Emotion Embodied: Unveiling the Expressive Potential of Single-Hand Gestures

Yuhan Luo
Department of Computer Science
City University of Hong Kong
Hong Kong SAR
yuhanluo@cityu.edu.hk

Junnan Yu
School of Design
The Hong Kong Polytechnic University
Hong Kong SAR
junnan.yu@polyu.edu.hk

Minhui Liang
Department of Computer Science
City University of Hong Kong
Hong Kong SAR
mhliang4-c@my.cityu.edu.hk

Yichen Wan
School of Design
The Hong Kong Polytechnic University
Hong Kong SAR
yichen.wan@connect.polyu.hk
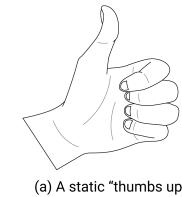
Kening Zhu
School of Creative Media
City University of Hong Kong
Hong Kong SAR
kening.zhu@cityu.edu.hk

Shannon Sie Santosa
Department of Computer Science
City University of Hong Kong
Hong Kong SAR
sssantosa2-c@my.cityu.edu.hk
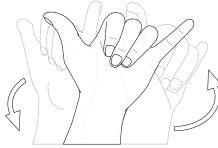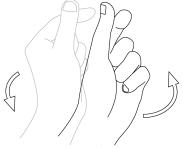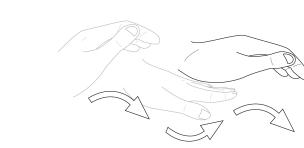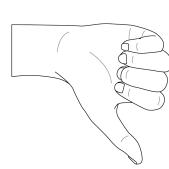
Figure 1: Examples of gestures collected from our survey study illustrate the expressive potential of a single-hand gesture: one emotion can be expressed differently with different finger-pointing directions, palm directions, movements, and strength.

ABSTRACT

Hand gestures are widely used in daily life for expressing emotions, yet gesture input is not part of existing emotion tracking systems. To seek a practical and effortless way of using gestures to inform emotions, we explore the relationships between gestural features and commonly experienced emotions by focusing on single-hand gestures that are easy to perform and capture. First, we collected 756 gestures (in photo and video pairs) from 63 participants who expressed different emotions in a survey, and then interviewed 11 of them to understand their gesture-forming rationales. We found that the valence and arousal level of the expressed emotions significantly correlated with participants' finger-pointing direction and their gesture strength, and synthesized four channels through which participants externalized their expressions with gestures. Reflecting on the findings, we discuss how emotions can be characterized and contextualized with gestural cues and implications for designing multimodal emotion tracking systems and beyond.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in HCI**; **Scenario-based design**.

KEYWORDS

Emotion capture, gesture interaction, gesture elicitation

1 INTRODUCTION

Emotion capture and tracking can benefit various domains, including health assessment [20, 58] and computer-mediated communication [17, 36]. Traditionally, emotion capture relies on self-reporting through diaries or questionnaires [18, 46, 85], which often impose a heavy input burden on people and can be susceptible to recall bias. Automated emotion tracking technologies such as facial expressions [7], speech analysis [49], and physiological signals detection (e.g., heart rate variability [21, 70], skin responses [42]) offer objective data but can be challenging to deploy in real-world settings, due to the wearing burden [86] and data interpretation difficulties [21].

Hand gestures, as part of our body language, are often used to express emotions in daily life [33, 56]. This connection between emotion and gesture is grounded in *Embodied Cognition*, a cognitive science theory that emphasizes the important roles of perception, action, and the environment in shaping our cognitive process [6, 79]. Rather than viewing cognition as amodal, abstract, and detached from the physical world, *Embodied Cognition* advocates that the mind should be understood in relation to the body's interaction with the physical and social surroundings [6, 74, 90]. Following *Embodied Cognition,* human gestures represent a form of physical embodiment of our mental activities. Consequently, they can convey our emotions and, conversely, serve as indicators of our emotional state. For example, people naturally move their hands while speaking to convey a feeling of excitement, hesitation, or confidence [33, 56]. Some gestures such as "thumbs up," "ok," and "victory" are widely recognized that people can easily interpret their intended messages without explicit explanation (e.g., a thumbs-up gesture is often associated with positive connotations, such as approval and agreement) [23, 52]. Other gestures, although without universal recognition, can still carry meanings within certain contexts (e.g., a clenched fist usually conveys a feeling of anger) [56].

In the HCI community, gesture is often seen as an interaction modality to facilitate the control of different objects, applications, and devices [87, 92], which has driven the development of technologies for precise gesture recognition and analysis [48, 89]. However, it remains unclear how these gesture recognition technologies can be applied to capture the emotions that people experience in daily life. Although prior work has investigated the connections between hand gestures and emotions [5, 35, 36], they largely focused on the symbolic meanings of the gestures (e.g., mapping gestures to different emojis [36]), and overlooked how various emotional states may affect the nuances in gestural features such as finger-pointing direction, palm direction, strength, and moving frequency [56]. Moreover, existing studies often examined emotion expression with gestures involving head, shoulder, and other body parts [15, 36, 68]. While conveying rich meanings, these body gestures can be difficult to perform and recognize in everyday life situations.

To seek a more practical and effortless way of using gestures to inform emotions, we focus on single-hand gestures that are often simple to perform and can be easily captured by lightweight and off-the-shelf technologies such as smart watches [48, 89] and sensing gloves [13]. As the first step, we are interested in understanding **how people express different types of emotions through single-hand gestures** by investigating the connections between gestural features and emotion dimensions (i.e., valence and arousal) [75], as well as the process by which individuals externalize their emotions using hand gestures.

We first conducted a survey study to collect photos and videos of individuals' single-hand gestures as they expressed different emotions. These emotions (e.g., happy, relaxed, angry, tired) are derived from the well-known Russell Emotion Circumplex covering a comprehensive range of valence (i.e., positiveness or negativeness) and arousal (i.e., intensity) levels [75]. To effectively elicit each emotion during the survey, the research team carefully selected 12 image stimuli from the OASIS database, an open-access online image stimulus depicting a broad spectrum of themes with normative ratings on valence and arousal [41]. Our survey gathered 756 single-hand gestures (captured by photos and video) from 63 individuals representing diverse ages, regions, and ethnic groups. To gain deeper insights into the reasoning behind how participants form the gestures to express their emotions, we conducted follow-up interviews with 11 of them.

We found the valence and arousal levels of the expressed emotions significantly correlated with participants' finger-pointing direction and gesture strength; and that younger individuals tended to employ more diverse gestures to express their emotions. The interviews revealed that our participants not only leveraged existing symbolic meanings associated with gestures, but also "self-created" a variety of gestures for more personalized expression. They added meanings and stories behind the gestures, and moved their hand in different angles and speeds to enrich the meanings. We further synthesized four channels through which participants externalized their emotions, including *communication norms*, *creative embodiment*, *physical expression*, and *abstract expression*. In particular, individuals' cultural practices and the visual elements in the image stimuli played parts in their gesture-forming process. These findings unfolded the dynamic and subjective nature of emotion expression, which can potentially unlock a new avenue for emotion capture and tracking with single-hand gestures.

This work contributes to the HCI community in three aspects: (1) Empirical evidence of how different gestural features are correlated with the valence and arousal associated with emotions; (2) In-depth understanding of how individuals express their emotions using single-hand gestures and their underlying mental models. To the best of the authors' knowledge, this is the first empirical study probing into the connections between emotions and single-hand gestures; and (3) design implications for efficient and inclusive emotion tracking systems leveraging gesture input.

2 RELATED WORK

This section covers related work on the role of gestures in emotion expression, existing emotion capture methods, and prior HCI research on gesture elicitation to motivate our study method.

2.1 Emotion Expression and Gestures

Emotion is generally recognized as mental states related to perceptions, thoughts, and behaviors [9]. The widely known Russell's

Circumplex Model of Affect posits a bipolar circular space in which human emotions are distributed, with valence (i.e., the positiveness or negativeness of that feeling) and the level of arousal (i.e., the intensity or activation level of feeling excited or alert) that jointly represent various affective states (e.g., being alarmed, sleepy, pleased, or miserable) [75]. In a recent review, Keltner et al. summarized that humans possess at least 20 distinct emotions conveyed through a wide range of indices [32], which encompasses not only facial expressions [7] but also vocalizations [93], textual communications [1, 38, 62], physiological signals [70], body movements [2], and gestures [19]. In other words, human emotions are a nuanced phenomenon that entails integration and synchronization of cognition, physiology, and behavior.

In expressing emotions, the role of physical movement can be understood through a series of somatic theories regarding the origin of emotions [12, 29, 43, 77]. The classical James-Lange Theory [29, 43] postulates that emotions are elicited by physiological reactions to certain events, with bodily actions rather than cognitive appraisal forming the basis of emotional experiences. As a particular type of bodily action, hand gestures have been recognized as a significant channel of self-expression [5], which is present in almost all cultures as a non-verbal language to facilitate communications [10, 28, 57, 65]. For example, politicians use the movements of their left and right hands to represent the left-wing and right-wing parties in political conversations [10].

Referring to McNeill's pioneering work [56], hand gestures can be broadly classified into two sorts: *representational gestures* with semantic-driven meanings (e.g., shapes, actions, and events), and *non-representational gestures* with no concrete meanings and are characterized by rapid, rhythmic hand movements. In sign languages, hand gestures have also been seen as an important tool to encode emotional signals [27, 71]. For instance, researchers found that people without prior knowledge of Finnish sign language were able to capture anger and neutrality via hand movements during signing, indicating that hand gestures have a rich capacity to communicate emotional signals. Accordingly, there are coding systems developed to code and analyze hand gestures depending on the particular research inquiries. As summed up by Kipp and Martin [35], the coding schemes can be categorized into a descriptive scheme that relies on an objective description of hand movements and an interpretative scheme focusing on examining the semantic meanings behind bodily movements [69, 88]. These coding strategies have driven the development of gesture databases for researchers to detect, classify, and understand hand gestures [47, 52, 53, 59, 84]. For example, Ma et al. developed a gesture database with 10,300 original photos from 129 participants [52], which contains seven well-known hand gestures used as communicative symbols associated with positive (e.g., "thumbs up": indicating recognition; "OK": showing agreement; "victory": demonstrating success), negative (e.g., "thumbs down": suggesting disapproval), and neutral emotions (e.g., "waving hand": attendance, farewell, or ignorance to criticism; "phone call": a later call back). While lacking diversity in participants' backgrounds [52]), this database has laid the groundwork for the later development of hand gesture recognition systems, serving as the pillar for emotion detection and beyond.

2.2 Emotion Capture and Tracking

In recent decades, there has been a growing recognition of the importance of tracking emotions, especially in the healthcare context [20]. Specifically, emotional states, such as stress and anxiety, can predict various health conditions, including mental issues [58] and physical illnesses [4], which highlights the need to effectively monitor emotions, or *emotion tracking*, as a first step in treatment intervention. Emotion tracking is an important tool for understanding patterns of emotion variability, promoting self-awareness, and facilitating emotion regulation and overall well-being [44]. Given the subjectivity of emotions, researchers commonly use an experience-sampling methodology that relies on self-report [85]. These tools can range from traditional methods such as online diaries [73], open-ended surveys [60], and Likert-scale questionnaires [46], to tangible tools like 3D clay pieces shaped to represent emotional experiences [44]. However, self-report assessments of emotions involve significant limitations, covering but not limited to report bias, where individuals may not possess a comprehensive and accurate understanding of their instant emotional experiences or report emotions in conformity to social or cultural expectations [63]. Additionally, people may lack motivation for long-term and intensive data collection in self-reported emotion-tracking tasks [8].

