed to complete microtasks (Martin et al. 2014; Gupta

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

et al. 2014; Gadiraju et al. 2017). Such differences in environments result in different interactions with tools, which can in turn have varying effects on workers' physical health (Ranasinghe et al. 2011; Woo, White, and Lai 2016). Following a study on work activity and muscular complaints, Luttmann, Schmidt, and Jäger (2010) concluded that workers should make an effort to change the pace and sequence of their work in order to decrease muscular activation throughout the day, thereby reducing muscular discomfort. Lee, Park, and Kim (2013) investigated the effects of neck exercises on high school students, focusing on the neck and shoulder area. Findings indicated that strengthening of the cranio-cervical flexor muscle has significant effect towards improving neck and shoulder posture. A broader study of stretches for the neck, shoulders, lower back, hands and wrists found that complaints of musculoskeletal discomfort reduced when workers performed the stretches (Mehrparvar et al. 2014). Converting targeted stretches that benefit muscles used in working environments into interpretable gestures can potentially lead towards maintained performance while also benefiting the health of workers.

Crowdsourcing researchers have demonstrated that fatigue, both mental and physical, can negatively affect crowd worker performance (Mao, Kamar, and Horvitz 2013; Zhang, Ding, and Gu 2018). Mental health is a growing topic of import within society, as it relates to people worldwide (Organization 2001). Previous studies in office work have investigated mental disorders and treatment techniques, such as meditation and relaxation therapies (Loewenstein 1991; Montero-Marin et al. 2019; Stein 2001; Tariq et al. 2006). As a means to assess the overall and mental health, the SF-36 survey has been extensively used in studies (Ware et al. 1993; Ware Jr 2000). The survey contains two sub-scales for measuring mental health, focused on mental well-being and energy/fatigue. Within crowdsourcing marketplaces, workers need to invest a large amount of time and effort into underpaid tasks due to factors such as power asymmetry and invisible labor (Gray and Suri 2019; Irani and Silberman 2013; Martin et al. 2014; Toxtli, Suri, and Savage 2021; Sannon and Cosley 2019). Such factors influence not only mental well-being of crowd workers, but also their psychosocial condition. To assess the psychosocial condition of crowd workers and their environments, the Copenhagen Psychosocial Questionnaire (COPSOQ) has been widely applied (Pejtersen et al. 2010). To the best of our knowledge, neither of these surveys has been applied in a crowdsourcing scenario using gestures as input, a gap which we address in this work.

Gestures as Input

Gestures as a mechanism for interaction between humans and machines is well researched, with many taxonomies defining the different gestures being presented. In some works, gestures are divided into "vision-based" and "sensor-based" (Al-Shamayleh et al. 2018; Cheok, Omar, and Jaward 2019; Sarma and Bhuyan 2021). Sensor-based solutions often take the form of wearable devices, such as smartwatches or gloves. Adams et al. (2018) introduced Keppi, a pressure-sensitive stick, that users can squeeze to self-report pain levels. Vision-based solutions rely on cameras, such as built-in webcams or externally connected devices (Kim et al. 2015; Shin and Kim 2017; Wachs et al. 2011). Quek (1995) separates gestures into "communicative" and "manipulative". Communicative gestures, as the name implies, are used to communicate. Manipulative gestures are used to interact with objects or systems and create an effect, such as moving an object. Each of these larger classifications have been further subdivided over time. Stern, Wachs, and Edan (2008) introduced the idea of conversational and control gestures. Conversational gestures are those that occur during speech activities. Control gestures are command-oriented gestures meant to give instructions. Further variations of a gesture taxonomy include those by (Karam and m. c. schraefel 2005) and (Rojas-Muñoz and Wachs 2019), which include Butterworth's, semaphoric, and linguistic gestures among those previously discussed. In an effort to provide uniformity across the many taxonomies, Vuletic et al. (2019) performed a systematic literature survey. Inspired by the survey conducted by Vuletic et al. (2019), Carfì and Mastrogiovanni (2021) performed a literature review to define what a gesture is in a concrete way, and also introduce a more flexible taxonomy of gestures. Gestures are defined as "body actions that humans intentionally perform to affect the behavior of an intelligent system". While there are many variations of gestures for human-machine interaction, none are explicitly defined for crowdsourcing microtasks. In this work, we address this gap by proposing gestures as an input modality for microtask execution.

Study Design

In this study, we aim to understand the efficacy of gestures (referred to henceforth as `Gesture`) as an alternative input modality to the conventional means of using a keyboard and mouse (Gadiraju et al. 2017) (referred to henceforth as `Standard`). We seek to understand what trade-offs may exist between the two modalities in terms of effectiveness and efficiency, and we also aim to understand how crowd workers perceive gestures. Due to the variety in task types and the potential interaction of task types with input modalities, we additionally consider different task types in our exploration. Towards these ends, we conducted a controlled 2×3 between-subject crowdsourced study. The independent variables for the study are the input modality (`Gesture`, `Standard`) and task type (sentiment analysis, classification, and information finding), resulting in a total of six experimental conditions. An example of what one of these experimental conditions looks like can be seen in Figure 1. For each condition, crowd workers are asked to complete a survey before and after their interaction with the microtask.

Surveys and Measures

Prior to completing the microtask, each worker is asked a set of 16 questions intended to gather demographic information, *e.g.*, their age, their mood, their experience on Prolific, how much and when they tend to work on the platform, and their earnings. To ensure that we gather high-quality responses, an attention check question was included in this survey (Gadiraju et al. 2015). Additionally, we inquire as to how workers are feeling, sectioned off into specific parts of the body. The demographic data affords us contextual information that is helpful in interpreting the responses to the post-task survey.

We aim to understand the trade-offs of using gestures (**`Gesture`**) as an input modality on worker engagement in comparison to using the conventional input modality of a keyboard and mouse (**`Standard`**). To capture engagement, we included six questions from the short form of the User Engagement Scale (UES-SF), specifically from the "perceived usability" and "reward" scales (O'Brien, Cairns, and Hall 2018).

Furthermore, we explore whether gestures as an input modality lead to a similar cognitive load, while still enabling workers to complete their tasks effectively and efficiently. To measure this, we include the six questions from NASA's Task Load Index (TLX) (Hart 2006). Finally, to explore whether using gestures results in any perceived health benefits from the standpoint of workers, we implemented questions from the COPSOQ (Pejtersen et al. 2010) and the SF-36 survey (Ware et al. 1993; Ware Jr 2000).

Considering the survey responses together with the accuracy of responses, *e.g.*, how many were responded to correctly, enables us to assess the effectiveness of the **`Gesture`** input modality. We also want to understand how **`Standard`** and **`Gesture`** inputs compare in terms of efficiency. Therefore, we also record how long it takes for workers to complete a micro task, excluding the time spent on pre- and post-task surveys.

