Emotional expressiveness and mimicry of online speakers

Amanveer Wesley, Panagiotis Tsiamyrtzis, Gloria Mark, and Ioannis Pavlidis, *Senior Member, IEEE*

Abstract—Facial communication between speakers and audience, in the form of measured emotional expressiveness and emotional mimicry, is important in establishing rapport. Here we present a multimodal method to quantify the role of stress, dispositional factors, and situational context in the emotional expressiveness and mimicry of online speakers. The said method uses a validated neural network to differentiate emotional from neutral facial expressions, a validated facial thermal imaging algorithm to measure stress, and the Big-Five inventory to determine dispositional traits. We used this method to investigate responses of not previously informed vs. duly informed speakers – scenarios that often arise in interviews and educational settings. The results suggest that stress, agreeableness, and unexpectedness positively correlate with speakers' emotional expressiveness. Interestingly, unexpectedness negatively correlates with speakers' emotional mimicry, putting not previously informed speakers at a distinct affiliative disadvantage. Our methodology opens the way for researching affiliative bonding under various virtual talk scenarios.

Index Terms—Online presentation, emotional mimicry, anticipation, facial expressions, agreeableness, stress, thermal imaging.

1 Introduction

It is hard to imagine a future of work without a significant remote component. As of the time of this writing, many businesses and educational institutions continue to conduct part of their operations remotely - a trend that appears to outlast the COVID-19 pandemic [1]. Synchronous meetings and presentations are a common component of remote work and education. The most widespread method for conducting synchronous meetings is videoconferencing. Compared to other media such as audioconferencing, video affords rich social cues for sharing audience feedback such as facial expressions [2]. Presenters can use such feedback to guide their speech. Because video of remote meetings primarily shows headshots or upper bodies (e.g., people seated around a conference table), facial expressions take on significance in influencing interaction.

Emotional expressiveness on individual faces has long been known to shape the impressions of observers [3]; in this respect, emotional expressiveness of virtual speakers and audience members should be no exception. Furthermore, facial expressions from an audience during a presentation can convey interest, disinterest, affirmation or disapproval to the speaker. Speakers adapt their behavior to audience feedback and one salient way is in the form of emotional mimicry. Emotional mimicry refers to when a speaker matches their facial expression to that of a speaking partner [4].

Emotional mimicry is shown to have a social function, to increase affiliation and build relationships [5]. Emotional mimicry is important to study during meeting presentations as it not only signals that the speaker and audience members are attentive and responding to each other, but mimicry can also be used to develop rapport. However, giving a virtual talk is not the same as delivering it in a face-to-face environment, where those present can see a range of nonverbal behaviors such as seating posture or gestures. In remote meetings, there are a number of constraints that can affect the presentation delivery compared to face-to-face environments, such as the limited screen size, distortion of the physics of space (people are displayed far away), and lack of gaze alignment. Provided that the listeners have their video turned on, and are not seated too far away, the most salient aspect of the remote audience is likely to be facial expressions. Although a great deal of research has examined emotional mimicry, to our knowledge, there has been no research examining how speakers react to audience facial expressions when they are delivering a talk via videoconferencing.

The goal of this study is to investigate factors that affect emotional expressiveness and mimicry among speakers in online settings. As in most behavioral studies, the role of dispositional and situational factors looms large. However, while key dispositional information can be universally captured with psychometric instruments such as Big Five [6], situational information is scenario dependent and calls for targeted experimental designs. Here we focus on a parallel group design, which tests the effect on online speakers of knowing vs. not knowing in advance about delivering a talk on a familiar topic. Our design is motivated by breadth of applications and reasonableness. The said scenario often arises in online job interviews, when the interviewees are asked to deliver a short talk on a topic that falls under the purview of the sought-after position. It also arises in educational settings, when the instructors randomly ask students to present on the topic of the day. The speakers are knowledgeable about the topic of the speech, but unsettled by the unexpected request - a situation that is challenging but not hopeless. The intended purpose of this scenario is to test if people can 'think on their feet'. The rather more unreasonable scenario, where the subjects are neither

- *A. Wesley and I. Pavlidis are with the Department of Computer Science, University of Houston, Houston, TX, 77004.E-mail: ipavlidis@uh.edu*
- *P. Tsiamyrtzis is with Milano Politecnico and Gloria Mark is with the University of California, Irvine.*

Manuscript received April 1, 2022.

familiar with the topic nor expecting to give a talk about it, predictably leads to negative outcomes. Such a scenario, although a popular design in certain lab studies (e.g., Trier Social Stress Test [7]), does not occur very often in the real world because it is practically meaningless.

In addition to dispositional and situational factors, stress is also associated with public speaking and thus, it is a variable that deserves consideration. In 2014, the number one fear expressed by Americans was public speaking, or *glossophobia* [8]. Whereas in recent years other topics have supplanted public speaking as the top fear, in 2019, 31.2% of respondents still expressed that they were afraid or very afraid of public speaking [9]. Hence, given the stressful nature of public speaking, it is important to understand how individual differences in stress responses can affect the all important facial communication channel in virtual talks.

Accordingly, we developed a general research methodology for virtual talks and then used it to conduct a case study. In this study, participants first wrote essays and then presented them to three reviewers through a videoconference link. We examined stress, individual personality differences, and unexpectedness as potential influencers of the speakers' emotional expressiveness and mimicry. The main contributions of our research are:

- A novel and validated multimodal method for investigating emotional communication in online talks. The method includes physiological, observational, and questionnaire channels, featuring unobtrusive sensing technology and AI tools. The code we developed for acquiring and analyzing data in accordance with this methodology is publicly available [10]. It will spur further research by facilitating the conduct of virtual talk studies under different experimental scenarios.
- The case study we conducted shows that people who know in advance about their online talk better manage visual communication with the audience, by exhibiting higher mimicry. The opposite is true for people who do not know in advance. This heretofore unknown effect of unexpectedness bears implications for rapport building between online speakers and reviewers in professional and educational settings. In the interest of transparency and reproducibility, the raw and annotated data of the study are publicly available [11].

2 Background: Emotional expressiveness and mimicry

Emotional facial expressions are shown to capture attention via neural mechanisms that involve limbic components in cingulate cortex and amygdala [12]. Exactly because humans attend to emotional (vs. neutral) stimuli faster, and process them more deeply, emotion priming is found to contribute to speaker comprehension [13]. In the arena of promotional videos, audiences appear to prefer certain types of emotional expressions over others, with neutrality being the least popular facial display [14]. Naturally, the optimal amount and type of facial expressiveness depends on the situational circumstances, but generally facial expressiveness leaves more favorable impressions than facial neutrality [3]. In fact, it has been shown that when people attempt to suppress emotional expressiveness during social interactions, this disrupts communication and increases stress levels [15].