To lower the burden of manual emotion tracking, recent advancements in wearable sensing have expanded the possibilities, particularly leveraging physical embodiment of emotional reactions. This class of emotion-capturing techniques works mainly through the collection of biological signals to mark the changes in the autonomic nervous system responsible for emotion responses [39], which typically includes the data of activities in muscle (e.g., electromyography signals [40]), skin (e.g., electrodermal activity [82], temperature [66], blood flow [45]), brain (e.g., electroencephalography signals [64]), and respiration indices [94]. However, collecting such indices, particularly biological signals, is usually limited to laboratory settings due to factors like high cost, complex setup, and lack of applicability in daily life [34]. Therefore, researchers and practitioners have been seeking more affordable tracking methods, resulting in the development of smart gadgets that use wearable devices like smartwatches, wristbands, and headbands to recognize and capture emotion-specific cues.

2.3 Gesture Elicitation Studies in HCI

Gesture elicitation, introduced by Wobbrock et al. in 2005 as a "guessability method" to collect users' intuitive preferences for symbolic input [91], has been widely used in HCI research for understanding and leveraging gestures as an interaction means between humans and computing systems [87]. As Villarreal-Narvaez et al. summarized from 216 gesture elicitation studies spanning 2009 to 2019 [87], existing research predominantly focused on utilizing gestures to improve task performance and interaction efficiency. For example, Wobbrock et al. collected a set of user-defined gestures as commands to control and manipulate objects on tablet surfaces [92]; Chan et al. studied how single-hand microgestures can help users better navigate small interaction spaces (e.g., mobile phones) [16], and Sharma et al. investigated how such mircogestures can be used to facilitate interaction with hand-hold objects [80]. Interestingly, despite the proliferation in the studies of physical embodiment of

emotions, gesture elicitation, and multimodal emotion tracking in the HCI community, there has been limited design and development work using hand gestures to infer people's emotional states. Most relatedly, Koh et al. leveraged the symbolic meaning of hand gestures to analogize emojis in the context of instant messaging [36]. The researchers conducted a series of studies, in which participants used two hands along with body movements to represent over 30 popular emojis through self-creation or learning from others. Their results showed that some emojis (e.g., "thumbs up," "victory," "pray") are intuitive to represent whereas others (e.g., "sob," "grin," "expressionless") posed challenges for participants to convey. Building on the findings, the research group later applied gesture interaction in virtual meetings, where people could use gestural cues for impromptu polling [37]. Other works focused on understanding how people communicate emotions through touch-based gestures such as squeezing, lifting, pushing, and tapping [25, 26], which is often applied in human-robot interaction [95, 96]. To the best of our knowledge, no research work has been done to examine whether and how people can use single-hand gestures to intuitively express emotions at different valence and arousal levels. This matter, as argued by Asalıoğlu and Göksun [5], possibly stems from difficulties in precisely transcribing and classifying the complicated attributes of hand gestures. To bridge this gap, we aim to understand the mapping between emotions and gestures, laying a foundation for future designs of gesture-facilitated emotion tracking.

3 METHOD

This section describes the procedures of our survey study and the follow-up interviews. We explain our rationale for the study design while describing the types of data we gathered. Next, we cover the data analysis methods, including gesture coding, statistical tests, and qualitative interview analysis.

3.1 Survey Data Collection

As an exploratory study, we chose to conduct a survey study because it allows us to gather a large amount of data from a diverse range of participants. Drawing from Russell Emotion Circumplex [75], we derived 12 emotions covering a comprehensive range of valence and arousal levels:

- Positive valence and high-arousal: *excited, delighted, happy*.
- Positive valence and low-arousal: *content, relaxed, calm*.
- Negative valence and high-arousal: *tense, angry, frustrated*.
- Negative valence and low-arousal: *depressed, bored, tired*.

3.1.1 Image Stimuli Selection. Prior to eliciting gestures for different emotion expressions, our first step is to elicit these 12 emotions. We used a set of images as emotional stimuli, which is a commonly used approach in social psychology studies [24]. To determine the appropriate stimuli for the 12 emotions, we searched through 900 images from the OASIS database, an open-access image source that was published in 2017 and widely used in numerous emotion elicitation studies [41]. The 900 images were randomly divided into six parts (each containing 150 images); each part was reviewed by one of the six researchers (three male and three female) independently to select the most representative image for each emotion, with notes explaining their choices. Next, the entire research group reviewed all the selected images and voted for the best one for each

emotion. Ultimately, we selected 12 image stimuli to be presented in the survey (see Figure 2). Note that we also compared the valence and arousal rating of each image from the data provided by OASIS, which consistently aligned with the emotion words that they were chosen to represent.

3.1.2 Survey Design. The survey started with a consent form describing the survey procedure and potential risks (e.g., emotional fluctuation during the survey). We also appended a video illustrating the process of capturing single-hand gestures from different angles and our requirements for the photos and videos regarding clarity and resolution. The instructions explicitly stated that all the photos and videos collected in the survey would be used exclusively for research purposes and would not be linked to participants' personal identities. Participants were strongly encouraged not to include any personally identifiable information, such as their faces, in the captured media. For each emotion, participants were first asked to select an option from the list of 12 words that best describes the emotion they felt, and to rate the valence and arousal of that emotion on a scale from 1 to 7. To ensure that participants choose the word solely based on their perceived emotions, we did not require them to choose different words for different image stimuli. The rating questions were designed to assess participants' understanding of the emotions portrayed by the image stimuli and to identify potential careless respondents for later exclusion. Next, participants were requested to upload a photo and a short video (2 to 5 seconds) to capture the static gesture and their gesture-forming process. The 12 image stimuli were presented in a random order to counterbalance any bias or influence that may arise from the specific sequence in which the images are shown.

For quality control, a test question was inserted randomly in the survey, asking participants for their year of birth. Inconsistent responses between the test question and the provided age would result in exclusion from the analysis. At the end of the survey, participants were asked basic demographic questions, including gender, age, occupation, and regions they live. On average, the survey took each participant 30 minutes to complete. To thank participants for their time, each of those who completed the survey received a USD 5 Amazon Gift Card.

3.1.3 Participants. We disseminated the survey through flyers on campus and online posts on Reddit and Facebook. Among the 457 participants who filled out the survey, 86 completed all the questions. After a thorough review of the uploaded photos and videos, we excluded the data from 23 participants due to irrelevant uploads (e.g., photos or videos not capturing any hand gestures), unclear visuals (e.g., difficulty in discerning gesture shape and position), repetitive uploads (e.g., we did not enforce participants to capture different gestures for each emotion, but we excluded those who uploaded multiple instances of the exact same content), contrast responses in valence/arousal choice compared to the associated emotion word (e.g., choosing "angry" with a high-valence rating, or choosing "happy" for a clearly negative image stimulus), or failure to pass the test question. As a result, we included data from 63 participants (39 females, 23 males, and one non-binary), whose ages ranged from 18 to 54 (M = 26, SD = 7.67). Our participants were from diverse regions including US (38), UK (11), Hong Kong (5), Kenya (3), Indian (3), Nigeria (1), Lithuania (1), and Mexico (1).

Figure 2: The image stimuli selected from the OASIS database to elicit 12 emotions (positive high arousal: (1)–(3); positive low arousal: (4)–(6); negative high arousal: (7)–(9); negative low arousal: (10)–(12)). The word under each image was used as a reference for the research team, rather than being provided by the participants. In the survey, participants could choose different words to represent their emotions elicited by the image stimuli.

3.2 Follow-Up Interviews

We sent an email to invite all 63 participants whose data were included in the survey analysis. Among the 19 participants who responded to our invitation, only 11 ultimately showed up during the interview. These participants were three females and eight males; their ages ranged from 21 to 30 (M = 28, SD = 2.44), and are from the US (8), Kenya (2), and Hong Kong (1). Each participant received a USD 5 Amazon Gift Card as a thanks for their time.

The interview took place within two weeks of the survey completion via Zoom and lasted from 20 to 40 minutes. To help participants better recall their survey experience, we conducted the interviews in a semi-structured manner. We shared our screen with a series of slides consisting of the 12 image stimuli that participants were presented with during the survey, along with the photos and videos of gestures they formed. For each stimulus, participants were given time to recollect their emotional responses and then explain their use of certain gestures to express that particular emotion. In this process, we prompted participants to elaborate on their understandings or interpretations of the emotions as well as nuances associated with their gestures, such as the features we coded (e.g., gesture strength, motion frequency) and reasons for using similar gestures to express different emotions (or different gestures to express similar emotions).

3.3 Data Analysis

3.3.1 Gesture Coding. Before mapping the gestures to different emotions, we first aimed to characterize the prominent gestural features from the collected photos and videos. This approach to manually characterize the gestures has also been widely used in prior work to ensure that subtleties and nuances in gestures are not overlooked [35, 69, 88]. Our gesture coding procedure involved the following steps:

- **Step 1. Initial coding**: To start, two experienced HCI researchers with prior experience in gesture interaction reviewed a sample of 176 (23%) gestures with a balanced distribution of the 12 emotions. In this step, we did not use any predefined codebook; instead, we took a bottom-up approach to observe the recurring patterns of the gestures and noted down all the features deemed important. Following our own feature list, we independently coded all the sample gestures.

- **Step 2. Codebook development**: The two experienced researchers compared the gestural features they generated. Through rounds of discussions, we merged similar features (e.g., "motion speed" and "motion frequency") and gathered their possible values—codes (e.g., "motion frequency" can be "high," "middle," or "low"). We also removed features that are deemed less relevant (e.g., "arm direction") or difficult to characterize (e.g., "motion fluidity," "motion range"). With the codebook, we re-coded all the sample gestures (the details of the codebook are described below).

- **Step 3. Coding training**: The two experienced researchers shared the codebook with the other two graduate research assistants (RAs). We explained the meaning of each feature with examples and then asked the two RAs to code the sample gestures independently following the codebook. Next, the RAs compared their coded gestural features with those coded by the two experienced researchers. During this process, we seek to establish a shared understanding of the gestural features among the research team.

- **Step 4. Reliability calculation**: To ensure that the two RAs can carry out the coding tasks in a way that consistently aligns with the codebook, they independently coded another set of gestures (163, 21.5%) and compared each other's codes. To calculate the inter-coder reliability, we used Perreault & Leigh's approach (I_r), which was developed for assessing the quality of nominal data based on qualitative judgments of multiple coders [67]. The result showed that the I_r values were above 0.80 for all seven gestural features (mean: 0.89, maximal: 0.98, minimal: 0.84), which passed the 0.70 thresholds suggested as high reliability [76]. While resolving the discrepancies in this step, the entire research team collectively reviewed all the codes, and then iteratively expanded or updated the possible values of each feature.

Table 1: The distribution of the words chosen by participants when presented with the 12 image stimuli (the percentages listed in this table have been adjusted to rounded values).

Positive valence & high arousal emotion (n = 243, 32.1%)			Positive valence & low arousal emotion (n = 141, 18.7%)			Negative valence & high arousal emotion (n = 211, 27.9%)			Negative valence & low arousal emotion (n = 161, 21.3%)		
excited	delighted	happy	content	relaxed	calm	tense	angry	frustrated	depressed	bored	tired
96 (12.7%)	53 (7.0%)	94 (12.4%)	32 (4.2%)	69 (9.1%)	40 (5.3%)	62 (8.2%)	67 (8.9%)	82 (10.8%)	82 (10.8%)	47 (6.2%)	32 (4.2%)

- **Step 5. Wrapping-up**: The two RAs evenly divided up the work to complete coding the remaining 417 (55.16%) gestures. In this step, the research team met on a weekly basis to discuss the codes, and resolve any discrepancies that arose.

Upon finalizing the codebook, we devised two parts of gestural features—static features based on the photos and motion features based on the videos. The static features included the following.