Tasks Design

We selected three task types as being well-suited for completion via gesture input, from the taxonomy of microtasks introduced by Gadiraju, Kawase, and Dietze (2014). The tasks we selected were: Sentiment Analysis, Classification, and Information Finding.

Figure 2: Interface for the Sentiment Analysis task.

The *Sentiment Analysis* microtasks (Figure 2) involve workers reading a review for a movie or TV show and assigning the star rating they deem most appropriate based on the sentiment of the review content. The reviews were sampled from the Amazon Review Dataset (Ni, Li, and McAuley 2019) such that there were two reviews for each rating level, from 1–5 stars, for a total of 10 reviews. Additionally, these reviews were limited to a maximum of 150 words to ensure a coherent and consistent display, and to maximize the ease of reading for workers.

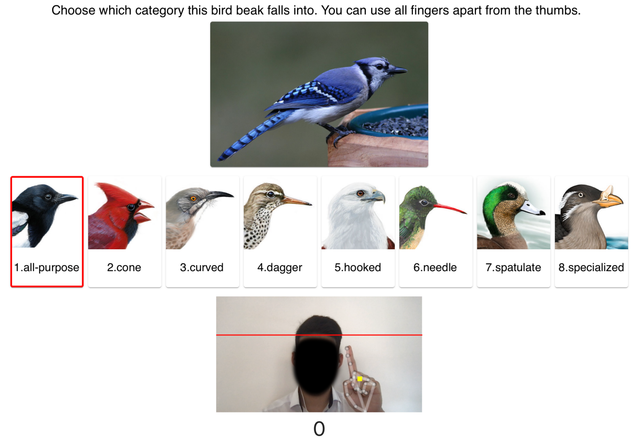


Figure 3: Interface for the Classification task.

The *Classification* microtasks (Figure 3) ask workers to classify the beak shape of a bird in an image using a set of example images for 8 different beak shapes. A series of 10 images are shown, sampled from the dataset originally produced by Balayn et al. (2022). For each image, the worker is asked to select the beak type from the candidates that matches the beak type in the image.

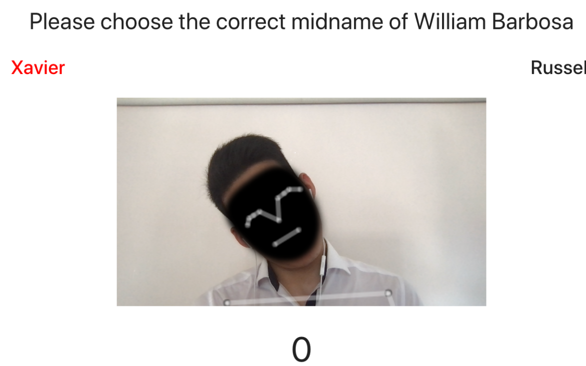


Figure 4: Interface for the Information Finding task.

Finally, the *Information Finding* microtasks (Figure 4) follow a similar design to the information finding tasks designed in (Gadiraju and Dietze 2017). This task involves workers being presented with the name of a famous person. Using the name and any search tool at their disposal, e.g., Google, Bing, or Wikipedia, workers are tasked with finding the middle name or profession of the famous person. In total, there are 10 names presented: five requiring searching for the profession and five for the middle name.

Combining the input modalities of **`Gesture`** and **`Standard`** and the three types of microtasks, we designed six separate experimental conditions: (*i*) Gesture-Sentiment, (*ii*) Standard-Sentiment, (*iii*) Gesture-Classify, (*iv*) Standard-Classify, (*v*) Gesture-Finding, and (*vi*) Standard-Finding.

When designing the gestures for each task, we made sure

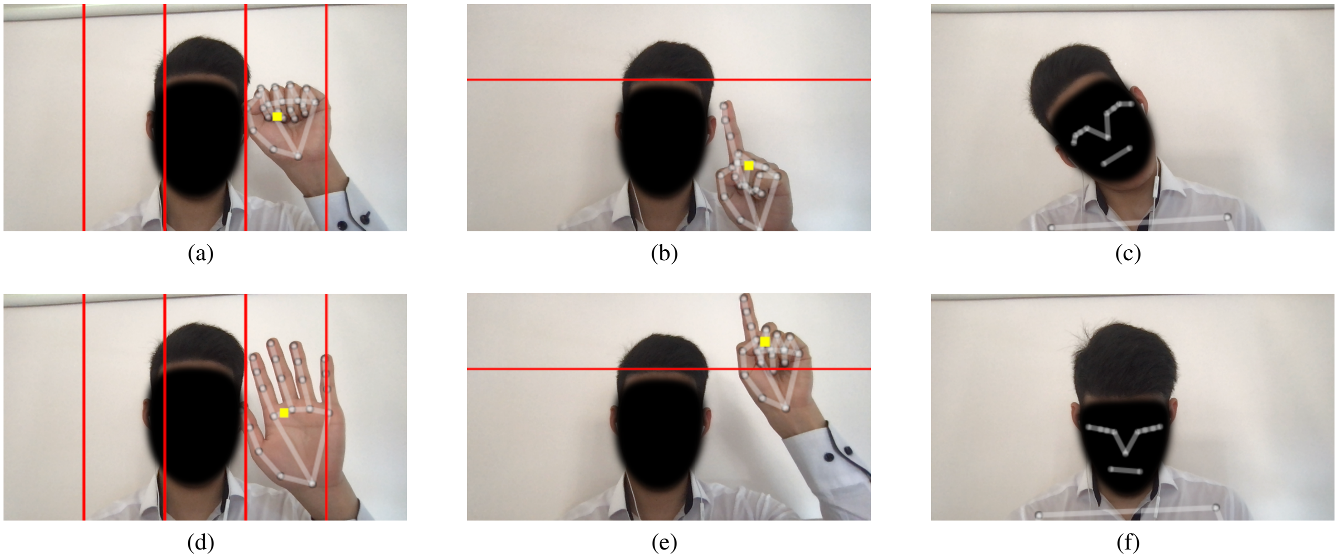


Figure 5: Example screenshots of gestures for selecting an answer. First row, left to right: (a) Sentiment, closed fist moving left/right to choose a rating. (b) Classification, number of extended fingers in the bottom 2/3 of view to choose an answer option. (c) Finding, tilting head left or right to choose between two options. Second row, left to right: (d) Sentiment, open fist to submit rating. (e) Classification, hand in top 1/3 of view to submit answer. (f) Finding, lowering head to submit answer.

that separate actions are used for selection and submission of an answer for both **`Standard`** and **`Gesture`** input modality, keeping them consistent. For the **`Standard`** tasks, it means that two mouse clicks are required to submit an answer: one for selecting the answer and one more for submitting. Similarly for **`Gesture`**, two actions are required to choose an answer and to submit it. The answer selection gesture for all tasks has immediate effect, whilst the confirmation gesture needs to be held for 2 seconds for the answer to be submitted. When the answer is changed or the confirmation gesture is interrupted, then the 2 second countdown will be reset, allowing workers to change the answer.