Not only the level of emotional expressiveness but also its time synchronization during an interaction play an important role in the quality of communication. Such synchronization, that is, when people match the facial expressions of others during an interaction, is known as mimicry, and when undisturbed by exogenous or endogenous factors, it can take place at the subconscious level [16]. Emotional mimicry has received much attention for over a century, first modeled by Lipp in 1907 [17]. Why do people mimic others? One theory is that people emotionally mimic others to promote a goal of affiliation with others [18]. If emotional mimicry is displayed during videoconferencing we might then expect that it is done to affiliate with the other remote participants.

Behavioral mimicry can occur fast. People can synchronize their movements with interacting partners in as little as 1000 ms [19]. When people were shown pictures of happy faces of others, they changed their facial muscles into smiles within this short time frame. Hess et al. [4] found evidence that the interaction context influences the mimicry behavior, which suggests that the emotional alignment does reflect a person's emotional state. A recent review also suggests that research has converged to accept that when people mimic another's emotional expression, they really do feel the emotions that are consistent with the facial expression they display [20].

Further evidence for mimicry as reflecting an emotional state is found by Bavelas et al. [21] who demonstrated that emotional mimicry is communicative. In other words, at least some of a speaker's facial expressions are influenced by the emotions displayed by the listeners. Emotional contagion occurs when one person's emotional state is adopted by another person who has observed that emotional facial expression [22]. Thus, expressions have a social component and lead to synchrony of behavior. When people interact with each other, they tend to automatically synchronize their facial expressions and movements. Hess and Fisher [23] propose that mimicry acts as a social regulator, indicating the role of context. Mimicry can also serve to show empathy with the speaking partner [24]. In other words, mimicry can serve for the speaker to try to bond with the partner.

While strong mimicry often occurs in face-to-face settings, videoconferencing among remote participants presents a different setting. Video does not create as strong of a sense of social presence of others as face-to-face interaction [25]. While enhanced video conferencing systems such as a 180° display can increase a sense of presence compared to a flat screen [26], typically videoconferences in most work settings use flat screens. If there are multiple participants in the video conference, the images can be small displays, making it hard to observe details of facial expressions. Sometimes video conference presentations deploy shared screens, where the focus is on the displayed content. Even worse, videoconferencing systems enable the speaker and audience to turn off their video feeds altogether. In this study, we focus on those cases where a person is making a presentation to a small set of reviewers without screen sharing and with all the video feeds on - a scenario that often arises in online job interviews and educational contexts. This enables us to investigate emotional mimicry during virtual talks without the confounding factors of shared content and large audiences. Our interest is in how remote speakers take emotional cues from clearly viewed audience members, how speakers might adjust and align their expressions accordingly, and what factors could disrupt this process.

3 Research Questions and Hypotheses

The role of stress on the emotional expressiveness and mimicry of virtual speakers

It is quite common for people to be stressed to one degree or another during public speaking [27]; we expect this to be no different during remote meetings. It has been shown that during bouts of stress, people tend to be more emotionally expressive. In fact, emotional expressiveness has neurophysiological utility because it acts as a stress reliever, leading to reductions in autonomic nervous system activity [28].

It is less clear, however, how stress affects mimicry, as reports in the literature do not always agree. Some research suggests that people with high social anxiety have difficulty encoding emotions of others, and therefore show less emotional mimicry [29]. Some other research found that people with high anxiety mimic emotions selectively, with Vrana and Gross [30] reporting that anxious people tend to mimic negative emotions only, while Dijk et al. reporting that anxious people tend to mimic positive emotions only [31]. In the same school of emotional selectivity, Nitschke et al. reported that people who experience an acute psychological social stressor show lower mimicry for positive emotions but not for negative emotions [32]. In contradistinction to all the previous accounts, Dimberg and Christmanson [19] found that anxious people do not mimic differently seeing positive versus negative emotions on faces. Therefore, to clarify the relationship of stress with emotional expressiveness and emotional mimicry in online speakers, we ask the following questions:

RQ1a: Do people with higher stress exhibit more emotion when delivering a remote talk?

RQ1b: Do people with higher stress show higher mimicry when delivering a remote talk?

The role of disposition on the emotional expressiveness and mimicry of virtual speakers

There is evidence relating both emotional expressiveness and emotional mimicry to personality factors [33], [34]. In a study of mimicry of 18 different behaviors (e.g., head nods, head touching, touching objects), Kurzius and Borkenau [35] found that neurotocism mimicked more negative behaviors; conscientiousness and openness showed lower rates of mimicry of evaluative behaviors; agreeableness showed higher rates of mimicry for positive behaviors; and extraversion showed higher rates of mimicry for negative behaviors. However, these were not mimics of facial expressions but rather mimics of movements. Affiliative mimicking, which corresponds to the Big-Five Agreeableness trait [6], was higher for positive social interactions [36]. A goal of mimicry is to increase affiliation [18]. Individuals who score high in the agreeableness personality trait get along well with others and are more empathic [37]. We therefore expect that scoring high in the personality trait of agreeableness will lead to higher emotional expressiveness and mimicking. We hypothesize the following:

H1a: Remote speakers with higher Agreeableness scores should display more emotions in their facial expressions.

H1b: Remote speakers with higher Agreeableness scores should show higher mimicry, as these speakers should desire to affiliate with the audience.

The role of situation on the emotional expressiveness and mimicry of virtual speakers - unexpectedness

Not only dispositional but also situational factors play a role in public speaking [38]. Situational factors abound and thus, not all of them could be investigated in a single study. Here we focus on unexpectedness, which often arises in practice. For example, in job interviews, the interviewers may ask interviewees to give unannounced short talks about a topic relevant to the sought after position. Interviews styled with such flexible agendas are meant to test if interviewees can ‘think on their feet’. In the realm of education, instructors may ask certain students to give unannounced speeches about the topic of the day. This is typically part and parcel of teaching techniques aiming to maintain attention and maximizing the learning experience [39]. Such rather unexpected presentation requests add to the challenges inherent to speech delivery and may temporarily disorient the speaker. Should this happen, in physical job interviews and classrooms there are follow-up opportunities for the speaker to partially make things up; for example, by further talking to the interviewers during lunch or engaging in a corridor conversation with the instructor after class. In online settings, however, such opportunities are non-existent, which renders synchronous online talks all the more critical.