- **Gesture name**: 20 labels that primarily depict the general shape of the hand. These labels are informed by commonly referred symbols (e.g., "thumbs up," "thumbs down," "victory," "finger heart"), descriptive terms (e.g., "open palm with fingers spread," "open palm with fingers pressed together"), or numeric values that do not typically convey other meanings (we have a few gestures labeled as "number three" and "number four", which are different from gestures with commonly referred meanings such as "ok" and "victory"). In cases where a gesture could not be clearly depicted, we labeled them as "others" but still coded their features described below.
- **Finger-pointing direction**: "up," "down," "side" (left or right), "towards the body," "outwards the body," or "none" (e.g., closed fist).
- **Palm direction**: "up," "down," "side" (left or right), "towards the body," or "outwards the body."
- **Gesture strength**: "tight," "loose," or "unclear." We coded this feature by replicating the gestures in the photo with our own hands to evaluate the amount of strength required. After multiple rounds of trials and team discussions, we reached an agreement that the strength of some gestures is clearly tight or loose. The remaining gestures are labeled with "unclear," because they can be replicated with either tight or loose strength, and it was difficult to determine their strength level.
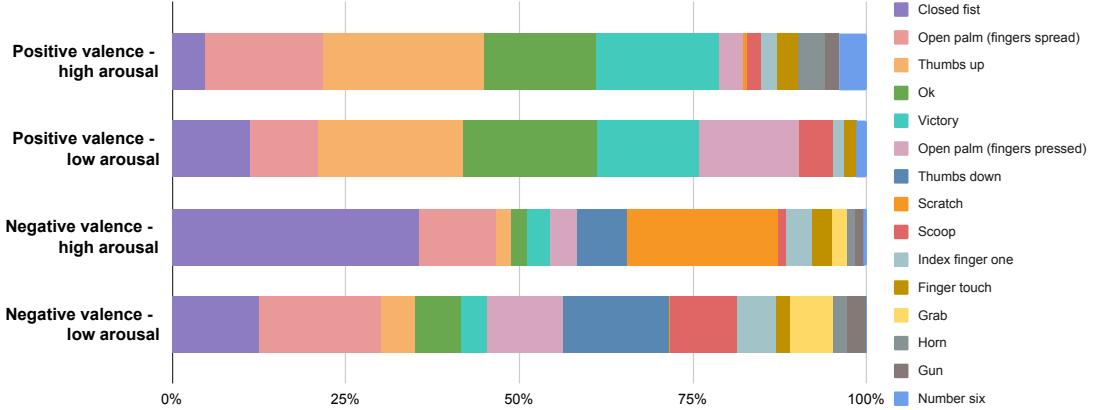
The motion features include the following.

- **Motion name**: 22 labels that depict the major movement of the hand, including 10 types of one-time motion (e.g., "finger flexion," "finger extension," "palm supination," "palm pronation") and 12 types of repeated motion (e.g., "repeated punching," "repeated shaking," "repeated finger flexion and extension"). Similar to the gesture name, if a gesture motion could not be clearly depicted (e.g., involving multiple motions that are difficult to discern one which was the primary), we labeled them as "others" but still coded their motion features; and for a gesture involving multiple motions without a clear message of which motion is the primary, we labeled with as "multiple others." Some gestures are static without any hand movement, which we labeled as "none."

- **Motion frequency**: "low," "middle," "high," or "none." Gestures with motion frequency "none" did not involve any hand movement in the videos.
- **Ending status**: "static" or "moving." At the end of the gesture-forming process, we examined whether the participant was moving their hands. Note that this feature focuses on how a gesture is ended rather than how it is formed. Thus, the motion names of those with "static" ending status may have a one-time motion name or be labeled as "none."

3.3.2 Statistical Tests. We used the words chosen by participants to represent their emotions in response to each stimulus. Rather than relying on the emotions that the original stimuli were intended to evoke, we considered the participants' own choice of their emotions as the basis for our analysis. In accordance with the Russell Emotion Circumplex model [75], we categorized the 12 elicited emotions based on the level of valence (positive or negative) and arousal (high or low). To examine the correlations between gesture features and the two dimensions of emotions (valence and arousal), we performed a Chi-square test [81], by treating emotion valence and arousal as independent variables and each gesture feature as a dependent variable. We then conducted a residual analysis for each significant test. Noting that participants sometimes formed similar gestures to express different emotions, we are interested in investigating if any individual traits contribute to the diversity of gestures they used. Thus, we conducted correlation analysis [55] to examine the relationships between the number of gestures that participants uploaded and their demographics such as age and gender. For participants whose gestures are labeled with multiple *others*, we specifically examined whether these gestures are in a similar shape; if not, they were further categorized with distinct labels (e.g., *other 1*, *other 2*) to facilitate the correlation analysis.

3.3.3 Interview Data Analysis. All the interviews are audio recorded and transcribed into text. To derive insights into participants' gesture-forming rationales, we analyzed the interview transcripts alongside the captured photos and videos of the gestures. Four researchers (two senior researchers, one graduate RA, and one undergraduate RA) first independently read and familiarized themselves with three out of the 11 transcripts and noted down the recurring patterns. Through multiple rounds of discussions, we identified the major themes from the interviews that are centered around the connections between image stimuli, emotion understanding, and gesture expression. Then, the two RAs collaboratively completed the analysis of the remaining transcripts. Next, the entire research team worked together to merge similar codes, address discrepancies, and refine the themes identified. Note that interrater reliability check was not performed for the interview analysis as we aimed to understand participants' rationale for their gesture formation in this

Figure 3: An overview of the gesture names commonly appeared (freq > 10) in our survey and their distribution across emotions in different valence and arousal levels.

particular study rather than to generalize the rationale patterns [54]. Instead, the analysis reliability was ensured through independent coding and cross-checking among the four coders. To provide a visual representation of the insights gained from the interviews, the graduate RA sketched all the gestures formed by the 11 participants who took part in the interviews.

4 RESULTS

In this section, we first report the results from the survey study focusing on the quantitative analysis of gestures. Then we detail our findings from the interviews focusing on participants' gesture-forming rationales and experiences.

4.1 Emotion and Gesture Distribution

4.1.1 Emotion Distribution. Our original image stimuli were selected to evoke 12 different emotions (as described in Section 3.1.1). However, participants sometimes chose the same word in response to different stimuli. As a result, each of them chose five to 12 distinct emotion words (M = 8.75, SD = 1.54). We found that participants tended to choose words representing high-arousal emotions such as *excited*, *happy*, *angry*, and *frustrated*, compared with those representing low-arousal emotions such as *content*, *calm*, *tired* (see Table 1). Thus, in our subsequent analysis to quantify gestural features across emotions, we focused on the two dimensions of emotions (valence and arousal), as their sample distributions are more balanced than those within individual emotion words.
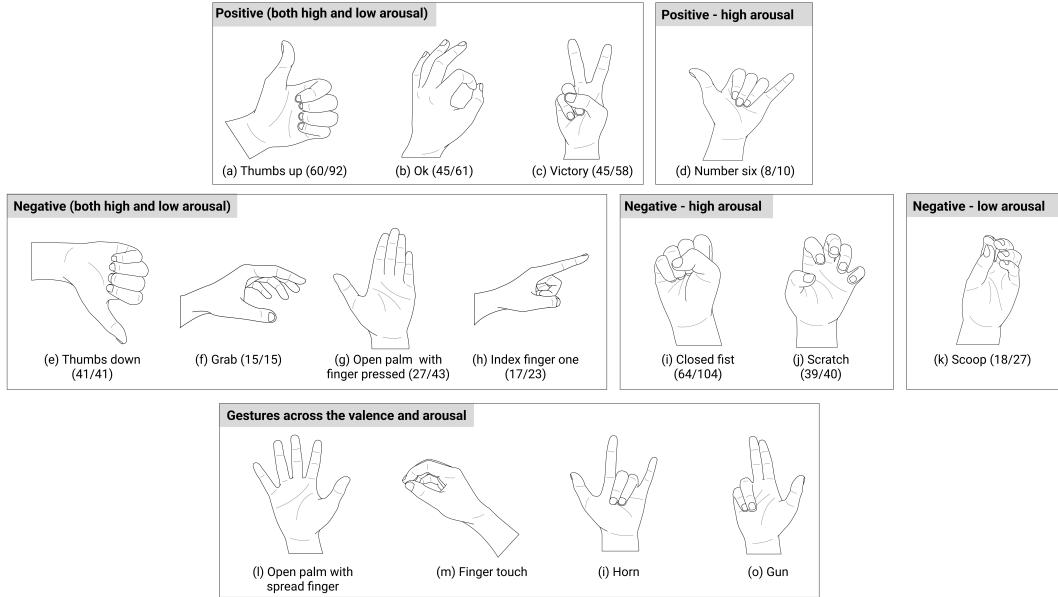
4.1.2 Gesture Overview. In response to each of the 12 image stimuli, each participant captured gestures in a photo and a short video; therefore, we collected 756 single-hand gestures from the 63 participants. Our participants used 4 to 12 unique gestures to express the emotions they experienced in the survey (M = 7.71, SD = 1.56). Interestingly, we found a moderate negative correlation between participants' age and the number of unique gestures they formed (*cor*: -.25, p < .05): the younger participants tended to use more diverse gestures. Most participants used the same hand (often the left one) to form all the gestures, while using the other hand to

take photos and record videos; only two participants alternated two hands throughout the survey. Five participants used a front-facing camera to capture their gestures, and others all used a back-facing camera.

4.1.3 Gesture Names Across Emotions. Among the collected gestures, 667 (88.23%) were assigned with one of the 20 gesture names, while the remaining gestures (89, 11.77%) were labeled with *others*. The most commonly appeared gesture names were illustrated in Figure 4, which include "closed fist" (104, 13.76%), "open palm with spread fingers" (92, 12.17%), "thumbs up" (73, 9.66%), "ok" (60, 7.94%), "victory" (58, 7.67%), "open palm with fingers pressed together" (43, 5.69%), "thumbs down" (41, 5.42%), "scratch" (40, 5.29%), "scoop (27, 3.57%), "index finger one" (23, 3.04%), "finger touch" (16, 2.12%), "grab" (15, 1.98%), "horn" (11, 1.85%), "gun" (11, 1.46%), and "number six" (11, 1.46%).

As shown in Figure 3, "closed fist," "scratch," and "thumbs down" are often used for expressing negative emotions, while "closed fist" appeared more frequently in high-arousal emotions (*tense*, *angry*, *frustrated*), "thumbs down" appeared more frequent in low-arousal emotions (*depressed*, *bored*, *tired*), and "scratch" only appeared in high-arousal emotions. On the other hand, "thumbs up," "ok," and "victory" are commonly used for expressing positive emotions no matter whether the arousal level is low or high. An open palm, either fingers pressed or spread, is seen in all types of emotions. In addition, some gestures, although not widely used by participants, seemed to be used more often in certain emotions. For instance, most "number six" gestures are used for expressing positive emotions with high arousal, most "grab" gestures are used for expressing negative emotions with low arousal, and most "scoop" gestures are used for expressing low arousal emotions.

4.1.4 Gesture Motions. Of the 756 gestures, 131 (17.32 %) were labeled with "none" as they did not involve any motions, 68 (8.99%) were labeled with "others," and 114 (15.08%) were labeled with "multiple others" which involved more than one motions. Among the remaining 443 gestures, 195 (44.02%) involved repeated motions,

Figure 4: Illustrations of the top 15 frequently observed gesture names gathered in the survey (static versions). The gestures are grouped based on whether they tend to appear within certain emotional dimensions (valence and arousal) or across different dimensions. The number under each gesture = the instances of that gesture appearing under the emotional dimension / the total observed instances of that gesture. Note that while we sketched the gestures in a standardized form, the original photos uploaded by participants may differ. For example, gestures with the same name can vary in other features such as finger-pointing direction and palm direction (See Appendix B for original photo examples).

meaning that their ending status as "moving"; 248 (55.98%) were one-time motion without repetition, meaning that their ending status is "static" (we still coded the motion frequency of this group depending on how fast the hand shape was formed). The most commonly observed repeated motions included "finger extension and flexion" (53, 11.96%), "arm extension and flexion" (41, 9.26%), "shaking" (34, 7.67%), "palm flipping" (33, 7.45%), "radial and ulnar deviation" (19, 4.29%), and "knocking" (18, 4.06%). The most commonly observed one-time motions included "finger flexion" (124, 28.00%), "finger extension" (37, 8.35%), "wrist flexion" (13, 2.93%), "palm supination" (11, 2.48%), "palm pronation" (10, 2.26%), and "squeeze" (9, 2.03%). Note that a continuous repeated motion can consist of two types of one-time motions (e.g., "finger extension and flexion" consists of repeated "finger flexion" and "finger extension;" "palm flipping" consists of repeated "palm supination" and "palm pronation"). We illustrated these motions in Figure 5.