In the sentiment analysis task, the answer can range between 1–5, thus the camera view is divided into 5 columns; with the leftmost column representing 1 to the rightmost column representing 5. When a hand enters the view, the centroid of its landmarks are denoted by a yellow square. With a closed fist, workers can move the hand left and right and change the selected answer based upon which column the hand centroid falls into (Fig. 5(a)). The confirmation gesture is an open hand, which needs to be held for longer than 2 seconds (Fig. 5(d)).

In the classification task, there are 8 bird classes to choose from, each denoted with a number from 1–8. To select an answer, the workers have to raise the respective number of fingers with one or both hands (Fig. 5(b)). The finger count is independent from the combination of fingers raised. All fingers apart from the thumb are taken into account. We opted to ignore the thumb because during testing we have found the recognition of thumb extension to be unreliable. To confirm the answer, any of the visible hands' centroid has to be raised above the line drawn on the camera view (Fig. 5(e)).

In the information finding task, there are only two answers; one on the left and one on the right. To select an answer, the worker has to tilt their head in the direction of the chosen answer (Fig. 5(c)). Once chosen, the worker can return to the neutral head position and lower the head to confirm the answer (Fig. 5(d)).

Examples of all the described gestures can be found in Figure 5. Although not shown in the figure, extra components are displayed on the task page to provide immediate feedback to the worker of their actions, such as the selected answer and the countdown timer. These can be found in the task screenshots, available on the companion page.

To limit familiarity biases, we included a tutorial stage for all task types. All tutorials include a text description of the task and instructional text. The classification task additionally includes example images of birds so workers can familiarise with the bird classes first. For the three **`Gesture`** task tutorials, workers also have to submit two specific answers with gestures before proceeding to the actual task. These are: answers 4 and 1 for the sentiment task, 5 and 8 for the classification task and left and right for the finding task. These answers were chosen specifically to confirm that workers have understood the gesture inputs correctly.

Participants and Quality Control

We carried out a power analysis using the *G*Power* tool (Erdfelder, Faul, and Buchner 1996), to determine the required sample size in our controlled study, resulting in 279 participants. To account for potential exclusion of data, we recruited 50 participants for each of the 6 experimental conditions, resulting in a total population sample of 300. Participants were recruited from the Prolific crowdsourcing plat-

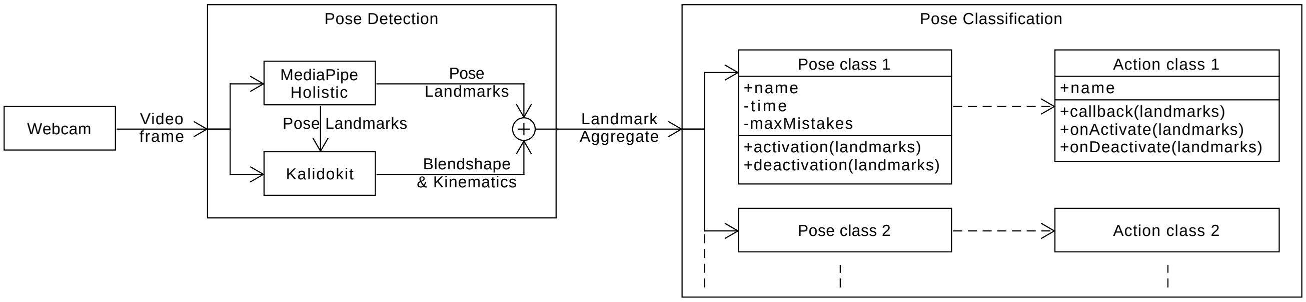


Figure 6: Simplified structural diagram of the gesture recognition pipeline we implemented.

form, while enforcing a set of pre-screening filters to ensure reliability of responses. We restricted participation to workers who were both fluent in English (since the task instructions and microtasks were entirely in English), and had an approval rating of 90% or above on Prolific. We also ensured that workers could not participate in more than a single experimental condition in our study. All workers were required to complete an informed consent form prior to participation. Upon completing an experimental condition, each worker was compensated at a rate of £8.00 per hour. Our study also received an institutional ethics approval.

Technical Implementation

All six experimental conditions were hosted on a custom developed website written using the React library, connected to a back-end server written using the NestJS framework. The back-end API manages the progress of each worker and the front-end is responsible for rendering the pages and handling gesture, mouse, and keyboard inputs. The communication between the front-end and back-end consists of an HTTP REST API with all the worker data being stored in a MongoDB NoSQL database. The input of gestures is enabled by two consecutive client-side stages: pose detection and pose classification, which is concisely represented in Figure 6.

Pose Detection The pose detection stage detects the body, face and hand landmarks of the worker on each frame. It utilises two libraries to do so: MediaPipe holistic[3] and Kalidokit[4]. First, the webcam video feed is passed to MediaPipe holistic, which uses multiple pre-trained models to perform real-time pose landmark estimation of the body, face and hands. We augment the result landmarks using Kalidokit, a utility library which converts them into more directly interpretable data such as simple Euler rotations of joints and blend shape face values. The combination of these two pieces of data is utilized to classify poses.

Pose Classification The pose classifier is defined by three components: a set of pose classes, a set of action classes (side effects) and the mapping between the two. A pose class, in part inspired by the "Repetition Counting" section in this Google solution[5], is described by the entry condition (activation), exit condition (deactivation), duration and maximum number of mistakes. When the classifier receives pose data from the previous stage, it passes them onto all the pose classes. Each pose class then evaluates its entry condition, which if truthful then the class enters the active state and saves the current timestamp. Once a pose class enters the active state, it will stay so until the exit condition evaluates to true and deactivate. Instead, if the class stays in the active state for longer than the duration time (which can be 0), then a callback function is triggered. An action class describes three functions: callback, onActivate and onDeactivate. These are called respectively when the duration of an active class elapses, when the class is activated and when it is deactivated. For example, we use the callback function to change the answer choice or to submit the answer, and the onActivate and onDeactivate functions to start and reset the timer. The separation between gesture classes and their actions allows for a modular setup and the ability to dynamically assign a gesture to an action.

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 enters a certain range of the webcam frame or if an angle of a joint exceeds a certain degree. Furthermore, by using the fingerpose package,[6] it is also possible to evaluate finger gestures, as we show in the classification task.