It is important to note that the situational effect we aim to investigate does not stem so much from lack of preparation or idle anticipation, but largely from unexpectedness or the lack thereof. Job candidates who pass screening and are invited for interview, in all likelihood have competencies; the same applies for the majority of students, who typically prepare before each class. Hence, the said subjects are knowledgeable about the topic of the talk. However, depending on the fixed or flexible agenda of the interview or class, the subjects either know about their talk well in advance (i.e., hours or days) or they are informed in real-time. In the former case, anticipation has largely dissipated [40], while in the latter case there is no time for anticipation to build up.

To our knowledge, no study has looked at the effects of unexpectedness on the emotional display and mimicry of virtual speakers. Although in our situational scenario not previously informed speakers are knowledgeable about the topic, they probably do not have ready a discourse and may experience a higher cognitive load during speech delivery. When executive function resources are being used in decision-making, prioritizing actions, managing attention, using working memory, and in self-regulating [41], then there are less cognitive resources available to devote to other activities. In other words, it is possible that not previously informed speakers will be using more cognitive resources in formulating their presentation, searching for words, remembering the content, and so on. Speakers who were informed well in advance, however, have already spent time formulating their presentation and will need fewer cognitive resources in the act of delivering their talk. There should then be fewer resources used in executive function of making decisions of what to say, retrieving content from memory, and in consciously regulating their actions.

How might the use of cognitive resources affect emotional display is an open question. Regarding mimicry, it was found that cognitive resources used in doing secondary tasks during an interaction resulted in less emotional mimicry [4], [42]. If people are unprepared for a presentation and are using up cognitive resources in formulating their speech delivery, it leaves less resources available to monitor the audience and engage in implicit mimicry [43]. Studies show that when attentional resources were not available, people were not able to process emotional faces and stimuli [44], [45]. If people are prepared for a presentation, then they have more cognitive resources available that can use to closely monitor the audience expressions, which should lead to higher mimicry. Nevertheless, it is not clear if cognitive resources are managed a bit differently in online settings vs. physical settings. We therefore ask the following research questions:

RQ2a: Do knowledgeable speakers exhibit more emotions when they deliver virtual talks upon unexpected requests?

RQ2b: Do knowledgeable speakers show less mimicry when they deliver virtual talks upon unexpected requests?

4 Experimental Design

4.1 Participants

The experimental protocol was approved by the institutional review boards of the University of Houston and the University of California, Irvine, and the Texas A&M University. The protocol was executed in these three universities in accordance with the approved guidelines, obtaining informed consent from each participant before conducting the experiments. In total, $n = 63$ volunteers (45 females/ 18 males), mostly upper-level undergraduate students, participated in the study. The study employed affective computing methods dependent upon a multimodal acquisition system that collected psychometric, observational, and physiological data. The synchronization module of this system malfunctioned in 11 participants. Thus, although data exist for these 11 participants, the various data channels cannot be accurately time-registered, and thus cannot be used in the present investigation. Accordingly, this left $n = 63 - 11 = 52$ participants with fully synchronized data channels. Among these $n = 52$ participants, 11 participants moved their chairs during the experiment in a way that resulted in near total loss of observational data from the monitoring sensors. For two other participants, the online survey system malfunctioned and their psychometric data were corrupted. Hence, the complete and fully synchronized dataset used for analysis included $n = 52 - 11 - 2 = 39$ participants (32 females/ 7 males; age 23.59 ± 8.62).

4.2 Experimental Protocol

The protocol prescribed a controlled experiment on various knowledge work activities for studying the associated behaviors. We reported the dataset of this study (known as Office Tasks 2019) in [46]. Here we focus on the analysis of the last activity in the said experiment, which was an online talk. In more detail, the Office Tasks 2019 protocol opened with a four minute baseline session, where participants were asked to close their eyes and think of something relaxing, such as a nature scene. This session meant to bring participants close to their tonic levels, which could then be used as normalizing anchors for the physiological measurements taken during subsequent treatments. The key treatments of the Office Tasks 2019 protocol included an essay writing task that lasted 50 min, followed by a short break and concluding with a 5 min online talk, where the participants presented the essay they have just written. The essay was on the issue of technological singularity, that is, when machines overtake human intelligence - a topic of broad interest, given the rapid advances in AI. During the essay writing, the participants were free to consult online resources about the topic.

Before commencing the experimental tasks, participants had to complete a demographic and a psychometric questionnaire. The demographic questionnaire collected personal information of importance to this experiment, including participants' native language, education, writing proficiency, and age/gender. Given the nature of the designed tasks, to ameliorate confounding factors, all participants had to have undergraduate education and be native English speakers - an inclusion criterion that we enforced and documented. The psychometric questionnaire was meant to capture personality traits that could affect participants' behaviors in public speaking, with a special interest in agreeableness. For that we chose the Big Five Inventory (B5) [6], which includes agreeableness in its set of sub-scales, scored in the range [9-45].

The experiment featured a parallel group design to facilitate the investigation of situational factors in knowledge work behaviors. Of chief interest were manifestations of valence and mimicry in virtual talks. Among the $n = 39$ participants, $n = 22$ (18 females/ 4 males) had advance notice that they had to present their essay to a panel of judges, while $n = 17$ (14 females/ 3 males) were not informed in advance about the presentation requirement. Specifically, the former group (referred as I) was informed prior to writing their essay, while the latter group (referred as NI) was informed only right after writing their essay, just a few minutes before they needed to deliver their speech. Irrespective of the timing of the notice, all participants delivered an oral presentation without visual aids in front of a three-judge panel, who attended remotely (via Skype). In fact, this was a pre-recorded video with actors that gave the impression of a live session. It is crucial to explain here the design choices with respect to the evaluative audience and the style of the public speaking test.