4.1.5 Gestural Features × Emotion Dimensions. The Chi-square test showed that the finger-pointing direction (χ^2 = 285.51, p < .001) is significantly correlated with the valence and arousal levels of the emotions. Through residual analysis [1], we found that when expressing negative emotions with low arousal, participants tended to point *downward* with their fingers (*residual* = 6.33); and when expressing positive emotions with high arousal, participants tended to point *upward* (*residual* = 5.87). Note that apart from "thumbs

up" and "thumbs down" which have a clear indication of the finger-pointing direction, other gestures with the same name may have different finger-pointing directions (e.g., we observed *downward* pointing gestures a shape of "ok," "victory," and "number six").

Gesture strength (χ^2 = 107.44, p < .001) is also significantly correlated with the valence and arousal levels of the emotions. The residual analysis showed that when expressing negative emotions with high arousal, the gesture strength is more likely to be tight (*residual* = 6.23), and when expressing negative emotions with low arousal, the gesture strength is more likely to be loose (*residual* = 4.54). However, no significant trend of gesture strength was observed within the positive emotions.
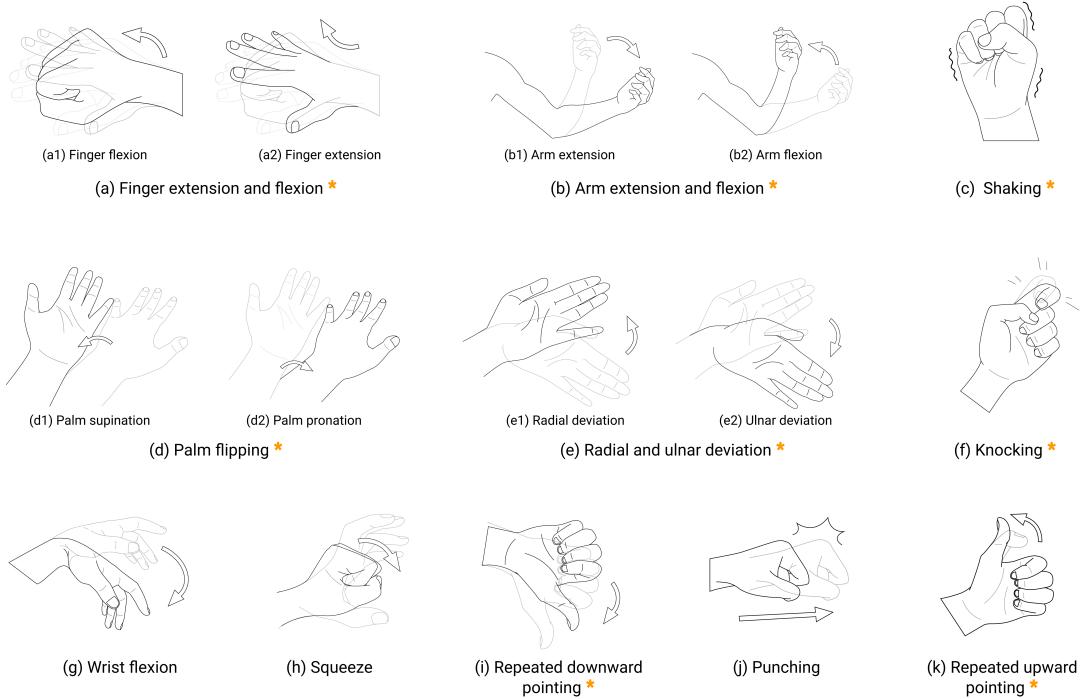
The Chi-square test did not yield significant results for other gestural features (e.g., palm direction, ending status, motion frequency), suggesting that in our study, these features are not correlated with the valence or arousal levels of the expressed emotions.

4.2 From Emotion Activation to Emotion Externalization With Gestures

Here, we describe how participants understood or felt the emotion from the provided stimuli, and then surface how they externalized the experienced emotions with single-hand gestures. These connections are illustrated in Figure 6. The 11 participants who took part in the interviews were denoted by P-#.

4.2.1 Emotion Activation With Image Stimuli. When recalling their survey experience, participants shared four ways that the stimuli activated their emotions. First, participants often relied on the *interpretation of how the people in the image felt* (P2, P3, P8, P6,

---

[1]Pearson residual, which is typically considered as a large deviation from the expectation if it exceeds ±2, suggesting a significant correlation between the independent and dependent variable [81].

Figure 5: Illustrations of most frequently observed motion names from the survey (a)–(h) and additional gestures brought up during the interviews (i)–(k). Motions marked with * are repeated motions with a "moving" ending status. Here we focused on the motion of the hand rather than its shape, thus gestures with the same motion name could be different gestures (e.g., both a "closed fist" and an "open palm" could perform (g) "wrist flexion"). From our observation, there are no significant relationships among the motion names with emotional dimensions, except that (a1) finger flexion (83/124) and (c) shaking (30/34) often appeared in high arousal emotions.
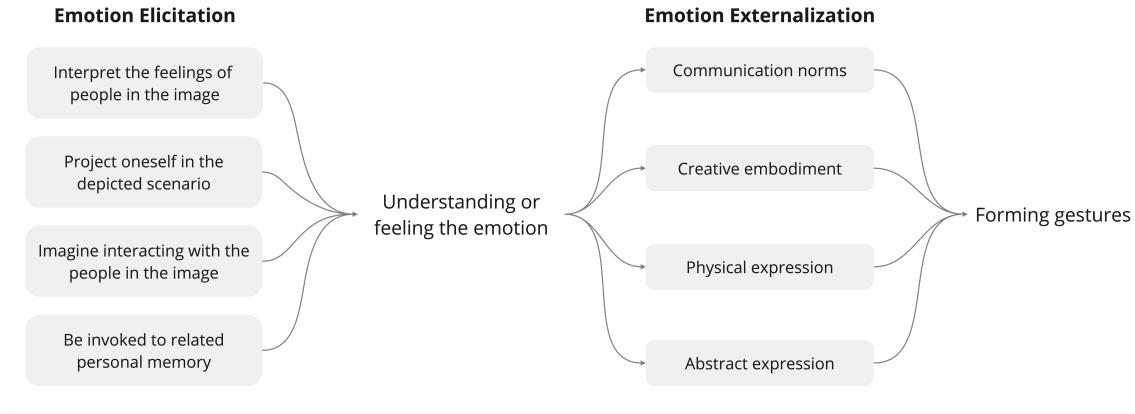
P10, P11). They mentioned analyzing the facial expressions and body language of the people in the images: "*she (the woman in Figure 2 (5)) is getting a massage. She's sleeping and smiling, and I was like, oh she's in a very relaxed state*" (P2). Second, participants tended to **project themselves in the depicted scenario** (P1, P2, P4, P7, P9, P10), as P1 recalled: "*The mountain and the lakes (Figure 2 (6)) were just so beautiful, it's like if I am in such a good view, maybe in New Zealand, I would feel peaceful and calm.*" Third, participants engaged in **imagining themselves interacting with the people in the image** (P1, P2, P5, P8). This not only involves projecting oneself in the depicted scenario but also considering how one would respond to the people and objects in that scenario. For example, seeing an old lady tired and upset, P9 imagined themselves talking to the lady: "*The old lady in the picture (Figure 2 (12)) kinda looks tired, and worn out like someone who has maybe spent all day with negative things. I felt like maybe I was talking to her and also felt depressed.*" Fourth, some participants were **invoked with related personal memory** (P8, P10): "*It reminds me of my sister. She just gave birth to a baby, so seeing this picture (Figure 2 (2)), I could totally feel how happy it is*" (P8).

4.2.2 Emotion Externalization Through Gestures. Upon understanding or experiencing the emotion, participants externalized that emotion by forming different gestures using a single hand. In particular, we identified four channels through which they made

a connection between the emotion and their gestures. Depending on the nuances of the emotions that participants experienced, it is common to express an emotion through multiple channels.

Communication Norms. Most commonly, participants relied on shared understandings and conventions regarding nonverbal communication to form a gesture (P1, P3, P5, P6, P7, P8, P9, P10, P11). The most representative examples are using "thumbs up," "ok," and "victory" to convey positive emotions, as P5 mentioned: "*I didn't think too much about it. The gesture (victory) for me basically means cool, that's what you show your friends when you feel awesome. so I just went for it.*" P3 also added: "*honestly these gestures (ok) are kind of like they came from everyday life.*" Similarly, a "closed fist" and "thumbs down" were considered as a conventional way to convey negative emotions, as P4 explained that "*thumbs down shows disagreement and I don't like it;*" and P11 believed that "*I think most people know, it (a closed fist) means frustrated or something.*"

Sometimes, the ways that participants formed the hand gestures were influenced by the communication norms in certain cultures (P1, P4, P5, P10). For instance, P1, a fan of street dance, used "number six" to express the feeling of excitement (Figure 4 (d)): "*my inspiration is from the street dance because when we see amazing poses, we respond just like this one (forming a "number six" gesture) with our hands.*" Likewise, P4 was inspired by the rock culture and used the "rock" gesture to express happiness (Figure 1 (b)): "*I like rock music and this just means let's rock or it is cool.*"

Figure 6: Four channels through which the emotions are elicited and understood, which led to four ways of expressing the emotion, and finally externalizing the emotion in the form of a single-hand gesture.

When asked about how to choose between existing gestures with similar meanings (e.g., "thumbs up" versus "number six"), participants found themselves being influenced by the elements in the image stimuli. For example, when presented with an image depicting a skydiving person and an image with two kids laughing, P1 felt *happy* (positive valence, high arousal) in both scenarios. However, they responded with "number six" to the former stimulus and "thumbs up" to the latter stimulus, because they considered "thumbs up" is more appropriate for kids.

Creative Embodiment. Beyond communication norms, participants also self-created gestures in non-conventional forms. They assigned meanings and narratives behind each gesture, forming a more personalized mode of expression (P1, P2, P3, P6, P7, P10). For example, when presented with the image stimulus of a person skydiving, P2 expressed the *excited* emotion (positive valence, high arousal) by quickly opening their palm to mimic the procedure of a parachute opening (Figure 5 (a2)). They explained: "*just the thought of skydiving also seems very adventurous so it's like Like there like a bird or letting go of stuff. I remember I did it (opening the palm) very fast because I think this is how fast it happens.*" P1 also created a narrative related to "letting it go" with an open palm (Figure 1 (j)), but in a different emotional state—*depressed* (negative valence, low arousal): "*When you feel depressed, usually that's because you lost something. So I'm opening my hand to let them go.*"

In another example, P7 expressed the *bored* emotion (negative valence, low arousal) with an open palm by repeatedly extending and flexing all the fingers to create an action of "pushing away" (Figure 5 (a)) as they explained: "*I used that gesture to get away the boredom that is making you feel bad, so like I am pushing it away.*" P6 also created a gesture of "pushing away" to express a feeling of *tense* (negative valence, high arousal) with a similar rationale, while they used a closed fist instead of an open palm: "*When you feel tense, you are like confronting whatever in front of you, so I'm trying to push it away a bit harder.*"

Furthermore, we found the elements in the image stimuli often served as a source to inspire participants' creation of gestures. For example, when presented with a scene featuring mountains and a lake (see Figure 2 (6)), several participants (P1, P2, P3, P11) tried to mimic the gentle wave of the water to convey the feeling of calmness (Figure 1 (g)), as P2 explained: "*This gesture allowed me to visually represent the serene and tranquil atmosphere depicted in the image.*" Sometimes, participants used their hands to mimic the posture of the person in the image. This was commonly observed in expressing *tired* and *depressed*, where the image stimuli included people who sat in corners alone, lowering their heads. In response to the stimuli, participants formed a loose "open palm", slowly "falling down" (Figure 5 (g)) to depict "*a person with no power*" (P1), "*a downhearted person*" (P7), and "*a depressed person who is going down and down*" (P11).