To protect the workers' privacy, we made sure to not collect any personally identifiable information; thus all user data is only linked to their prolific id, including survey responses, question answers, and pose data. Participants also have the possibility to revoke their consent whenever they want during the entirety of the study; once revoked, any data that was generated from that participant is then removed. Furthermore, the webcam images used for gesture recognition are shown and processed all on the participants' device and never sent to the back-end or stored anywhere. We do collect pose landmarks on some actions, but they are not per-

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

[5]https://google.github.io/mediapipe/solutions/pose_classification.html#repetition-counting
[6]https://github.com/andypotato/fingerpose

sonally identifiable information.

Results and Analysis

With this study we explore two key factors. First, we want to determine if gestures can serve as a viable alternative input strategy to a mouse and keyboard for crowdsourced microtasks. Second, we want to know how crowd workers perceive the gestures. To enable fair comparison between conditions, we recruited 50 participants per experimental condition. Across all conditions, we had to remove six participants for failing two attention checks. We removed one additional participant for clicking through the post-task survey, *i.e.*, providing the same answer to all questions on the post-task questionnaire in a short period of time. Finally, we removed two responses that were duplicated. As a result, we were left with a total of 294 responses. The distribution across experimental conditions can be seen in Table 1.

Experimental Condition	**#Participants**
Gesture-Sentiment	46
Standard-Sentiment	49
Gesture-Classify	51
Standard-Classify	49
Gesture-Finding	49
Standard-Finding	50

Table 1: Number of participants per experimental condition.

Worker Demographics and Background

We required participants to complete a pre-task survey before interacting with a microtask to gather demographic-related data. Figure 7 shows this demographic information including age, gender, experience, and self-reported health factors of the crowd workers. The majority of workers are aged 18–35 (89%), while the remaining are aged 36–65 (11%). Of the 294 recruited participants, 157 were female and 135 male, one preferred not to specify, and one reported as non-binary or other. The experience of the workers varied with most reporting having between six months and one year of experience on Prolific, as seen in Figure 7(e). Additionally, the workers reported working ten hours or less each week (Figure7(g)). This is reflected in the yearly income shown in Figure 7(f) with most respondents making $25,000 and under from their crowd work.

When reporting the comfort and health level of their work environments, the crowd workers tend to consider their environments somewhere between slightly comfortable/healthy and neither comfortable/healthy or uncomfortable/unhealthy. As seen in Figure 7(d), similar results were reported for the comfort level of the eyes and the neck and shoulders. Knees and feet were reported as the most comfortable, with the back being the least comfortable.

Efficiency and Effectiveness

Gaining a full picture of how well gestures serve as an alternative input modality requires examining both the effectiveness and the efficiency workers exhibit while using the modality. To this end, we measure the accuracy of workers for each task, i.e., how many of the 10 tasks they completed correctly, and how quickly workers complete the tasks.

Experimental Condition	**Accuracy (%)**
Gesture-Sentiment	36.74
Standard-Sentiment	41.22
Gesture-Classify	61.37
Standard-Classify	68.98
Gesture-Finding	76.73
Standard-Finding	82.8

Table 2: Accuracy of the crowd workers across different experimental conditions.

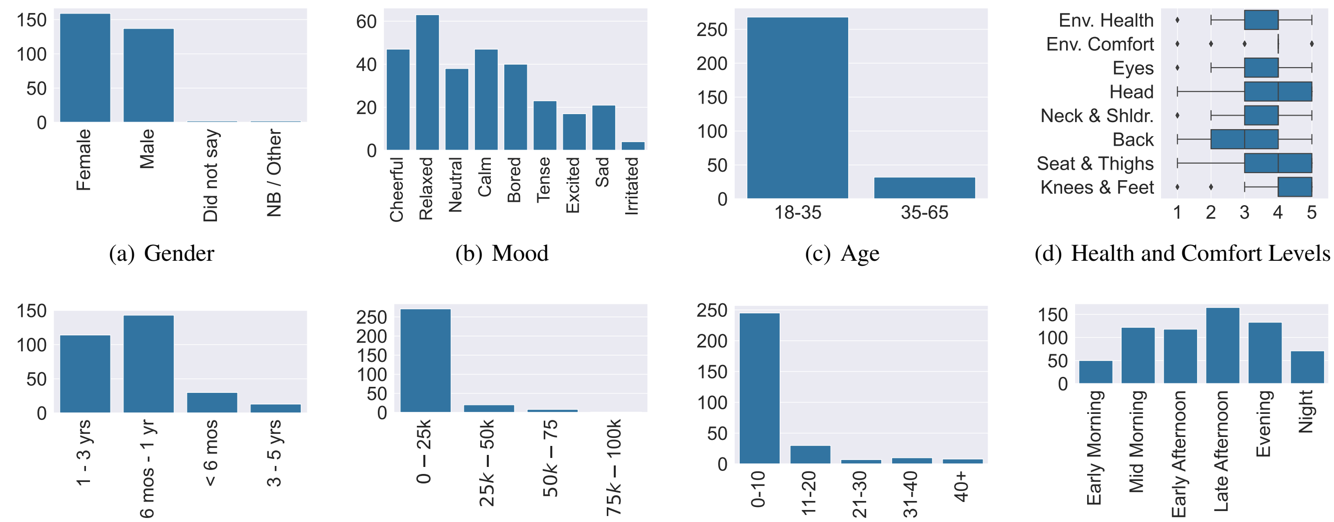
In exploring the effectiveness of gestures, we measure the accuracy across all workers for each experimental condition, reported in Table 2. The results show a significant main effect for the input modality, *$F(1, 288)=9.56$, $p=.002$* such that workers achieved higher accuracy in tasks with the **`Standard`** input modality (*$M=.64$, $SD=.15$*) in contrast to those with the **`Gesture`** input modality (*$M=.58$, $SD=.18$*). The main effect of task types was also significant, *$F(2, 288)=146.53$, $p<.001$*. To see where the significance resides with respect to the three task types, we implement a post-hoc Tukey test. We found significant differences between all combinations of task types. The interaction effect between the input modality and task type was not significant.

As an indicator of efficiency, we measured the time it takes workers to complete the task in each experimental condition. Note that the time spent on pre- and post-task surveys is not included in this measurement. We found that workers spent more time to complete the tasks when using the **`Standard`** input modality (*$M=213.82$, $SD=92.01$*), in contrast to the **`Gesture`** modality (*$M=207.04$, $SD=113.63$*). However, this difference was not statistically significant.

Worker Perceptions

Aiming for a broader overview of the effectiveness beyond accuracy, we investigated workers' perceptions. Depending on the underlying data and dependent variables of interest, we conducted one-way/two-way ANOVAs or the Kruskal-Wallis H test. To correct for Type-I error inflation, we report statistical significance with respect to adjusted *p*-values.