Evaluative audience: It is impossible to control the composition and behavior of a live audience in order to create a standardized situation for each participant. Accordingly, using a pre-recorded video is a method rooted in the literature for controlling confounding factors in public speaking experiments [47]. In fact, it has been shown that when implemented properly, such a research design allows both realism and stimulus consistency to be maintained across subjects [48]. A key challenge in this approach is to convince the speakers that are presenting in front of a live evaluative audience. In the exit interviews, all participants stated that they believed the audience was live. This success was due to technical excellence and careful thematic choice. The topic of the participants' essay and subsequent talk was about machine intelligence overtaking human intelligence – a sobering prospect. Accordingly, the actors who posed as judges were instructed to maintain a sober attitude. Figure 1 shows that the emotions of the recorded audience largely fluctuate between neutral and sad, that is, an emotional display that fits the topic of the talks. For instance, positive affect is less than 5%, associated with welcoming the speaker the first few seconds. A typical live audience would have exhibited a similar emotional profile. Figure 1 also shows the summary emotional distributions of the speakers. Both the speaker and judge panels in Fig. 1 appear to conform to similar mean patterns: they are dominated by neutral and sad displays, giving a general impression of mood locking, and pointing to the possibility of mimicry. Indeed, prior work suggests that subjects as passive receivers of realistic pre-recorded facial displays, would develop high levels of mimicry in response [21].

Public speaking test: Our public speaking test is closer to the Leiden Public Speaking Task (Leiden-PST) [48] rather than the Trier Social Stress Test (TSST) [7]. In the TSST, participants are given short notice (~ 10 min) to prepare to give a speech on a topic that is usually neither relevant nor engaging. This is a useful laboratory procedure to induce stress to participants, but is not a scenario that often arises in practice. Asking people to deliver talks about things they know little about and for which they did not have time to prepare, is tantamount to setting them up for failure. This is typically not the intent in job interviews, educational processes, and other professional contexts. In contradistinction, in the Leiden-PST the participants know well ahead of time about their upcoming talk and have sufficient time to prepare for it. The I arm of our experiment conforms exactly to the Leiden-PST design. The NI arm of our experiment, which tests the element of unexpectedness, gives short instead of advance notice to the speakers. However, in contrast to the TSST, the NI speakers, although hit with an unexpected request, have already developed competency about the topic of the talk, and thus are not given a hopeless task. In both arms, anticipation stress in blunted by design. In the case of the I group, stress due to the long essay writing task is expected to override anticipatory stress. In support of this expectation, a meta-review of 186 public speech studies found that stress peaked at 38 minutes after subjects were told they would deliver a speech, with stress declining after [40]. In the alternative case of the NI group, the speech is delivered upon notification, leaving little time for anticipatory stress to develop. Hence, no group anticipatory stress effect is expected and stress should be likely driven by individual responses to the writing and speaking tasks per se.

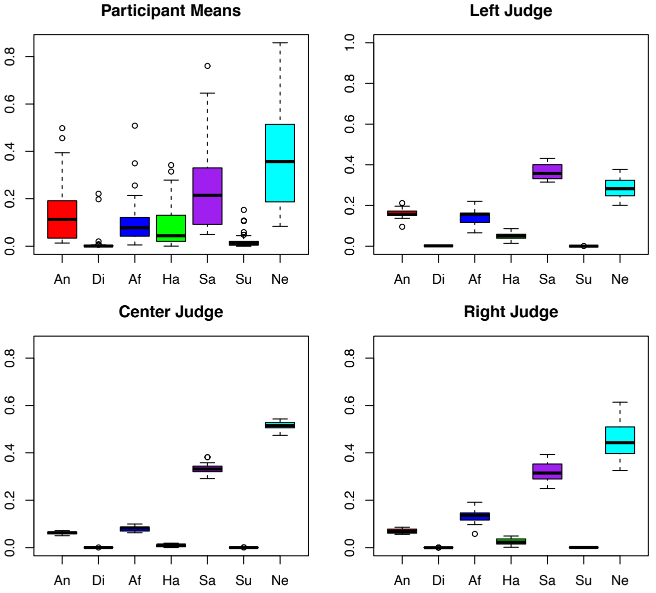


Fig. 1: Distributions of displayed emotions. The first panel shows the distributions of the representative values (means) of all $n = 39$ participants for each emotion. The remaining three panels show the distributions of the momentary values (second by second) for each emotion for the Left, Center, and Right judges, respectively. The judge identifiers denote their location in the Skype window. The abbreviations of the categorical levels in the horizontal axes should be interpreted as follows: An≡Angry; Di≡Disgusted; Af≡Afraid; Ha≡Happy; Sa≡Sad; Su≡Surprised; Ne≡Neutral.

4.3 Experimental setup

The participants used a desktop computer in a typical private office to perform the experimental tasks (Fig. 2). The computer was a Dell OptiPlex 7050 connected to a Dell U2417H-Ultrasharp 24 in display. During the presentation, the face of participants was recorded via a Logitech HD Pro - C920 camera (Logitech, Newark, CA) with spatial resolution 1920×1080, tucked atop their computer screen. The synced data streams from the participant web camera and the judge panel video were fed to a neural network for quantifying displayed emotions (valence) on both sides of the communication channel. This two-channel emotionality was the study's primary response-explanatory variable pair.

For the participant faces, in addition to video recording there was also thermal recording (Fig. 2). Thermal imagery was used to extract a measure of physiological stress (arousal), an important factor in public speaking [49]. This

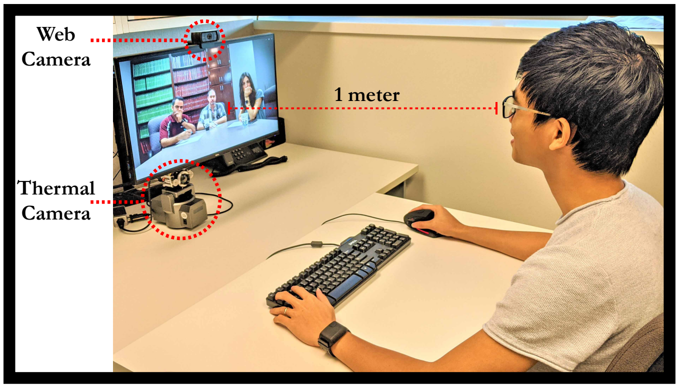


Fig. 2: Experimental setup. Participants were delivering their oral presentation in front of a desktop computer, which was showing in full screen mode the Skype window with the three-judge panel. A web camera atop the computer screen and a thermal camera at the bottom of the computer screen were recording in sync the participants' face.

was carried out with a Tau 640 long-wave infrared (LWIR) camera (FLIR Systems, Wilsonville, OR), featuring a small size ($44 \times 44 \times 30$ mm) and adequate thermal ($50°$ mK) and spatial resolution (640×512 pixels). It was outfitted with a LWIR 35 mm lens f/1.2, controlled by a custom auto-focus mechanism. The Tau 640 camera was located under the participant's computer screen.