Physical Expression. Physical expression, differing from communication norms or creative embodiment, refers to a process of directly externalizing emotion based on one's physical instinct, and oftentimes, as a way to vent out the emotion. All the participants mentioned such physical expressions during the interviews. For example, in an attempt to express an emotion of negative valence and high arousal, we found several instances with a closed fist heavily shaking (Figure 5 (c)) or punching forward (Figure 5 (j)). As participants explained, this gesture serves as a physical manifestation of the intense emotions when they felt angry or frustrated (P1, P2, P3, P5, P6, P7, P8, P9, P10, P11): "*Personally, when I'm angry, I like to just hit some stuff like no point*" (P9). Another example is the use of a "scratch" gesture (Figure 4 (j)) to express frustration (P1, P3, P6, P7, P10, P11), as P3 elaborated: "*When you feel tense, lost, and don't know what to do, you just keep scratching your head, your leg, or somewhere else. So yeah, this is what I would do if I'm frustrated*" (P3). The "squeeze" motion (Figure 5 (h)) we found from the survey data was also used to convey similar emotions.

Abstract Expression. Occasionally, participants may resort to abstract expressions, where they could not articulate why and how they formed the gestures to express certain emotions (P2, P8, P10, P11). For example, P8 formed a "finger touch" gesture (Figure 4 (m)) to express *bored* emotion but found it difficult to provide a clear explanation: "*there's not much meaning. I thought actually, I was trying to mimic something that feels bored on my own.*" One reason

for this abstract expression was "*running out of gestures*" at the end of the survey (P2, P10), leading participants to create gestures that might not have a direct connection to the specific emotions they were trying to express. When P10 used a "finger touch" gesture to express the feeling of anger, they acknowledged that the gesture was not that intuitive, and partly it was due to the lack of sources: "*I wanted to show something strong but did not know how to do it. Also, I don't want to repeat my gestures.*"

4.3 Experience of Emotion Expression Using Single-Hand Gestures

Overall, participants found that using single-hand gestures to express different emotions is "*intuitive,*" "*convenient,*" and "*fun*" (P1, P2, P5, P6, P7, P9, P10, P11). They recognized the potential for single-hand gestures to serve as "*a new communication channel*" for individuals who cannot use both hands at the same time (P1, P7). However, participants also highlighted it can be challenging to express emotions with only a single hand at times, because "*one hand is too limited to present rich emotions, compared with both hands or other body languages*" (P1) and "*the gestures could be hard to understand by others because some of them are very personal*" (P3). Below, we elaborate on additional considerations that participants mentioned regarding expressing emotions at different valence and arousal levels.

4.3.1 Expressing Positive versus Negative Emotions. Reflecting on their own emotion expressions with single-hand gestures, participants realized that when expressing positive emotions, besides "*upward*" gestures such as thumbs up and victory, they also preferred "*more open and large*" gestures such as an open palm (P1, P2, P7). P2 explained this was because they would like to "*share the happiness and let others know*" and P7 believed that "*positive feelings are inherently open and welcoming.*" Negative emotions, on the contrary, were considered "*less open*" (P1, P2, P3) and "*personal*" (P7). Thus, participants tended to use a closed fist or scratch gesture to "close" themselves: "*I don't think I want to share anybody about my negative feelings.*" (P1).

4.3.2 Expressing High-Arousal versus Low-Arousal Emotions. Although our statistical test did not yield any significant results with motion-related features, several participants mentioned that they liked to add movements for emotions with high arousal levels (P1, P2, P3, P7, P8, P10). During the interviews, P3 explained the reasons for quickly shaking their fist to express the angry emotion (Figure 5 (c)): "*Because it is so intense, I just wanna shake my fist as fast as I could to express such intensity.*" P7 and P10, chose to repeat the action of downward pointing when they used "thumbs down" to convey a feeling of anger or frustration (Figure 5 (i)): "*compared with tired and depressed (emotions) where I did not move my hand a lot, this time I kept pointing down to express the anger*" (P10).

In expressing positive emotions, similarly, participants mentioned adding movements when they felt extremely excited or happy. A typical example is the "repeated upward pointing" with "thumbs up" (Figure 5 (k)), as P8 noted: "*I would associate something slow with being tired or sad or lazy, but if it's like fast, it's more positive or more intense.*" In another example, P1 kept rotating their hand while making a "number six" gesture (Figure 1 (c)) to express the high arousal emotion (e.g., happy). The observation of

repeated movements, in our analysis, was attributed to physical expression, as participants made the movements largely driven by their physical instincts.

4.3.3 Using Same Gestures to Express Different Emotions. We found several instances where participants used the same gestures, mostly commonly—an "open palm" to express different emotions. As we mentioned earlier, some participants felt an open palm conveys openness and positivity (P1, P2, P7), whereas others used an open palm to express a negative feeling (e.g., waving the palm to "*let it go*"). In part, participants acknowledged that the distinctions between emotions within the same valence-arousal quadrant were not always obvious (P5, P9, P11). For example, P5 used the same "open palm" gesture to express the feeling of calm and explained: "*But you know, being relaxed and being calm, is more of the same thing. I cannot really get a difference between them.*" Likewise, P9 used the "victory" gesture for both happiness and excitement, noting that "*maybe feeling excited is more intense but to me they are basically the same, which is that you feel good about something.*"

5 DISCUSSION

In this section, we first reflect on the connections between gestures and emotions revealed in our findings, then share the implications of these connections for designing single-hand gesture-based emotion-tracking technologies and beyond.

5.1 The Connections Between Gestures and Emotions

This work provides empirical evidence demonstrating that hand gestures, as a part of body language and mental expression, are intricately connected to the inner emotional world of human beings. These connections not only lie in the commonly known symbolic meanings associated with gestures, but also in multiple gestural features that are inherently linked to individuals' mental models.

First, aligning with prior work in gesture-based communication [23, 52, 56], our findings showed that certain gestures can serve as symbols for positive versus negative emotions. Moving one step further, we delved deeper into the relationships between different gestural features and emotion dimensions, including statistically significant trends in finger-pointing directions and gesture strength. We found an *upward* finger-pointing direction is related to positive valence and high-arousal emotions, while a *downward* finger-pointing direction is related to negative valence and low-arousal emotions. For example, a "thumbs up" is likely to convey an intense and strong positive feeling, and a "thumbs down," on the other hand, is likely to convey a negative feeling with less intensity. These findings echo previous findings on the association between the direction of bodily reactions (e.g., hand upwards, head upright, and raised upper lip) and positive emotions like joy and cheerfulness [32], but also extend to the motor movements of more specific body pars–hand gestures. Additionally, when it comes to gesture strength, we observed a significant trend only within negative emotions: a tight gesture was linked to high-arousal emotions, whereas a loose gesture was linked to low-arousal emotions. Thus, gesture strength can be useful for detecting the arousal level of negative emotions, but may not help discern the arousal level of positive emotions. These findings also suggested that gestures with different

names can exhibit the same finger-pointing direction and strength (e.g., "index finger one" can also point downward as "thumbs down;" both a "scratch" and a "grab" can be loose or tight). As such, relying solely on gesture names (shape and form) may not be adequate for determining the corresponding emotions.

Second, we did not find any significant relationships between emotion dimensions and other gestural features, including palm direction, gesture ending status, and motion frequency. However, during the follow-up interviews, several participants mentioned adding additional movements or speeding up the movements to express high-arousal emotions. We suspect that this gap was partly due to individual preferences for employing moving gestures, as shown in our data, nearly 40% of the gestures did not involve motions. Also, we relied on human judgments to code the gestural features. Despite the efforts to establish a shared understanding, the coding results, such as motion frequency, may not be perfect (the limitations and practical challenges of the gesture coding procedure are described in Section 6). Building upon the interview findings, future work can expand the understanding of gesture-emotion connection primarily on gestural features such as the attributes of hands [5, 56] by revealing how hand movements, such as ending status and motion frequency, are related to emotions.

Third, we identified four channels through which participants externalized their emotions in response to the stimuli with hand gestures: communication norms, creative embodiment of an emotion's specific features, physical expression as if venting out the emotion, and representing the emotion based on intuition in abstract ways. These findings, on the one hand, resonate with prior work on the categorization of gestures as two types – *representational gestures* and *non-representational gestures* [5, 56]. That is, leveraging communication norms and creative embodiment belong to *representational gestures*, while physical expression and abstract expression belong to *non-representational gestures*. On the other hand, we expanded the understanding of this dichotomous gesture categorization [56]. In part, emotion is often externalized through multiple channels, where one may leverage communication norms and physical expression (e.g., a "thumbs down" repeatedly pointing downward). Additionally, individuals' gender, age, cultural practices, and emotional contexts (e.g., the stimuli influence) played important roles in shaping their gestures. In our survey, one related finding is that younger participants tended to employ more diverse gestures, suggesting that younger generations may be more willing to adopt and engage with gesture-based emotion expression technologies; however, more research needs to be done to understand the underlying reasons of this phenomenon (e.g., the influence of pop cultures such as music, movies, and social media [3]).

5.2 Design Implications for Emotion Characterization and Contextualization

In this section, we reflected on how participants chose the words representing their experienced emotions and used gestures to convey the meanings behind these emotions, which led to important implications on how designers of emotion tracking systems could better characterize and contextualize the captured emotions.

5.2.1 Emotion Characterization. Our survey results showed that participants tended to choose words representing high-arousal emotions rather than low-arousal emotions. One explanation could be that high-arousal emotions are more intense and captivating in nature [31], and therefore are more noticeable. Also, the word choice may be influenced by participants' personal experiences— the frequency of encountering certain emotions in daily life. Words such as *happy*, *excited*, and *angry* are more commonly used and discussed in daily life, compared with those such as *content*, *calm*, and *tired*. As a result, participants might be more familiar with these words and find them readily available for expressing their emotions [61]. During the interviews, participants mentioned not being able to discern the difference between certain emotions (e.g., *relaxed* versus *calm*). This could be due to the inherent similarity of emotions in the same valence-arousal quadrant [75], or participants' infrequent exposure to the words in life.

While not all participants ended up choosing 12 different emotions, the average and minimum number of emotion words chosen were 8 and 5, respectively. This finding suggests that the spectrum of emotional responses extends beyond the simplistic categorization of emotions by valence (positive, neutral, and negative), which are commonly employed by commercial emotion tracking systems [11] or questionnaires [46]. Future designs for emotion tracking should take into account this diversity in framing emotions.

Taken together, the nuances in participants' word choice in our survey led us to ask: *how to better characterize the captured emotions in a way that matches individuals' mental models?* Although valence and arousal are commonly used in academic research to describe emotions, they may not necessarily align with individuals' everyday expressions of emotion. That being said, while these two concepts are theoretically able to reflect on any emotions in a joint manner, it is hardly possible to accurately locate one's everyday emotions that vary in a subtle and instant manner in real-world scenarios. One opportunity is to gather their preferences and prior experiences and provide personalized options. This process can be facilitated by starting with high-frequency words from each valence-arousal quadrant (e.g., more words representing high-arousal emotions) and gradually guiding individuals to customize their preferences thereafter. Due to the intricate and complex nature of emotions, further research is needed to explore how individuals express emotions associated with additional meanings from other emotion models (e.g., the six emotions specified in the Basic Emotion Theory involving sadness and disgust [22]).

5.2.2 Emotion Contextualization with Gestures. As the source of emotion elicitation, the presented image stimuli inevitably added nuances to the emotions that participants experienced and to some extent, influenced their emotional responses and subsequent expressions through gestures. This influence was observed in all the participants, highlighting the contextual nature of emotions that are affected by various factors, including external stimuli, personal experiences, and individual interpretations. Therefore, when designing emotion tracking systems, capturing only the words or numeric values of the experienced emotions is not sufficient, and the importance of capturing what individuals were experiencing when certain emotions arose has been emphasized in prior work [11, 72, 78]. This emotional experience involves different contextual information, which can be automatically tracked (e.g., date, weather, physical activities, and location) or requires manual annotation (e.g., events, people, and elaboration on other details).