Through a two-way ANOVA, we explored the impact of input modality and task type on the perceived usability among workers. The results show a significant main effect for the input modality, *$F(1, 288)=13.26$, $p<.001$* such that workers perceived a higher usability in tasks with the **`Standard`** input modality (*$M=4.06$, $SD=.66$*) in contrast to those with the **`Gesture`** input modality (*$M=3.71$, $SD=.96$*). The main effect of task types was also significant, *$F(2, 288)=13.72$, $p<.001$*. A post-hoc Tukey test uncovered significant differences between the *classification* and *information finding* tasks. Significance between the *sentiment analysis* and *information finding* task types was also found. The interaction effect was not significant.



(a) Gender (b) Mood (c) Age (d) Health and Comfort Levels

(e) Experience (f) Microtask yearly income (g) Weekly hours (h) Working times

Figure 7: Distribution of workers across the demographics of age, gender, experience on Prolific, income from crowd work, and self-reported health and comfort.

We also investigate how the workers perceived the reward of completing the microtasks with different inputs. We found a significant main effect for the input modality, *F(1, 288)=6.33, p=0.01* such that workers perceived a higher reward in tasks with the **`Gesture`** input modality (*M=3.8, SD=.78*) in contrast to those with the **`Standard`** input modality (*M=3.59, SD=.69*). The main effect of task types was also significant, *F(2, 288)=3.19, p=0.042*. A post-hoc Tukey test of task types revealed no significant difference. The interaction effect was not significant.

Through a two-way ANOVA we investigated the impact of input modality and task type on the cognitive load of workers. We found a significant main effect for the input modality, *F(1, 288)=28.49, p<0.001* such that workers experienced lower cognitive load when using the **`Standard`** input (*M=1.82, SD=0.49* versus when using the **`Gesture`** input (*M=2.17, SD=0.61*). The main effect of task type was also significant, *F(2, 288)=10.61, p<0.001*. A post-hoc Tukey test uncovered significant differences between classification and finding as well as between the sentiment and finding task types. The interaction effect was not significant.

We explored the impact of input modality and task type on the self-reported general health of workers with a Kruskal-Wallis H-test. We found the main effect for the input modality was significant, (*H(1)=4.59, p=0.03*), such that workers report better general health when using the **`Standard`** input modality (*M=2.51, SD=.94*) over the **`Gesture`** modality (*M=2.28, SD=.94*). Neither the main effect for task type nor the interaction effect were significant.

Looking into the impact of the input modality and task type on the work pace of workers, we found that the main effect for input type was not significant. On the other hand, the main effect for task type was significant, *H(2)=6.05, p=0.049*. Applying a Dunn post-hoc test for the task types, we find that there is a significant difference between the sentiment and metadata task types, such that workers feel the need to work at a faster pace when completing the sentiment analysis task (*M=3.25, SD=.87*) in contrast to the metadata task (*M=2.93, SD=.94*).

There were no statistically significant effects found in our exploration of the main effects of input modality and task type on the emotional demands, emotional well-being, quantitative demands, and perceived error of workers.

Finally, we wanted to learn how workers perceive their success rate at completing the various microtasks, thus we asked them to estimate how many of the 10 questions they believe they answered correctly. Using the Kruskal-Wallis H-test, we found a significant main effect for the input modality, *H(1)=7.88, p=0.005* indicating that workers perceive themselves to be more accurate when using the **`Standard`** input modality (*M=7.56, SD=1.62*) versus when using the **`Gesture`** modality (*M=6.98, SD=1.85*). We also found a significant main effect for the task type, *H(2)=22.02, p<0.001*. A post-hoc Dunn test revealed that workers completing the classification task have significant difference in their perceived accuracy when compared with those that completed the sentiment analysis task. Workers that completed the sentiment analysis task also estimated significantly different accuracy when compared to those that performed the information finding task.

In summary, we investigated the effectiveness and efficiency of `Gesture` inputs for crowdsourcing microtasks and found no significant difference in the efficiency of workers in comparison to the `Standard` input method. We did discover significant main effects from the input modality for perceived usability, perceived reward, cognitive load, self-reported general health, and for perceived accuracy. Workers found `Standard` to be more usable and to require less cognitive effort. They also estimated more accurate performance and perceived themselves to have better general health while using the `Standard` input. Of note is workers found `Gesture` to be a more rewarding input modality. We additionally found significant main effects from the task types in the same measures, as well as for work pace.

Discussion and Implications

In an increasingly complex landscape of crowdsourcing marketplaces, sweeping changes driven by policy or platform regulations require the benefit of time. We argue that crowd worker health is an urgent issue that cannot afford the privilege of time. Thus, we envision gestures as a dual-purpose input modality, that can allow workers to complete their tasks while also serving the alternative purpose of providing short to long-term health benefits through the types of movements used. However, the first step towards that goal is to fully understand the trade-offs concomitant with using gestures as an input modality for microtask crowdsourcing — the main objective of this study.

The Usability of Gestures. When prompted on usability across the experimental conditions, workers responded that `Standard` inputs were more usable. This comes as no surprise, given the ubiquitous presence of the modality. In addition to prevalence, `Standard` inputs have accrued a large selection of accessibility support systems, such as stabilization for tremors and zoom options for those with poor visual acuity. Yet, `Gesture` inputs were viewed as being more rewarding despite leading to higher perceived cognitive loads. The increased cognitive load can be a result of workers' lack of familiarity with the novel input modality. For each gesture task, we provided a short training at the beginning to demonstrate the gestures, provided workers with an opportunity to familiarize themselves with the modality, and practice the gestures. This was intended to limit the impact of familiarity. However, a brief training session may not always be sufficient to even out workers' prior use and experience with `Standard` inputs.

Trade-offs When Using Gestures. Workers performed with a higher accuracy when using the `Standard` input modality (64%) in contrast to the `Gesture` modality (58%). Despite generating slightly more accurate responses (around 6% better), workers spent nearly identical amounts of time completing tasks across both modalities. Considering the potential benefits of gestures as an input, our results suggest a realistic trade-off between worker accuracy in exchange for potential short to long-term health benefits. The performance differences measured in this study may be offset over time as workers become more familiar with the `Gesture` modality, reducing the cognitive effort needed. Having said that, repetitive gesture-based input may be just as damaging as the existing standard. Therefore, we envision the role of gesture-based input as one that can be used at intervals or sporadically, either in a push-based fashion (where crowd workers are required to complete specific tasks using gestures) or a pull-based fashion (where crowd workers can opt to use gestures to complete tasks), at the discretion of the task requester or crowdsourcing platforms. Further research is required to best understand how we can integrate gestures as an input modality in everyday crowdsourcing marketplaces to maximize the health benefits of crowd workers, and increase the sustainability of the paradigm.