Additional physiological variables associated with arousal were recorded through wearable sensors. Specifically, an E4 wristband device (Empatica Inc., Milano Italy) recorded electrodermal activity (EDA) and heart rate on the participants' non-dominant hand, while a Zephyr Bio-Harness 3.0 device (Zephyr Technology, Annapolis, MD) recorded heart rate and breathing rate on the participants' chest. Among the physiological channels employed in the Office Tasks 2019 experiment, the only one that passed all validation tests was the thermal imaging channel (detailed validation report in [46]). For this reason, we chose to use in our analysis the thermal imaging measurements as the most reliable proxies of arousal levels.

Facial thermal imaging is a form of EDA, which is on par with the gold standard [50], that is, palm EDA sensed via a contact probe [51]. Unlike palm EDA, facial thermal EDA is unobtrusive and has no usability issues. For instance, in the context of the Office Tasks 2019 experiment, palm EDA would have been a poor option, as the experimental protocol included writing tasks. Sensing EDA on the wrist solves the usability issue [52], but has accuracy and reliability problems [53], as also the validation report of the the Office Tasks 2019 experiment documented [46]. Hence, this experiment adds to the successful record of thermal facial EDA as a reliable arousal measurement channel in affective computing studies [54], [55], [56].

All video and thermal streams were time registered for precise syncing. To maximize object resolution, the thermal and video cameras aimed at participants were calibrated so that their faces fit nearly the entire frame. High object resolution is especially important in thermal imagery, as it improves the quality of physiological signal extraction. In the case of judges, who were sitting at a conference room table in a remote location, the video frame encompassed all faces. Thus, participants could see at once the facial expressions of all three judges via their Skype window (Fig. 2). This arrangement reduced the resolution per judge face with respect to participant faces, but did not create any problems for the operation of the neural network.

5 Methods

5.1 Visual imaging and extraction of valence

We used a convolutional neural network (CNN) [57] to obtain a probabilistic estimate of participant's S emotional mix, based on her/his facial image at time t. In more detail, we employed a Keras implementation of CNN by Serengil [58], which was trained on the challenging FER dataset [10]. For each participant S, the outcome for the CNN-processed facial frame at time t is a vector $\overrightarrow{V}_{S,t} = \{$Neutral, Surprised, Sad, Happy, Afraid, Disgusted, Angry$\}$. In this vector, each component $v_{S,t,i}$ represents the probability of the corresponding basic emotion being momentarily manifested on the participant's face; thus, $\sum_{i=1}^{7} v_{S,t,i} = 1.0$. We applied the same CNN to obtain momentary emotion vectors $\overrightarrow{V}_{J_{\cdot},t}$ for each judge $J_{\cdot}$, where $\cdot \in \{L, C, R\}$ for left, center, and right judge, respectively. Following the lead from previous research, we set the temporal resolution t at 1 second, as it was found that people can synchronize with interacting partners approximately at this speed [19].

Figure 1 indicates, neutral and sad emotions dominate the facial displays of participants and judges. The remaining minority emotions exist in very small amounts and all but one match sadness in negativity. This configuration motivates binarization of the emotional vectors to simplify analysis without incurring significant loss of information. Accordingly, we binarized the emotion vectors $\overrightarrow{V}_{S,t}$ of participants as follows:

$$\overrightarrow{V}_{S,t} \rightarrow \begin{cases} N, & \text{if } \max\limits_{1 \leq i \leq 7} v_{S,t,i} = \text{Neutral} \\ E, & \text{if } \max\limits_{1 \leq i \leq 7} v_{S,t,i} \neq \text{Neutral} \end{cases} \tag{1}$$

where N indicates a largely neutral facial display, while E indicates a facial display dominated by one of the six basic emotions. Similarly, for the emotion vectors $\overrightarrow{V}_{J_{\cdot},t}$ of the judges:

$$\overrightarrow{V}_{J_{\cdot},t} \rightarrow \begin{cases} N, & \text{if } \max\limits_{1 \leq i \leq 7} v_{J_{\cdot},t,i} = \text{Neutral} \\ E, & \text{if } \max\limits_{1 \leq i \leq 7} v_{J_{\cdot},t,i} \neq \text{Neutral} \end{cases} \tag{2}$$

Finally, we consolidated the judges' responses into a composite response as follows:

$$\overrightarrow{V}_{J,t} \rightarrow \begin{cases} N, & \text{if } V_{J_L,t} \rightarrow N \wedge V_{J_C,t} \rightarrow N \wedge V_{J_R,t} \rightarrow N \\ E, & \text{otherwise} \end{cases} \tag{3}$$

Accordingly, if all judges appeared neutral, then the composite response was labeled neutral N. If at least one of the judges had a non-neutral (that is, emotional) response E, then the composite response was labeled as such. The rationale behind this classification scheme is that participants had in their field of view all three judges (Fig. 2). Hence, emotional expression even by a single member of the judge panel could not easily escape the attention of participants, indelibly coloring their momentary perception.

Validation of CNN: On the one hand, our experimental dataset features high resolution facial imagery in well-lit interior spaces that facilitate emotion recognition by deep learning algorithms. On the other hand, our dataset features talking subjects, which present challenges to emotion recognition networks, such as the CNN we use; the reason is that the said neural networks have been largely trained on non-talking faces. To ensure that the performance of our CNN rises to the task, we ran a validation test on the Ryerson Audio-Visual Database of Emotional Speech and Song database (RAVDESS) [60]. The RAVDESS database contains 24 professional actors (12 female, 12 male), vocalizing two lexically-matched statements in a neutral North American accent (same accent as the participants in our experiment). Speech includes calm, happy, sad, angry, fearful, surprise, and disgust expressions, and song contains calm, happy, sad, angry, and fearful emotions. We only used the speech files, as the song files were not relevant to our experimental data. From the speech files, we excluded the calm files, as these did not have a direct analog in the emotional results reported by the CNN. Please note that each expression is produced at two levels of emotional intensity (normal, strong), with an additional neutral expression. In total, we used 64,708 RAVDESS facial frames. We ran the CNN algorithm on this benchmark dataset and applied formula (1) for binary classification of the results into neutral vs. emotional classes. Using the RAVDESS ground-truth information, we computed an overall accuracy of 77.82%. We are not aware of any neural network performance benchmarks on emotional datasets of speaking subjects. Nevertheless, the near $\sim 80\%$ accuracy of our CNN on the RAVDESS dataset is good enough to stand on its own, rendering validity to the results of the present study. Moreover, this accuracy is at least 10% better than the reported accuracy of state-of-the-art algorithms on the non-speaking FER 2013 dataset [61].