During the interviews, we found that even with a single hand, participants were able to convey rich meanings alongside the image stimuli, which pinpoints how representational gestures provide an implicit way to understand the underlying mechanism of semantic-driven information carried by hand gestures. This brought up a new possibility of contextualizing emotions with gestures. For example, a slowly waving "open palm" mimicking the gentle water waves can refer to an unspoken yet personalized record to depict a relaxed vacation; a hand gradually falling down may reflect a tiring person grappling with work stress; and a rotating "number six" could resemble an exhilarating concert memory, etc. As these gestures are easy to perform, they can serve as situated references that provide contextual cues of the experienced emotions. Through creative embodiment, individuals can incorporate their personal experience into a simple hand gesture, which can accommodate the diverse ways in which individuals express their emotions and potentially foster self-awareness as a way to enhance mental wellbeing [44]. It is also worth pointing out, that younger participants were more likely to employ more diverse gestures. This observation suggested that younger individuals may find it more natural and accessible to engage in gesture-based emotion contextualization. Therefore, the age of target users and their preferences for and behavioral patterns of gestures should be fully understood to inform the design of user-centered gesture-based emotion-tracking systems.

5.3 Design Implications for Single-Hand Gesture Input in Emotion Tracking

While our work opens up opportunities for single-hand gestures to support emotion tracking, we are not ready to develop such a system for daily use due to several unanswered questions and challenges. Below, we discuss the scenarios where gesture input can be helpful in emotion tracking and implications for technologies.

5.3.1 Enhancing Multimodal Emotion Tracking. Although many wearable and sensing technologies can accurately detect emotion arousal, capturing the valence has been more challenging [21, 64, 70, 86, 94]. For instance, heart rate (HR) or heart rate variability (HRV) captured by a smart watch can suggest when a person feels aroused, but it is unclear whether the arousal is from a feeling of happiness or anger [21, 70]. As such, some mobile health apps (e.g., Cardiogram [14]) allow users to manually label their emotions alongside their recorded heart rate data, which provides a better understanding of how one's HR is correlated with different emotional states, but imposes a burden on recollecting their emotional experiences. In this case, a single-hand gesture can come into play. When an abnormal HR or HRV is detected, the watch can alert the wearer through a vibrate notification with a short message asking how they are doing at the moment. To respond, the wearer could simply perform a "thumbs up," "thumbs down," or "open palm" to indicate if they are in a positive, negative, or neutral mood. This effortless interaction enables quick integration of one's subjective feelings into the emotion recognition system in real-time, lowering the labeling effort and reducing recall bias. Although some gesture-based expressions can be abstract and hard to interpret, they still offer an avenue for individuals to convey their emotional state without explicitly disclosing sensitive information, which to some extent preserves individuals' privacy. To better inform the recognition of individual emotions with gesture input, future systems can provide an entry for users to revisit and edit the meanings of their gestures through text or speech input [50, 51], as a way to reinforce personalized learning.

Gesture input can also be integrated into multimodal systems to enrich the meaning of emotions. Drawing from our findings on the relationships between emotion dimensions and gestural features, the nuances in finger-pointing direction and gesture strength can be used to infer the valence and arousal of the corresponding emotion. Besides, several participants preferred using "larger and open" gestures to express positivity and willingness for sharing. Interestingly, a recent work by Asaliouglu et al. also analyzed the relationships between gesture size and emotional states and showed that narrow gestures were connected to a higher level of arousal [5]. Taking this work together with ours points to opportunities for examining the connections between emotional dimensions and gestural features that were not characterized in this study. In addition to static features such as gesture size, motion features such as moving range and fluidity can add insights into the emotion externalization process, which warrants lab experiments equipped with precise motion and wearable sensors.

5.3.2 Beyond Emotion Tracking. As Koh et al. noted in their mapping between symbolic gestures and commonly used emojis [36, 37], gesture input has great potential to enhance the communication experience in instant messaging. Our work offers the possibility to expand such communication on multiple platforms and devices such as wearable devices (similar to emotion labeling mentioned above) and motion sensors at a short distance [83], as well as in a wide range of contexts especially when typing or speaking is not convenient (e.g., public spaces). Our participants also brought up the idea of using single-hand gestures to facilitate communication when one of their hands is not available. Moreover, the insights gained from this study can be applied in human-robot communication. As Zhou et al. found in their work, humans perceive and interpret robots' emotions through tactile gestures (e.g., shaking, pushing, patting) in a similar manner to how they interact with other humans [95, 96]. While our work focused on nontactile gesture interaction, we see the opportunities to incorporate tactile stimulation in the future and empower individuals to establish a more intuitive and personalized means of communicating with robots. For example, one could define their own gesture sets as commands that instruct a robot to perform certain tasks, and the robot could simulate its user's gesture patterns as part of responding actions, further enhancing the engagement of the interaction.

6 LIMITATIONS AND FUTURE WORK

To extract the gestural features, we relied on human judgments which may introduce inaccurate interpretations. We acknowledge that, due to the diverse shapes and forms of the collected gestures, it was challenging to directly extract these gestural features automatically using existing technologies (e.g., computer vision [30]). However, this method has been widely used in prior research [35, 69, 88], and we followed a rigorous procedure to iteratively develop the codebook, train the research assistant, and calculate the reliability scores. These steps ensured consistency and minimized subjective biases to the best extent possible.

Our findings around the connections between gestural features and emotional dimensions may not be generalizable, due to the subjective and creative nature of the collected gestures. Relatedly, although our participants came from different regions, the sample distribution may not be adequately diverse regarding cultural backgrounds, as those living in Western countries took up over 77%. Nevertheless, our analysis of the 756 gesture photos and videos provided rich insights on how single-hand gestures could convey the ways individuals experience and express their emotions.

Additionally, among the 63 participants who took part in the survey, we interviewed only 11 of them. Thus, we do not have an explanation of the gesture-forming rationales from all the participants. Given our aim to qualify the gesture-forming rationales instead of quantifying them, our interview data has provided a deep understanding to complement the survey data regarding how participants externalized their emotions through gestures and the factors influencing the externalization.

Going forward, we plan to expand this work by incorporating a more balanced sample from diverse regions, which would allow us to quantitatively examine the geographic and cultural differences in people's gestural features. Another important next step involves developing research prototypes for emotion tracking that incorporate gesture input, along with other input modalities such as heart rate sensor, text entry, and speech.

7 CONCLUSION

Towards an applicable and effortless way of using gestures to inform emotion capture, this work set out to gain an empirical understanding of how individuals express their emotions through single-hand gestures. We conducted a survey study that collected 756 gestures (captured in photos and videos) from 63 participants who were presented with 12 image stimuli for emotion elicitation and then interviewed 11 of these participants to understand their gesture-forming rationales. Our studies revealed that the valence and arousal of emotions had a significant impact on participants' finger-pointing direction and gesture strength. Additionally, we unfolded four channels through which participants externalized their experienced emotions through single-hand gestures, including communication norms, creative embodiment, physical expression, and abstract expression. With the lessons learned, we discuss the implications for single-hand gestures to facilitate emotion tracking in different scenarios and opportunities for leveraging this understanding of embodied cognition in research areas such as computer-mediated communication, and human-robot interaction.

REFERENCES

[1] Francisca Adoma Acheampong, Chen Wenyu, and Henry Nunoo-Mensah. 2020. Text-based emotion detection: Advances, challenges, and opportunities. *Engineering Reports* 2, 7 (2020), e12189. https://doi.org/10.1002/eng2.12189

[2] Ferdous Ahmed, ASM Hossain Bari, and Marina L Gavrilova. 2019. Emotion recognition from body movement. *IEEE Access* 8 (2019), 11761–11781. http://doi.org/10.1109/ACCESS.2019.2963113

[3] Donna E Alvermann. 2011. 23 Popular Culture and Literacy Practices. *Handbook of Reading Research, Volume IV* (2011), 541. https://journals.sagepub.com/doi/pdf/10.2304/ciec.2001.2.3.3

[4] Allison A Appleton, Stephen L Buka, Eric B Loucks, Stephen E Gilman, and Laura D Kubzansky. 2013. Divergent associations of adaptive and maladaptive emotion regulation strategies with inflammation. *Health Psychology* 32, 7 (2013), 748. https://pubmed.ncbi.nlm.nih.gov/23815767/

[5] Esma Nur Asalıoğlu and Tilbe Göksun. 2023. The role of hand gestures in emotion communication: Do type and size of gestures matter? *Psychological Research* 87, 6 (2023), 1880–1898. https://doi.org/10.1007/s00426-022-01774-9

[6] Lawrence W. Barsalou. 2008. Grounded Cognition. *Annual Review of Psychology* 1 (2008), 617–645. https://doi.org/10.1146/annurev.psych.59.103006.093639

[7] John N Bassili. 1979. Emotion recognition: the role of facial movement and the relative importance of upper and lower areas of the face. *Journal of personality and social psychology* 37, 11 (1979), 2049. https://pubmed.ncbi.nlm.nih.gov/521902/

[8] Alexandra H Bettis, Taylor A Burke, Jacqueline Nesi, and Richard T Liu. 2022. Digital technologies for emotion-regulation assessment and intervention: A conceptual review. *Clinical Psychological Science* 10, 1 (2022), 3–26. http://doi.org/10.1177/21677026211011982

[9] Michel Cabanac. 2002. What is emotion? *Behavioural processes* 60, 2 (2002), 69–83. https://www.sciencedirect.com/science/article/pii/S0376635702000785

[10] Geneviève Calbris. 2008. From left to right...: Coverbal gestures and their symbolic use of space. In *Metaphor and gesture*. John Benjamins, 27–53.

[11] Clara Caldeira, Yu Chen, Lesley Chan, Vivian Pham, Yunan Chen, and Kai Zheng. 2017. Mobile apps for mood tracking: an analysis of features and user reviews. In *AMIA Annual Symposium Proceedings*, Vol. 2017. American Medical Informatics Association, 495. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5977660/

[12] Walter B Cannon. 1927. The James-Lange theory of emotions: A critical examination and an alternative theory. *The American journal of psychology* 39, 1/4 (1927), 106–124. https://www.jstor.org/stable/1415404

[13] Nicola Carbonaro, Alberto Greco, Gaetano Anania, Gabriele Dalle Mura, Alessandro Tognetti, EP Scilingo, Danilo De Rossi, Antonio Lanata, et al. 2012. Unobtrusive physiological and gesture wearable acquisition system: a preliminary study on behavioral and emotional correlations. *Global Health* (2012), 88–92. https://arpi.unipi.it/handle/11568/220745

[14] Cardiogram, Inc. 2023. Cardiogram. https://cardiogram.com/.