Barriers to Participation. Gestures as an input modality also have the potential to lower the barriers to participation in crowd work. Narula et al. (2011) highlighted that microtask crowdsourcing platforms are often inaccessible to workers in developing countries, and proposed a mobile-crowdsourcing platform for OCR tasks, to lower the barrier to participation. Khanna et al. (2010) studied usability barriers that were prevalent on Amazon's Mechanical Turk (AMT), which prevented workers with little digital literacy skills from participating and completing work on AMT. In this context, using intuitive gestures that crowd workers are familiar with as a means for input acquisition, offers a great potential for lowering barriers to participation and easing the onboarding of novice workers to crowd work.

Mitigation of Cognitive Biases. Performing research involving crowd workers has the potential to introduce numerous cognitive biases depending on task design and workflow. If ignored, such biases can have negative effects on study results. Therefore, we analyzed our experimental design by using the Cognitive Biases Checklist introduced by Draws et al. (2021) and took steps to mitigate any biases found.

Self-interest bias is possible due to the monetary compensation of crowd workers we recruited from the Prolific crowdsourcing platform. To mitigate this, we rejected a single submission that showed obviously low effort, *i.e.,* same answer for every survey question and an unreasonably short completion time – a fair rejection as described by Gadiraju and Demartini (2019). The comparative nature of the study comes with the potential for an affect heuristic, in the form of a *familiarity bias*. Workers will have a higher level of familiarity with the `Standard` input, therefore we include and provide time for a tutorial to workers using the `Gesture` inputs. Through clear instructions and a detailed task description, we attempt to address the presence of *optimism bias* by ensuring the workers are as informed as they can be before selecting our task. On the other side of this coin, the *sunk cost fallacy* is also potentially present in our study. We conducted a small pilot study to get an informed estimate of how long each task would take in order to minimize the effect. Finally, there is the possibility of *disaster neglect*, or workers being improperly informed of the consequences of the task. The requirement of workers to complete an informed consent form addresses this bias.

Caveats and Other Considerations. In our task instructions to workers in the `Gesture` experimental conditions, we did not mention the health-inspired motivation behind the gestures to avoid priming workers towards favorable perceptions. Therefore, through the questionnaires focused on worker engagement and perceived health benefits, we found that crowd workers gauge microtasks from the perspective of completing the tasks, and not necessarily from the angle of how gestures may help their health.

It is important to understand that gesture-based input may not equally benefit all crowd workers. Crowd workers may possess a variety of ailments that can impact their ability to perform tasks (Uzor et al. 2021). Individuals with existing physical ailments may be ill-suited to provide gesture-based input. However, if crowd workers are given the agency to self-select into specific input modalities such mismatches can be avoided. We also envision the further development of an "*inventory of gestures*" that can be used to suit different worker preferences and capabilities, creating opportunities to enhance worker experiences through personalization.

There are further considerations from a human-computer interaction perspective regarding our task interface. In our task design, we have provided feedback that immediately reacts to the worker's actions, but this feedback is only visual in nature. To improve the usability of the interface, further feedback of other types such as acoustic and intermediary can be added. For example in the information finding task, where workers have to open a second tab to search for information, audio cues can help them input their answers even when the browser tab containing the task is not in focus. Similarly, workers do not have intermediary feedback on how close they are to the answer selection and submission gestures, namely to what degree to tilt their head or how much to lower their chin to submit. Feedback on these factors can potentially increase the overall usability of `Gesture` inputs.

Limitations. In our work, we did not explore the role of task complexity (Yang et al. 2016) in shaping the quality of responses when using gesture-based input. It is likely that more complex tasks cannot easily be transformed to suit gesture-based input acquisition. This can be particularly true in the absence of dedicated tools that can help requesters design tasks for gesture-based input. As much as this is a challenge, it is a vital opportunity for future research targeting worker health and well-being. Our findings can inform the design of software tools that can help automatically transform tasks to suit gesture-based input from crowd workers.

Conclusions and Future Work

Crowdsourcing marketplaces are continuing to thrive nearly two decades since Amazon's MTurk broke a path for microtask crowdsourcing. Worker health and well-being, however, has received limited attention from researchers and practitioners despite being essential for the sustainability of the paradigm. It is well-known that many crowd workers around the world deal with sub-optimal work environments and harrowing work dynamics, that leave them straining to earn their livelihoods. Workers however are the most important actors in the ecosystem. To improve workers' physical and thereby mental health, we envision a dual-purpose input modality that workers can use to (a) execute microtasks, and one that is informed by workplace ergonomics and can thereby (b) have a positive influence on worker health in the short to long-term. With that vision in mind, we introduce a novel, modular gesture capture pipeline that is flexible and works in real time. With this pipeline, we carried out a controlled study to better understand the efficiency and effectiveness of ergonomically-informed gestures as a dual-purpose input modality for crowdsourced microtasks. We found that workers can efficiently execute microtasks using gestures with an accuracy that is on average around 6% below a conventional baseline without any difference in the task execution time. Across several types of microtasks, workers perceived the conventional input modality of using a mouse and keyboard as being more engaging and less cognitively taxing. Workers also tended to view input modalities in light of how the inputs will help them complete tasks, rather than from the perspective of the health benefits that they may provide. Through an analysis of various worker perceptions that shed light on the effectiveness of this novel input modality, we share insights leaving important implications for the design of gesture-based input for crowd work.

As a result of the findings in our work, we have uncovered important questions that need to be considered in the future. Our task design utilized gestures that focused on the head and hands, using small movements. Such physical activity can have a positive effect on one's mood if done in moderation (Peluso and De Andrade 2005). Zhuang and Gadiraju (2019) explored the effect of a worker's mood on their performance and perception of microtasks, finding that there is an effect on the perception of engagement. Is it possible to create a sustainable, beneficial circle of health benefits and positive moods through gestures as inputs? Further studies targeting these questions can advance our understanding of the potential physical benefits for crowd workers and improve the sustainability of crowd work. In our study, workers reported using more cognitive effort with gestures. In a longer-term study, allowing workers more time to familiarize themselves with the gesture inputs would elucidate the varying cognitive load requirements of gestures in comparison to the conventional mouse and keyboard input.

Acknowledgements

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

Adams, A. T.; Murnane, E. L.; Adams, P.; Elfenbein, M.; Chang, P. F.; Sannon, S.; Gay, G.; and Choudhury, T. 2018. Keppi: A tangible user interface for self-reporting pain. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–13.

Al-Shamayleh, A. S.; Ahmad, R.; Abushariah, M. A.; Alam, K. A.; and Jomhari, N. 2018. A systematic literature review on vision based gesture recognition techniques. *Multimedia Tools and Applications*, 77(21): 28121–28184.

Allan, B. A.; Dexter, C.; Kinsey, R.; and Parker, S. 2018. Meaningful work and mental health: Job satisfaction as a moderator. *Journal of Mental Health*, 27(1): 38–44.

Balayn, A.; Rikalo, N.; Lofi, C.; Yang, J.; and Bozzon, A. 2022. How can Explainability Methods be Used to Support Bug Identification in Computer Vision Models? In *CHI Conference on Human Factors in Computing Systems*, 1–16.