5.2 Thermal imaging and extraction of arousal

The primary location of EDA activity on the face is the perinasal region [62]. Accordingly, we applied algorithmic processing on the thermal imagery to quantify perinasal perspiration. The physiological algorithm was assisted by a virtual tissue tracker that kept track of the region of interest, despite participants' small head motions. This ensured that perspiration signal extraction was taking place on consistent and valid sets of data over the presentation's timeline.

5.2.1 Tissue tracking

We used the tissue tracker reported by Zhou et al. [59]. On the initial frame, the experimenter initiated the tracking algorithm by selecting a broad facial area that included the participant's perinasal region. The tracker estimated the best matching block in every next frame of the thermal clip via spatio-temporal smoothing. A visual example of the tissue tracking operation is shown in the TOP row of Fig. 3.

5.2.2 Quantification of perinasal perspiration

In facial thermal imagery, activated perspiration pores appear as 'cold' (dark) spots, amidst 'hot' surrounding tissue. Accordingly, we applied Shastri's clinically validated morphology algorithm on the measurement region of interest (MROI) to compute the perspiration signal [50]; MROI refers to the upper orbicularis oris portion of the tracked perinasal tissue. The MIDDLE row of Fig. 3 shows the evolving thermal signature of perspiration spots in the MROI of participant S046, as he undergoes moments of low and high arousal. The said algorithm quantifies the manifested spatial frequency pattern, extracting a power signal $\mathbf{PP}(S, m)$, indicative of perspiration activity in the perinasal MROI of participant S, for treatment m (BOTTOM panel of Fig. 3). Any high-frequency noise in this signal is suppressed by a Fast Fourier Transformation (FFT) filter. The units of $\mathbf{PP}(S, m)$ are in $^{\circ}\mathrm{C}^2$, expressing the mean latent heat power released by the transient flare-up of sudomotor responses. The numbers are small due to the physical scale of the phenomenon. To facilitate interpretation of the perinasal perspiration measurements, one needs to take their ratios; then it becomes apparent that changes sometimes are massive (e.g., 2× and 3× difference).

To ameliorate measurement bias due to variation in baseline EDA levels, we normalized the $\mathbf{PP}(S, m = PR)$ signal of participant S during the presentation session PR by subtracting $\overline{\mathbf{PP}}(S, BL)$, that is, the participant's mean non-aroused level captured during her baseline session BL:

$$\Delta\mathbf{PP}(S, PR) = \mathbf{PP}(S, PR) - \overline{\mathbf{PP}}(S, BL). \tag{4}$$

Finally, we computed $\overline{\Delta\mathbf{PP}}(S, PR)$, that is, the mean normalized arousal of each participant S during the presentation session PR; these mean delta values were used as the treatment stress levels in subsequent analytic modeling. Zero value for $\overline{\Delta\mathbf{PP}}(S, PR)$ - statistically speaking - indicates baseline level arousal, suggesting that the treatment produced no stress; values higher than statistical zero suggest that the treatment was stressful.

6 Results

6.1 Checking for biases

The presentation was based on the essay the participants wrote, and all tasks were in English. As all participants were native English speakers and had undergraduate education, they were equally qualified to perform the task at hand. This is exactly how the participants felt; the distribution of their writing proficiency self-assessment score was very narrow and on the high end of the 7 point Likert scale: 6.89 ± 0.73.

To objectively ensure that the quality of the essay was within acceptable standards, we rated the participants' essays using the *e-rater* scoring engine of the Educational Testing Service (ETS) [63]. The most important ETS measure is the Criterion Score, which rates the overall essay quality, with score range [1-6]. The Criterion Scores the essays received largely validated the participants' proficiency claims. Figure 4a shows the distributions of the *e-rater* Criterion Score for the participant essays in group I (3.59 ± 1.05) and NI (3.71 ± 1.57); there were no significant differences between the essay scores of the two groups (t-test, $p = 0.797$). Per ETS guidelines, scores 3.5 and higher indicate that essay writers can provide competent analysis of ideas [64]. Accordingly, the participants wrote good essays and thus, they