[15] Ginevra Castellano, Loic Kessous, and George Caridakis. 2008. Emotion recognition through multiple modalities: face, body gesture, speech. *Affect and Emotion in Human-Computer Interaction: From Theory to Applications* (2008), 92–103. https://doi.org/10.1007/978-3-540-85099-1_8

[16] Edwin Chan, Teddy Seyed, Wolfgang Stuerzlinger, Xing-Dong Yang, and Frank Maurer. 2016. User elicitation on single-hand microgestures. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 3403–3414. https://doi.org/10.1145/2858036.2858589

[17] Chih-Hao Chen, Wei-Po Lee, and Jhih-Yuan Huang. 2018. Tracking and recognizing emotions in short text messages from online chatting services. *Information Processing & Management* 54, 6 (2018), 1325–1344. https://doi.org/10.1016/j.ipm.2018.05.008

[18] James A Cranford, Patrick E Shrout, Masumi Iida, Eshkol Rafaeli, Tiffany Yip, and Niall Bolger. 2006. A procedure for evaluating sensitivity to within-person change: Can mood measures in diary studies detect change reliably? *Personality and Social Psychology Bulletin* 32, 7 (2006), 917–929. https://doi.org/10.1177/0146167206287721

[19] Nele Dael, Martijn Goudbeek, and Klaus R Scherer. 2013. Perceived gesture dynamics in nonverbal expression of emotion. *Perception* 42, 6 (2013), 642–657. https://journals.sagepub.com/doi/abs/10.1068/p7364

[20] Karina W Davidson, Elizabeth Mostofsky, and William Whang. 2010. Don't worry, be happy: positive affect and reduced 10-year incident coronary heart disease: the Canadian Nova Scotia Health Survey. *European heart journal* 31, 9 (2010), 1065–1070. https://academic.oup.com/eurheartj/article/31/9/1065/590670

[21] Xianghua Ding, Shuhan Wei, Xinning Gui, Ning Gu, and Peng Zhang. 2021. Data Engagement Reconsidered: A Study of Automatic Stress Tracking Technology in Use. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–13. https://doi.org/10.1145/3411764.3445763

[22] Paul Ekman. 1984. Expression and the nature of emotion. *Approaches to emotion* 3, 19 (1984), 344. https://api.semanticscholar.org/CorpusID:140982571

[23] Lauren Gawne and Gretchen McCulloch. 2019. Emoji as digital gestures. *Language@ Internet* 17, 2 (2019). https://www.languageatinternet.org/articles/2019/gawne/index_html

[24] James J Gross and Robert W Levenson. 1995. Emotion elicitation using films. *Cognition & emotion* 9, 1 (1995), 87–108. https://doi.org/10.1080/02699939508408966

[25] Steven C Hauser, Sarah McIntyre, Ali Israr, Håkan Olausson, and Gregory J Gerling. 2019. ncovering human-to-human physical interactions that underlie emotional and affective touch communication. In *2019 IEEE world haptics conference (WHC)*. IEEE, 407–412. https://doi.org/10.1109/WHC.2019.8816169

[26] Matthew J Hertenstein, Dacher Keltner, Betsy App, Brittany A Bulleit, and Ariane R Jaskolka. 2006. Touch communicates distinct emotions. *Emotion* 6, 3 (2006), 528. https://doi.org/10.1037/1528-3542.6.3.528

[27] Jari K Hietanen, Jukka M Leppänen, and Ulla Lehtonen. 2004. Perception of emotions in the hand movement quality of Finnish sign language. *Journal of nonverbal behavior* 28 (2004), 53–64. https://link.springer.com/article/10.1023/B:JONB.0000017867.70191.68

[28] Autumn B Hostetter, Martha W Alibali, and Sotaro Kita. 2007. I see it in my hands' eye: Representational gestures reflect conceptual demands. *Language and cognitive processes* 22, 3 (2007), 313–336. https://www.tandfonline.com/doi/full/10.1080/01690960600632812

[29] William James. 1884. What is an Emotion? *Mind* 9, 34 (1884), 188–205. http://www.jstor.org/stable/2246769

[30] Muskan Jindal, Eshan Bajal, and Shilpi Sharma. 2023. A Comparative Analysis of Established Techniques and Their Applications in the Field of Gesture Detection. *Machine Learning Algorithms and Applications in Engineering* 73 (2023). https://www.taylorfrancis.com/chapters/edit/10.1201/9781003104858-5/comparative-analysis-established-techniques-applications-field-gesture-detection-muskan-jindal-eshan-bajal-shilpi-sharma

[31] Georgios Karafotias, Akiko Teranishi, Georgios Korres, Friederike Eyssel, Scandar Copti, and Mohamad Eid. 2017. Intensifying emotional reactions via tactile gestures in immersive films. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)* 13, 3 (2017), 1–17. https://doi.org/10.1145/3092840

[32] Dacher Keltner, Disa Sauter, Jessica Tracy, and Alan Cowen. 2019. Emotional expression: Advances in basic emotion theory. *Journal of nonverbal behavior* 43 (2019), 133–160. https://link.springer.com/article/10.1007/s10919-019-00293-3

[33] Adam Kendon. 1988. How gestures can become like words.. In *This paper is a revision of a paper presented to the American Anthropological Association, Chicago, Dec 1983.* Hogrefe & Huber Publishers. https://psycnet.apa.org/record/1992-98173-004

[34] Jonghwa Kim and Elisabeth André. 2008. Emotion recognition based on physiological changes in music listening. *IEEE transactions on pattern analysis and machine intelligence* 30, 12 (2008), 2067–2083. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4441720

[35] Michael Kipp and Jean-Claude Martin. 2009. Gesture and emotion: Can basic gestural form features discriminate emotions?. In *2009 3rd international conference on affective computing and intelligent interaction and workshops.* IEEE, 1–8. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5349544

[36] Jung In Koh, Josh Cherian, Paul Taele, and Tracy Hammond. 2019. Developing a hand gesture recognition system for mapping symbolic hand gestures to analogous emojis in computer-mediated communication. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 9, 1 (2019), 1–35. https://doi.org/10.1145/3297277

[37] Jung In Koh, Samantha Ray, Josh Cherian, Paul Taele, and Tracy Hammond. 2022. Show of Hands: Leveraging Hand Gestural Cues in Virtual Meetings for Intelligent Impromptu Polling Interactions. In *27th International Conference on Intelligent User Interfaces.* 292–309. https://doi.org/10.1145/3490099.3511153

[38] Uros Krcadinac, Philippe Pasquier, Jelena Jovanovic, and Vladan Devedzic. 2013. Synesketch: An open source library for sentence-based emotion recognition. *IEEE Transactions on Affective Computing* 4, 3 (2013), 312–325. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6589580

[39] Sylvia D Kreibig. 2010. Autonomic nervous system activity in emotion: A review. *Biological psychology* 84, 3 (2010), 394–421. https://www.sciencedirect.com/science/article/pii/S0301051110000827

[40] Janina Künecke, Andrea Hildebrandt, Guillermo Recio, Werner Sommer, and Oliver Wilhelm. 2014. Facial EMG responses to emotional expressions are related to emotion perception ability. *PLoS one* 9, 1 (2014), e84053. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0084053

[41] Benedek Kurdi, Shayn Lozano, and Mahzarin R Banaji. 2017. Introducing the open affective standardized image set (OASIS). *Behavior research methods* 49 (2017), 457–470. https://doi.org/10.3758/s13428-016-0715-3

[42] Krzysztof Kutt, Grzegorz J Nalepa, Barbara Giżycka, Pawel Jemiolo, and Marcin Adamczyk. 2018. Bandreader-a mobile application for data acquisition from wearable devices in affective computing experiments. In *2018 11th International Conference on Human System Interaction (HSI).* IEEE, 42–48. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8431271

[43] Carl Georg Lange. 1885. The mechanism of the emotions. *The classical psychologists* (1885), 672–684. https://psychclassics.yorku.ca/Lange/

[44] Kwangyoung Lee and Hwajung Hong. 2017. Designing for self-tracking of emotion and experience with tangible modality. In *Proceedings of the 2017 Conference on Designing Interactive Systems.* 465–475. https://dl.acm.org/doi/abs/10.1145/3064663.3064697

[45] Min Seop Lee, Yun Kyu Lee, Dong Sung Pae, Myo Taeg Lim, Dong Won Kim, and Tae Koo Kang. 2019. Fast emotion recognition based on single pulse PPG signal with convolutional neural network. *Applied Sciences* 9, 16 (2019), 3355. https://www.mdpi.com/2076-3417/9/16/3355

[46] Stephanie Lichtenfeld, Reinhard Pekrun, Robert H Stupnisky, Kristina Reiss, and Kou Murayama. 2012. Measuring students' emotions in the early years: the achievement emotions questionnaire-elementary school (AEQ-ES). *Learning and Individual differences* 22, 2 (2012), 190–201. https://www.sciencedirect.com/science/article/pii/S1041608011000586

[47] Jia-Qing Liu, Kotaro Furusawa, Seiju Tsujinaga, Tomoko Tateyama, Yutaro Iwamoto, and Yen-Wei Chen. 2019. MaHG-RGBD: A multi-angle view hand gesture RGB-D dataset for deep learning based gesture recognition and baseline evaluations. In *2019 IEEE International Conference on Consumer Electronics (ICCE).* IEEE, 1–4. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8661941

[48] Yang Liu, Chengdong Lin, and Zhenjiang Li. 2021. WR-Hand: Wearable armband can track user's hand. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 3 (2021), 1–27. https://doi.org/10.1145/3478112

[49] Hong Lu, Denise Frauendorfer, Mashfiqui Rabbi, Marianne Schmid Mast, Gokul T Chittaranjan, Andrew T Campbell, Daniel Gatica-Perez, and Tanzeem Choudhury. 2012. Stresssense: Detecting stress in unconstrained acoustic environments using smartphones. In *Proceedings of the 2012 ACM conference on ubiquitous computing.* 351–360. https://dl.acm.org/doi/abs/10.1145/2370216.2370270

[50] Yuhan Luo, Young-Ho Kim, Bongshin Lee, Naeemul Hassan, and Eun Kyoung Choe. 2021. FoodScrap: Promoting Rich Data Capture and Reflective Food Journaling Through Speech Input.. In *Proceedings of the 2021 Conference on Designing Interactive System.* ACM. https://doi.org/10.1145/3461778.3462074

[51] Yuhan Luo, Bongshin Lee, Young-Ho Kim, Eun Kyoung Choe, et al. 2022. NoteWordy: Investigating Touch and Speech Input on Smartphones for Personal Data Capture. *proc. of the ACM on Human-Computer Interaction, Interactive Surfaces and Spaces (ISS), (To appear)* (2022). https://doi.org/10.1145/3567734

[52] Jiahao Ma, Tailun Li, and Jingfei He. 2020. A comprehensive hand gesture dataset. In *2020 International Conference on Computing and Data Science (CDS).* IEEE, 328–333. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9275984

[53] Leela Surya Teja Mangamuri, Lakshay Jain, and Abhishek Sharmay. 2019. Two hand Indian sign language dataset for benchmarking classification models of machine learning. In *2019 International conference on issues and challenges in intelligent computing techniques (ICICT),* Vol. 1. IEEE, 1–5. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8977713

[54] Nora McDonald, Sarita Schoenebeck, and Andrea Forte. 2019. Reliability and Inter-Rater Reliability in Qualitative Research: Norms and Guidelines for CSCW and HCI Practice. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 72 (nov 2019), 23 pages. https://doi.org/10.1145/3359174

[55] Andrew T McKenzie, Igor Katsyv, Won-Min Song, Minghui Wang, and Bin Zhang. 2016. DGCA: a comprehensive R package for differential gene correlation analysis. *BMC systems biology* 10 (2016), 1–25. https://doi.org/10.1186/s12918-016-0349-1

[56] David McNeill. 1992. *Hand and mind: What gestures reveal about thought.* University of Chicago press.

[57] David McNeill. 2019. *Gesture and thought.* University of Chicago press.