Carfì, A.; and Mastrogiovanni, F. 2021. Gesture-Based Human-Machine Interaction: Taxonomy, Problem Definition, and Analysis. *IEEE Transactions on Cybernetics*.

Cheok, M. J.; Omar, Z.; and Jaward, M. H. 2019. A review of hand gesture and sign language recognition techniques. *International Journal of Machine Learning and Cybernetics*, 10(1): 131–153.

Demartini, G.; Difallah, D. E.; Gadiraju, U.; and Catasta, M. 2017. An introduction to hybrid human-machine information systems. *Foundations and Trends in Web Science*, 7(1): 1–87.

Difallah, D.; Filatova, E.; and Ipeirotis, P. 2018. Demographics and dynamics of mechanical Turk workers. In *Proceedings of the eleventh ACM international conference on web search and data mining*, 135–143.

Difallah, D. E.; Catasta, M.; Demartini, G.; Ipeirotis, P. G.; and Cudré-Mauroux, P. 2015. The dynamics of micro-task crowdsourcing: The case of amazon mturk. In *Proceedings of the 24th international conference on world wide web*, 238–247.

Dontcheva, M.; Morris, R. R.; Brandt, J. R.; and Gerber, E. M. 2014. Combining crowdsourcing and learning to improve engagement and performance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 3379–3388.

Draws, T.; Rieger, A.; Inel, O.; Gadiraju, U.; and Tintarev, N. 2021. A checklist to combat cognitive biases in crowdsourcing. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, volume 9, 48–59.

Dubey, N.; Dubey, G.; Tripathi, H.; and Naqvi, Z. 2019. Ergonomics for desk job workers-an overview. *Int J Health Sci Res*, 9(7): 257–266.

Erdfelder, E.; Faul, F.; and Buchner, A. 1996. GPOWER: A general power analysis program. *Behavior research methods, instruments, & computers*, 28(1): 1–11.

Erlei, A.; Das, R.; Meub, L.; Anand, A.; and Gadiraju, U. 2022. For What It's Worth: Humans Overwrite Their Economic Self-interest to Avoid Bargaining With AI Systems. In *CHI Conference on Human Factors in Computing Systems*, 1–18.

Erlei, A.; Nekdem, F.; Meub, L.; Anand, A.; and Gadiraju, U. 2020. Impact of algorithmic decision making on human behavior: Evidence from ultimatum bargaining. In *Proceedings of the AAAI conference on human computation and crowdsourcing*, volume 8, 43–52.

Fan, S.; Gadiraju, U.; Checco, A.; and Demartini, G. 2020. CrowdCO-OP: Sharing Risks and Rewards in Crowdsourcing. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW2): 1–24.

Gadiraju, U.; Checco, A.; Gupta, N.; and Demartini, G. 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): 1–29.

Gadiraju, U.; and Demartini, G. 2019. Understanding Worker Moods and Reactions to Rejection in Crowdsourcing. In *Proceedings of the 30th ACM Conference on Hypertext and Social Media*, 211–220.

Gadiraju, U.; and Dietze, S. 2017. Improving learning through achievement priming in crowdsourced information finding microtasks. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*, 105–114.

Gadiraju, U.; Kawase, R.; and Dietze, S. 2014. A taxonomy of microtasks on the web. In *Proceedings of the 25th ACM conference on Hypertext and social media*, 218–223.

Gadiraju, U.; Kawase, R.; Dietze, S.; and Demartini, G. 2015. Understanding malicious behavior in crowdsourcing platforms: The case of online surveys. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 1631–1640.

Gadiraju, U.; and Yang, J. 2020. What can crowd computing do for the next generation of AI systems? In *CSW@ NeurIPS*, 7–13.

Gray, M. L.; and Suri, S. 2019. *Ghost work: how to stop Silicon Valley from building a new global underclass*. Eamon Dolan Books.

Gupta, N.; Martin, D.; Hanrahan, B. V.; and O'Neill, J. 2014. Turk-life in India. In *Proceedings of the 18th International Conference on Supporting Group Work*, 1–11.

Hart, S. G. 2006. NASA-task load index (NASA-TLX); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting*, volume 50, 904–908. Sage publications Sage CA: Los Angeles, CA.

Irani, L. C.; and Silberman, M. S. 2013. Turkopticon: Interrupting worker invisibility in amazon mechanical turk. In *Proceedings of the SIGCHI conference on human factors in computing systems*, 611–620.

Jung, J.-Y.; Qiu, S.; Bozzon, A.; and Gadiraju, U. 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.

Karam, M.; and m. c. schraefel. 2005. A Taxonomy of Gestures in Human Computer Interactions. Project report, University of Southampton.

Khanna, S.; Ratan, A.; Davis, J.; and Thies, W. 2010. Evaluating and improving the usability of Mechanical Turk for low-income workers in India. In *Proceedings of the first ACM symposium on computing for development*, 1–10.

Kim, K.; Kim, J.; Choi, J.; Kim, J.; and Lee, S. 2015. Depth camera-based 3D hand gesture controls with immersive tactile feedback for natural mid-air gesture interactions. *Sensors*, 15(1): 1022–1046.

Kittur, A.; Nickerson, J. V.; Bernstein, M.; Gerber, E.; Shaw, A.; Zimmerman, J.; Lease, M.; and Horton, J. 2013. The future of crowd work. In *Proceedings of the 2013 conference on Computer supported cooperative work*, 1301–1318.

Lascau, L.; Gould, S. J.; Brumby, D. P.; and Cox, A. L. 2022. Crowdworkers' Temporal Flexibility is Being Traded for the Convenience of Requesters Through 19 'Invisible Mechanisms' Employed by Crowdworking Platforms: A Comparative Analysis Study of Nine Platforms. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*, 1–8.

Lee, M.-H.; Park, S.-J.; and Kim, J.-S. 2013. Effects of neck exercise on high-school students' neck–shoulder posture. *Journal of physical therapy science*, 25(5): 571–574.

Loewenstein, R. J. 1991. An office mental status examination for complex chronic dissociative symptoms and multiple personality disorder. *Psychiatric Clinics*, 14(3): 567–604.

Luttmann, A.; Schmidt, K.-H.; and Jäger, M. 2010. Working conditions, muscular activity and complaints of office workers. *International Journal of Industrial Ergonomics*, 40(5): 549–559.

Mao, A.; Kamar, E.; and Horvitz, E. 2013. Why stop now? predicting worker engagement in online crowdsourcing. In *First AAAI Conference on Human Computation and Crowdsourcing*.

Martin, D.; Hanrahan, B. V.; O'Neill, J.; and Gupta, N. 2014. Being a turker. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*, 224–235.