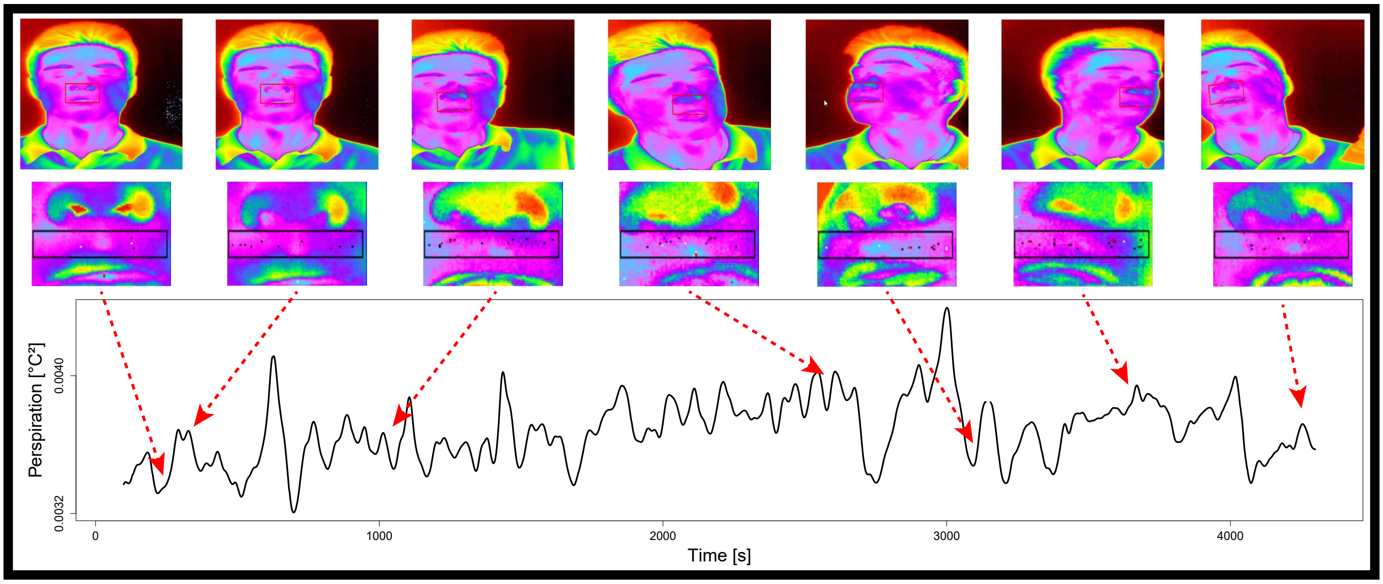
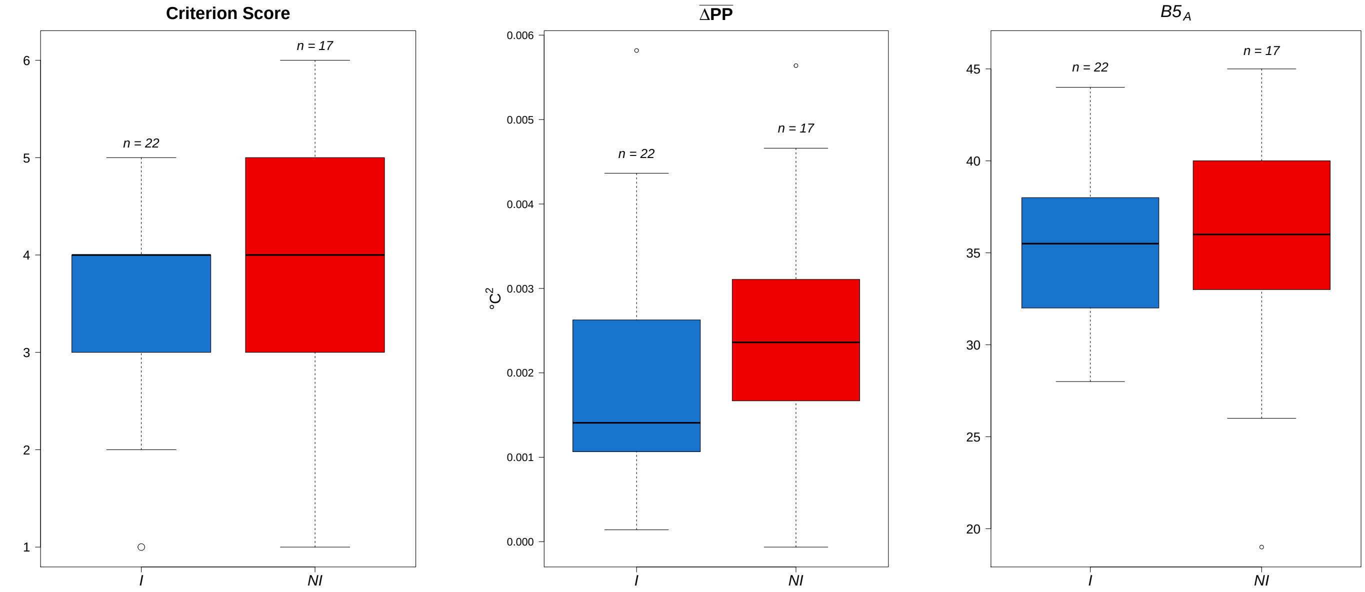


Fig. 3: Motion tracking of the perinasal region of interest or TROI (red rectangle) [59] from where the perspiration signal is extracted during the course of the presentation for subject S046. The thermal facial snapshots are accompanied by the zoomed-in perinasal measurement regions of interest or MROIs, where black dots manifest active perspiration pores detected by the algorithm [50]. This algorithm turns the spatial perspiration pattern into a signal by applying a morphological filter.



(a) (b) (c)

Fig. 4: Descriptive statistics for participant groups I (informed) and NI (not-informed). a. Group distributions of the Criterion Score, an essay quality score assigned by ETS's *e-rater* engine. b. Group distributions of normalized mean stress levels $\overline{\mathbf{\Delta PP}}$ during the presentation, as measured via the perinasal perspiration channel. c. Group distributions of agreeableness scores $B5_A$, as measured by the B5 inventory.

had good background material for their oral presentation, independent of grouping.

In this paper we focus in understanding the relationship between the displayed emotions of the presenting participants and those of the judges, who played the role of critical audience. As described in our Research Questions and Hypotheses (section 3), in addition to the emotional triggers emanated from the judges and the participants' informed/not-informed status, we consider the contribution of participants' stress levels and personality traits in forming their emotionality. Hence, it is important to check if the groups I and NI were equitable with respect to these two covariates. Figure 4b shows the distributions of stress levels for group I ($1.86 \pm 0.14 \times 10^{-3}$ °C^2) and NI

($2.33 \pm 1.60 \times 10^{-3}$ °C^2); there were no significant differences (t-test, $p = 0.348$). Furthermore, the stress distributions suggest that almost all participants had arousal significantly above the baseline level (zero line). Hence, on average the presentation was an equally stressful experience for both groups of participants. Figure 4c shows the distributions of agreeableness for group I (35.5 ± 4.32) and NI (35.24 ± 6.48); the distributions covered a healthy range and there were no significant differences between groups (t-test, $p = 0.886$).

6.2 Mixed effects logistic regression modeling

Having established that the two groups of participants did not differ in terms of background material, stress levels, and relevant traits, we can proceed with behavioral modeling. Equation (5) is the mixed effects logistic regression model that we used in our analysis. The binary response variable P and its transformation $\text{logit}(P) \equiv \ln\left(\frac{P}{1-P}\right)$ measures whether participants had emotional E or neutral N momentary expression. We consider as random effects $1|S$ the participants (Fig. 5a) and β_0 denotes the intercept. The remaining predictors are the fixed effects (Fig. 5b). These include the composite judge display J, which is a binary variable indicating whether the collective facial expression of the three-judge panel was neutral N or emotional E. Additional fixed effects include participants' grouping Gr based on advance knowledge of the event (levels I, NI), their stress level as expressed by the perinasal perspiration measurement $\overline{\mathbf{\Delta PP}}$, and their agreeableness $B5_A$ score from the Big Five Inventory. The model also includes the interactions of the composite judge display with all the other predictors. For the response variable P and the key predictor variable J, the neutral expression level N serves as the reference cell. For the variable Gr, the level NI (not informed) serves as the reference cell. Coefficients for categorical variables are listed as $\beta.$, while for continuous variables are listed as γ.

$$
\begin{aligned}
\text{logit}(P) \sim & \beta_0 + \beta_J J + \beta_{Gr} Gr + \gamma_{\overline{\mathbf{\Delta PP}}} \overline{\mathbf{\Delta PP}} + \gamma_{B5_A} B5_A \\
& + \beta_{J \times Gr} J \times Gr + \gamma_{J \times \overline{\mathbf{\Delta PP}}} J \times \overline{\mathbf{\Delta PP}} \\
& + \gamma_{J \times B5_A} J \times B5_A + 1|S.
\end{aligned} \tag{5}
$$

All four explanatory variables produce significant main effects. From the interactions of the judges' emotionality J with the other three explanatory variables, only the interaction with participants' group status Gr is significant (Table 1). This interaction, which captures the effect of grouping in emotional mimicry, is the most insightful result of model (5). In more detail:

RQ1a: Do people with higher stress exhibit more emotion when delivering a remote talk?