[58] Douglas S Mennin, Richard G Heimberg, Cynthia L Turk, and David M Fresco. 2005. Preliminary evidence for an emotion dysregulation model of generalized anxiety disorder. *Behaviour research and therapy* 43, 10 (2005), 1281–1310. https://www.sciencedirect.com/science/article/pii/S0005796704002323

[59] Javier Molina, José A Pajuelo, Marcos Escudero-Viñolo, Jesús Bescós, and José M Martínez. 2014. A natural and synthetic corpus for benchmarking of hand gesture recognition systems. *Machine Vision and Applications* 25, 4 (2014), 943–954. https://link.springer.com/article/10.1007/s00138-013-0576-z

[60] Kevin W Mossholder, Randall P Settoon, Stanley G Harris, and Achilles A Armenakis. 1995. Measuring emotion in open-ended survey responses: An application of textual data analysis. *Journal of management* 21, 2 (1995), 335–355. https://www.sciencedirect.com/science/article/pii/0149206395900616

[61] Constantino Méndez-Bértolo, Miguel A. Pozo, and José A. Hinojosa. 2011. Word frequency modulates the processing of emotional words: Convergent behavioral and electrophysiological data. *Neuroscience Letters* 494, 3 (2011), 250–254. https://doi.org/10.1016/j.neulet.2011.03.026

[62] Pansy Nandwani and Rupali Verma. 2021. A review on sentiment analysis and emotion detection from text. *Social Network Analysis and Mining* 11, 1 (2021), 81. https://doi.org/10.1007/s13278-021-00776-6

[63] Richard E Nisbett and Timothy D Wilson. 1977. Telling more than we can know: Verbal reports on mental processes. *Psychological review* 84, 3 (1977), 231. https://psycnet.apa.org/record/1978-00295-001

[64] D Oude Bos. 2006. EEG-based emotion recognition-The Influence of Visual and Auditory Stimuli. *Capita Selecta (MSc course)* (2006), 1–17. https://api.semanticscholar.org/CorpusID:16285681

[65] Aslı Özyürek. 2014. Hearing and seeing meaning in speech and gesture: Insights from brain and behaviour. *Philosophical Transactions of the Royal Society B: Biological Sciences* 369, 1651 (2014), 20130296. https://royalsocietypublishing.org/doi/full/10.1098/rstb.2013.0296

[66] Min Woo Park, Chi Jung Kim, Mincheol Hwang, and Eui Chul Lee. 2013. Individual emotion classification between happiness and sadness by analyzing photoplethysmography and skin temperature. In *2013 Fourth World Congress on Software Engineering.* IEEE, 190–194. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6754284

[67] William D. Perreault and Laurence E. Leigh. 1989. Reliability of Nominal Data Based on Qualitative Judgments. *Journal of Marketing Research* 26 (1989), 135 – 148. https://api.semanticscholar.org/CorpusID:144279197

[68] Stefano Piana, Alessandra Stagliano, Francesca Odone, Alessandro Verri, and Antonio Camurri. 2014. Real-time automatic emotion recognition from body

gestures. *arXiv preprint arXiv:1402.5047* (2014). https://doi.org/10.48550/arXiv.1402.5047

[69] Isabella Poggi. 2007. *Mind, hands, face and body: a goal and belief view of multi-modal communication.* Weidler.

[70] Raj Rakshit, V Ramu Reddy, and Parijat Deshpande. 2016. Emotion detection and recognition using HRV features derived from photoplethysmogram signals. In *Proceedings of the 2nd workshop on Emotion Representations and Modelling for Companion Systems.* 1–6. https://dl.acm.org/doi/abs/10.1145/3009960.3009962

[71] Judy S Reilly, Marina L McIntire, and Howie Seago. 1992. Affective prosody in American sign language. *Sign Language Studies* (1992), 113–128. https://www.jstor.org/stable/26204636

[72] Verónica Rivera-Pelayo, Angela Fessl, Lars Müller, and Viktoria Pammer. 2017. Introducing mood self-tracking at work: Empirical insights from call centers. *ACM Transactions on Computer-Human Interaction (TOCHI)* 24, 1 (2017), 1–28. https://doi.org/10.1145/3014058

[73] Alfredo Rodríguez-Muñoz, Mirko Antino, Paula Ruiz-Zorrilla, and Eric Ortega. 2021. Positive emotions, engagement, and objective academic performance: A weekly diary study. *Learning and Individual Differences* 92 (2021), 102087. https://www.sciencedirect.com/science/article/pii/S1041608021001242

[74] Tim Rohrer. 2006. *The Body in Space: Dimensions of Embodiment.* Vol. 1. 339–378.

[75] James A Russell. 1980. A circumplex model of affect. *Journal of personality and social psychology* 39, 6 (1980), 1161. https://doi.org/10.1037/h0077714

[76] Roland T. Rust and Bruce Cooil. 1994. Reliability Measures for Qualitative Data: Theory and Implications. *Journal of Marketing Research* 31 (1994), 1 – 14. https://api.semanticscholar.org/CorpusID:149884833

[77] Stanley Schachter and Jerome Singer. 1962. Cognitive, social, and physiological determinants of emotional state. *Psychological review* 69, 5 (1962), 379. https://psycnet.apa.org/record/1963-06064-001

[78] Stephen M Schueller, Martha Neary, Jocelyn Lai, and Daniel A Epstein. 2021. Understanding people's use of and perspectives on mood-tracking apps: interview study. *JMIR mental health* 8, 8 (2021), e29368. https://mental.jmir.org/2021/8/e29368

[79] Lawrence Shapiro. 2014. *The routledge handbook of embodied cognition.* Routledge. 1–400 pages. https://doi.org/10.4324/9781315775845

[80] Adwait Sharma, Joan Sol Roo, and Jürgen Steimle. 2019. Grasping microgestures: Eliciting single-hand microgestures for handheld objects. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems.* 1–13. https://doi.org/10.1145/3290605.3300632

[81] Donald Sharpe. 2015. Chi-square test is statistically significant: Now what? *Practical Assessment, Research, and Evaluation* 20, 1 (2015), 8. https://doi.org/10.7275/tbfa-x148

[82] Jainendra Shukla, Miguel Barreda-Angeles, Joan Oliver, Gora Chand Nandi, and Domenec Puig. 2019. Feature extraction and selection for emotion recognition from electrodermal activity. *IEEE Transactions on Affective Computing* 12, 4 (2019), 857–869. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8653316

[83] Stuart Taylor, Cem Keskin, Otmar Hilliges, Shahram Izadi, and John Helmes. 2014. Type-hover-swipe in 96 bytes: A motion sensing mechanical keyboard. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* 1695–1704. https://doi.org/10.1145/2556288.2557030

[84] Jochen Triesch and Christoph Von Der Malsburg. 2001. A system for person-independent hand posture recognition against complex backgrounds. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23, 12 (2001), 1449–1453. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=977568

[85] Timothy J Trull and Ulrich W Ebner-Priemer. 2020. Ambulatory assessment in psychopathology research: A review of recommended reporting guidelines and current practices. *Journal of Abnormal Psychology* 129, 1 (2020), 56. https://psycnet.apa.org/record/2019-79779-007

[86] Goran Udovičić, Jurica Đerek, Mladen Russo, and Marjan Sikora. 2017. Wearable emotion recognition system based on GSR and PPG signals. In *Proceedings of the 2nd international workshop on multimedia for personal health and health care.* 53–59. https://dl.acm.org/doi/abs/10.1145/3132635.3132641

[87] Santiago Villarreal-Narvaez, Jean Vanderdonckt, Radu-Daniel Vatavu, and Jacob O Wobbrock. 2020. A systematic review of gesture elicitation studies: What can we learn from 216 studies?. In *Proceedings of the 2020 ACM designing interactive systems conference.* 855–872. https://doi.org/10.1145/3357236.3395511

[88] Rebecca A Webb. 1997. *Linguistic features of metaphoric gestures.* University of Rochester.

[89] Hongyi Wen, Julian Ramos Rojas, and Anind K Dey. 2016. Serendipity: Finger gesture recognition using an off-the-shelf smartwatch. In *Proceedings of the 2016 CHI conference on human factors in computing systems.* 3847–3851. https://doi.org/10.1145/2858036.2858466

[90] Margaret Wilson. 2002. Six views of embodied cognition. *Psychometric Bulletin & Review* 9 (2002), 625–636. Issue 4. https://link.springer.com/article/10.3758/BF03196322

[91] Jacob O Wobbrock, Htet Htet Aung, Brandon Rothrock, and Brad A Myers. 2005. Maximizing the guessability of symbolic input. In *CHI'05 extended abstracts on Human Factors in Computing Systems.* 1869–1872. https://doi.org/10.1145/1056808.1057043

[92] Jacob O. Wobbrock, Meredith Ringel Morris, and Andrew D. Wilson. 2009. User-defined gestures for surface computing. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2009). https://api.semanticscholar.org/CorpusID:1341945

[93] Siqing Wu, Tiago H Falk, and Wai-Yip Chan. 2011. Automatic speech emotion recognition using modulation spectral features. *Speech communication* 53, 5 (2011), 768–785. https://www.sciencedirect.com/science/article/pii/S0167639310001470

[94] Qiang Zhang, Xianxiang Chen, Qingyuan Zhan, Ting Yang, and Shanhong Xia. 2017. Respiration-based emotion recognition with deep learning. *Computers in Industry* 92 (2017), 84–90. https://www.sciencedirect.com/science/article/pii/S0166361516303104

[95] Ran Zhou, Harpreet Sareen, Yufei Zhang, and Daniel Leithinger. 2022. EmotiTactor: Exploring How Designers Approach Emotional Robotic Touch. In *Designing Interactive Systems Conference.* 1330–1344. https://doi.org/10.1145/3532106.3533487

[96] Ran Zhou, Zachary Schwemler, Akshay Baweja, Harpreet Sareen, Casey Lee Hunt, and Daniel Leithinger. 2023. TactorBots: A Haptic Design Toolkit for Out-of-lab Exploration of Emotional Robotic Touch. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems.* 1–19. https://doi.org/10.1145/3544548.3580799

8 APPENDIX A: SURVEY QUESTIONS

For each of the 12 image stimuli, we asked participants to answer 4 questions. The images are presented in a random order.

1. Which of the following words **best** describes how the image makes you feel?

A. Excited	B. Delighted	C. Happy	D. Content
E. Relaxed	F. Calm	G. Tense	H. Angry
I. Frustrated	J. Depressed	K. Bored	L. Tired

2. How positive or negative does the picture make you feel, from 1 (very negative) to 7 (very positive)? Please use the full range of the scale to mark your responses rather than relying on only a few points.

3. Please rate how strong the feeling above is from 1 (low arousal) to 7 (high arousal)? Here, you can neglect how positive or negative the feeling is but focus on the *****intensity***** of the feeling. Please use the full range of the scale to mark your responses rather than relying on only a few points.

4. Now think about how you would use a *****single-hand***** gesture to express this emotion. There is no right or wrong answer. You can be as creative as possible.

- Upload a static gesture photo for the emotion (.gif, .jpg, .png, .bmp). Note that you do not have to show your face in the photo.
- Upload a short video to capture the motion of forming the gesture in the photo you just uploaded (recommended: 2 − 5s; supported format: m4v, mp4, mov, flv, avi, m4a, webm). Note that you do not have to show your face in the video.

9 APPENDIX B: EXAMPLES OF CODED GESTURAL FEATURES WITH THE ORIGINAL PHOTOS

Gesture photo	Emotion word chosen	Valence–Arousal	Gesture name	Palm direction	Finger pointing direction	Strength	Motion name	Ending status	Motion freq.
	Excited	Positive high	Number six	Towards the body	Up	Unclear	Others-multiple (finger flexion + repeated palm flipping)	Moving	High
	Delighted	Positive high	Horn	Towards the body	Up	Unclear	Finger extension	Static	Middle
	Happy	Positive high	Victory	Down	Outwards the body	Unclear	Pronation	Static	Middle
	Content	Positive low	Open palm with fingers pressed together	Down	Side (left)	Unclear	Repeated "water wave" (hand gently waving up and down)	Static	Middle
	Calm	Positive low	Ok	Side(left)	Up	Unclear	Repeated knocking	Moving	Middle
	Relaxed	Positive low	Finger heart	Side(left)	Up	Unclear	Repeated palm flipping	Moving	High
	Tense	Negative high	Closed fist	Outwards the body	None	Unclear	Finger flexion	Static	Middle
	Angry	Negative high	Scratch	Towards the body	None	Tight	Finger flexion	Static	Middle
	Frustrated	Negative high	Index finger one	Up	Outwards the body	Tight	Finger extension	Static	Middle
	Depressed	Negative low	Scoop	Towards the body	Towards the body	Loose	Finger flexion	Static	Low
	Bored	Negative low	Thumb down	Outwards the body	Down	Unclear	Repeated finger pointing downward	Moving	Middle
	Tired	Negative low	Grab	Down	Down	Loose	Wrist flexion	Static	Low