Mehrparvar, A. H.; Heydari, M.; Mirmohammadi, S. J.; Mostaghaci, M.; Davari, M. H.; and Taheri, M. 2014. Ergonomic intervention, workplace exercises and musculoskeletal complaints: a comparative study. *Medical journal of the Islamic Republic of Iran*, 28: 69.

Montero-Marin, J.; Garcia-Campayo, J.; Pérez-Yus, M. C.; Zabaleta-del Olmo, E.; and Cuijpers, P. 2019. Meditation techniques v. relaxation therapies when treating anxiety: A meta-analytic review. *Psychological medicine*, 49(13): 2118–2133.

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

Narula, P.; Gutheim, P.; Rolnitzky, D.; Kulkarni, A.; and Hartmann, B. 2011. Mobileworks: A mobile crowdsourcing platform for workers at the bottom of the pyramid. In *Workshops at the twenty-fifth AAAI conference on artificial intelligence*.

Ni, J.; Li, J.; and McAuley, J. 2019. Justifying recommendations using distantly-labeled reviews and fine-grained aspects. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 188–197.

Organization, W. H. 2001. *The World Health Report 2001: Mental health: new understanding, new hope*. World Health Organization.

O'Brien, H. L.; Cairns, P.; and Hall, M. 2018. A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short form. *International Journal of Human-Computer Studies*, 112: 28–39.

Patel, V.; Chesmore, A.; Legner, C. M.; and Pandey, S. 2022. Trends in Workplace Wearable Technologies and Connected-Worker Solutions for Next-Generation Occupational Safety, Health, and Productivity. *Advanced Intelligent Systems*, 4(1): 2100099.

Pejtersen, J. H.; Kristensen, T. S.; Borg, V.; and Bjorner, J. B. 2010. The second version of the Copenhagen Psychosocial Questionnaire. *Scandinavian journal of public health*, 38(3_suppl): 8–24.

Peluso, M. A. M.; and De Andrade, L. H. S. G. 2005. Physical activity and mental health: the association between exercise and mood. *Clinics*, 60(1): 61–70.

Qiu, S.; Bozzon, A.; Birk, M. V.; and Gadiraju, U. 2021. Using worker avatars to improve microtask crowdsourcing. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2): 1–28.

Quek, F. K. 1995. Eyes in the interface. *Image and vision computing*, 13(6): 511–525.

Ranasinghe, P.; Perera, Y. S.; Lamabadusuriya, D. A.; Kulatunga, S.; Jayawardana, N.; Rajapakse, S.; and Katulanda, P. 2011. Work related complaints of neck, shoulder and arm among computer office workers: a cross-sectional evaluation of prevalence and risk factors in a developing country. *Environmental Health*, 10(1): 1–9.

Rojas-Muñoz, E.; and Wachs, J. P. 2019. Magic: A fundamental framework for gesture representation, comparison and assessment. In *2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019)*, 1–8. IEEE.

Rzeszotarski, J. M.; Chi, E.; Paritosh, P.; and Dai, P. 2013. Inserting micro-breaks into crowdsourcing workflows. In *First AAAI conference on human computation and crowdsourcing*.

Sannon, S.; and Cosley, D. 2019. Privacy, Power, and Invisible Labor on Amazon Mechanical Turk. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–12.

Sarma, D.; and Bhuyan, M. K. 2021. Methods, databases and recent advancement of vision-based hand gesture recognition for hci systems: A review. *SN Computer Science*, 2(6): 1–40.

Shin, J.; and Kim, C. M. 2017. Non-touch character input system based on hand tapping gestures using Kinect sensor. *IEEE Access*, 5: 10496–10505.

Smith, C. 2014. *The effects of an ergonomic workstation intervention and exercise on office workers with neck pain*. Master's thesis, The University of Cape Town.

Stawarz, K.; and Cox, A. L. 2015. Designing for health behavior change: HCI research alone is not enough. In *CHI'15* *workshop: Crossing HCI and Health: Advancing Health and Wellness Technology Research in Home and Community Settings*.

Stein, F. 2001. Occupational stress, relaxation therapies, exercise and biofeedback. *Work*, 17(3): 235–245.

Stern, H. I.; Wachs, J. P.; and Edan, Y. 2008. Designing hand gesture vocabularies for natural interaction by combining psycho-physiological and recognition factors. *International Journal of Semantic Computing*, 2(01): 137–160.

Tariq, S. H.; Tumosa, N.; Chibnall, J. T.; Perry III, M. H.; and Morley, J. E. 2006. Comparison of the Saint Louis University mental status examination and the mini-mental state examination for detecting dementia and mild neurocognitive disorder—a pilot study. *The American journal of geriatric psychiatry*, 14(11): 900–910.

Toxtli, C.; Suri, S.; and Savage, S. 2021. Quantifying the Invisible Labor in Crowd Work. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2): 1–26.

Uzor, S.; Jacques, J. T.; Dudley, J. J.; and Kristensson, P. O. 2021. Investigating the accessibility of crowdwork tasks on Mechanical Turk. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–14.

Vaughan, J. W. 2017. Making Better Use of the Crowd: How Crowdsourcing Can Advance Machine Learning Research. *J. Mach. Learn. Res.*, 18(1): 7026–7071.

Vuletic, T.; Duffy, A.; Hay, L.; McTeague, C.; Campbell, G.; and Grealy, M. 2019. Systematic literature review of hand gestures used in human computer interaction interfaces. *International Journal of Human-Computer Studies*, 129: 74–94.

Wachs, J. P.; Kölsch, M.; Stern, H.; and Edan, Y. 2011. Vision-based hand-gesture applications. *Communications of the ACM*, 54(2): 60–71.

Ware, J.; Snow, K. K.; Kosinski, M.; and Gandek, B. 1993. SF-36 health survey. *Manual and interpretation guide. Boston: The Health Institute, New England Medical Center*, 10–6.

Ware Jr, J. E. 2000. SF-36 health survey update. *Spine*, 25(24): 3130–3139.

Woo, E.; White, P.; and Lai, C. 2016. Ergonomics standards and guidelines for computer workstation design and the impact on users' health–a review. *Ergonomics*, 59(3): 464–475.

Yang, J.; Redi, J.; Demartini, G.; and Bozzon, A. 2016. Modeling task complexity in crowdsourcing. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, volume 4, 249–258.

Zhang, Y.; Ding, X.; and Gu, N. 2018. Understanding fatigue and its impact in crowdsourcing. In *2018 IEEE 22nd International Conference on Computer Supported Cooperative Work in Design ((CSCWD))*, 57–62. IEEE.

Zhuang, M.; and Gadiraju, U. 2019. In What Mood Are You Today? An Analysis of Crowd Workers' Mood, Performance and Engagement. In *Proceedings of the 10th ACM Conference on Web Science*, 373–382.