The stress of presenting participants significantly influences their expressions (row $\overline{\mathbf{\Delta PP}}$ in Table 1: $\gamma_{\overline{\mathbf{\Delta PP}}} = 0.787$, $p = 0.034$). Given that the coefficient estimate for $\overline{\mathbf{\Delta PP}}$ is $\gamma_{\overline{\mathbf{\Delta PP}}} = 0.787$, the odds ratio is $e^{0.787} = 2.20$, that is, for every standardized unit increase in stress, the odds that the participants turn emotional are 2.20 to 1 (Fig. 5b). To compute the exact effect of stress $\overline{\mathbf{\Delta PP}}$ on participants S when all other factors are held constant, we use the reduced model $P = \beta_0 + \gamma_{\overline{\mathbf{\Delta PP}}} \overline{\mathbf{\Delta PP}}$. As $\overline{\mathbf{\Delta PP}}$ is a continuous variable, a range of values are applied in this reduced model, yielding the line plot in Fig. 6d, complete with the 95% confidence band. The plot clearly shows that the higher the stress of participants, the higher the chance to exhibit emotionality during their presentation. Thus, participants with higher stress exhibited more emotion during the delivery of the online presentation.

Predictor	$\beta.$ or $\gamma.$	Std. Error	z value	Pr(> $\lvert z \rvert$)
Intercept	1.434	0.537	2.669	0.008 **
$J[E]$	-0.520	0.239	-2.180	0.029 *
$Gr[I]$	-1.726	0.709	-2.436	0.015 *
$\overline{\mathbf{\Delta PP}}$	0.787	0.371	2.123	0.034 *
$B5_A$	1.216	0.363	3.352	<0.001 ***
$J[E] \times Gr[I]$	0.914	0.311	2.936	0.003 **
$J[E] \times \overline{\mathbf{\Delta PP}}$	-0.284	0.197	-1.422	0.149
$J[E] \times B5_A$	-0.300	0.179	-1.671	0.095

TABLE 1: Results of mixed effects logit model (5). Significance levels have been set as follows: *: $p \leq 0.05$, **: $p \leq 0.01$, ***: $p \leq 0.001$.

RQ1b: Do people with higher stress show higher mimicry when delivering a remote talk?

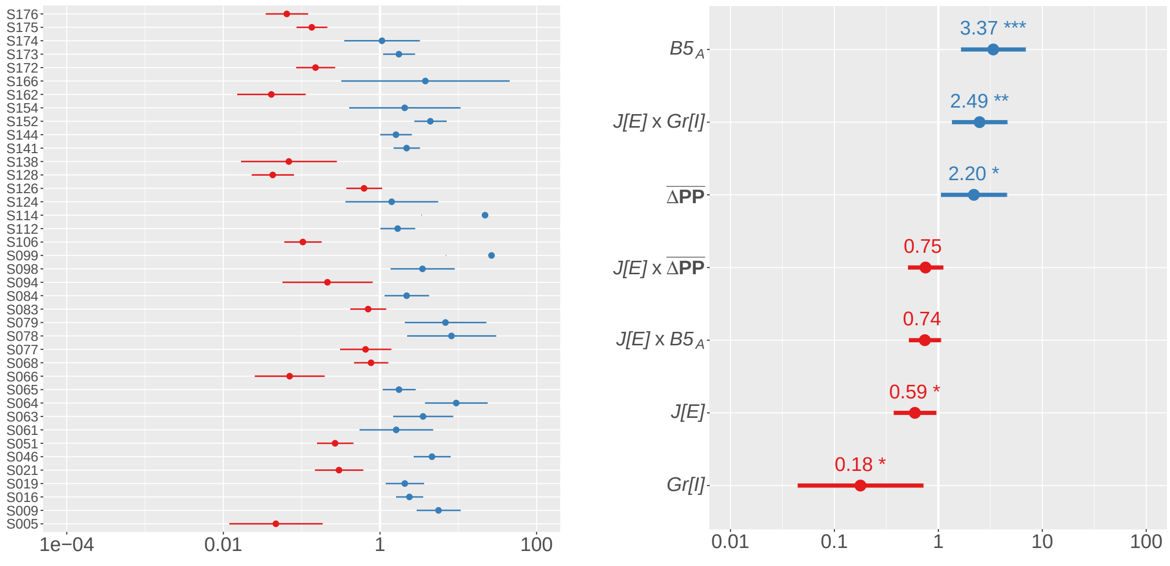
The stress level of presenting participants does not significantly influence their emotional mimicry (row $J[E] \times \overline{\mathbf{\Delta PP}}$ in Table 1: $p = 0.149$).

H1a: Remote speakers with higher Agreeableness scores should display more emotions in their facial expressions

The agreeableness personality score of presenting participants significantly influences their expressions (row $B5_A$ in Table 1: $\gamma_{B5_A} = 1.216$, $p < 0.001$). Given that the coefficient estimate for $B5_A$ is $\gamma_{B5_A} = 1.216$, the odds ratio is $e^{1.216} = 3.37$, that is, for every standardized unit increase in agreeableness, the odds that the participants turn emotional are 3.37 to 1 (Fig. 5b). To compute the exact effect of agreeableness $B5_A$ on participants S when all other factors are held constant, we use the reduced model $P = \beta_0 + \gamma_{B5_A} B5_A$. As $B5_A$ is a continuous variable, a range of values are applied in this reduced model, yielding the line plot in Fig. 6e, complete with the 95% confidence band. The plot clearly shows that the more agreeable the participants are, the higher the chance to exhibit emotionality during their presentation. Thus, we find support for H1a: participants with higher scores in agreeableness show higher emotion when speaking.

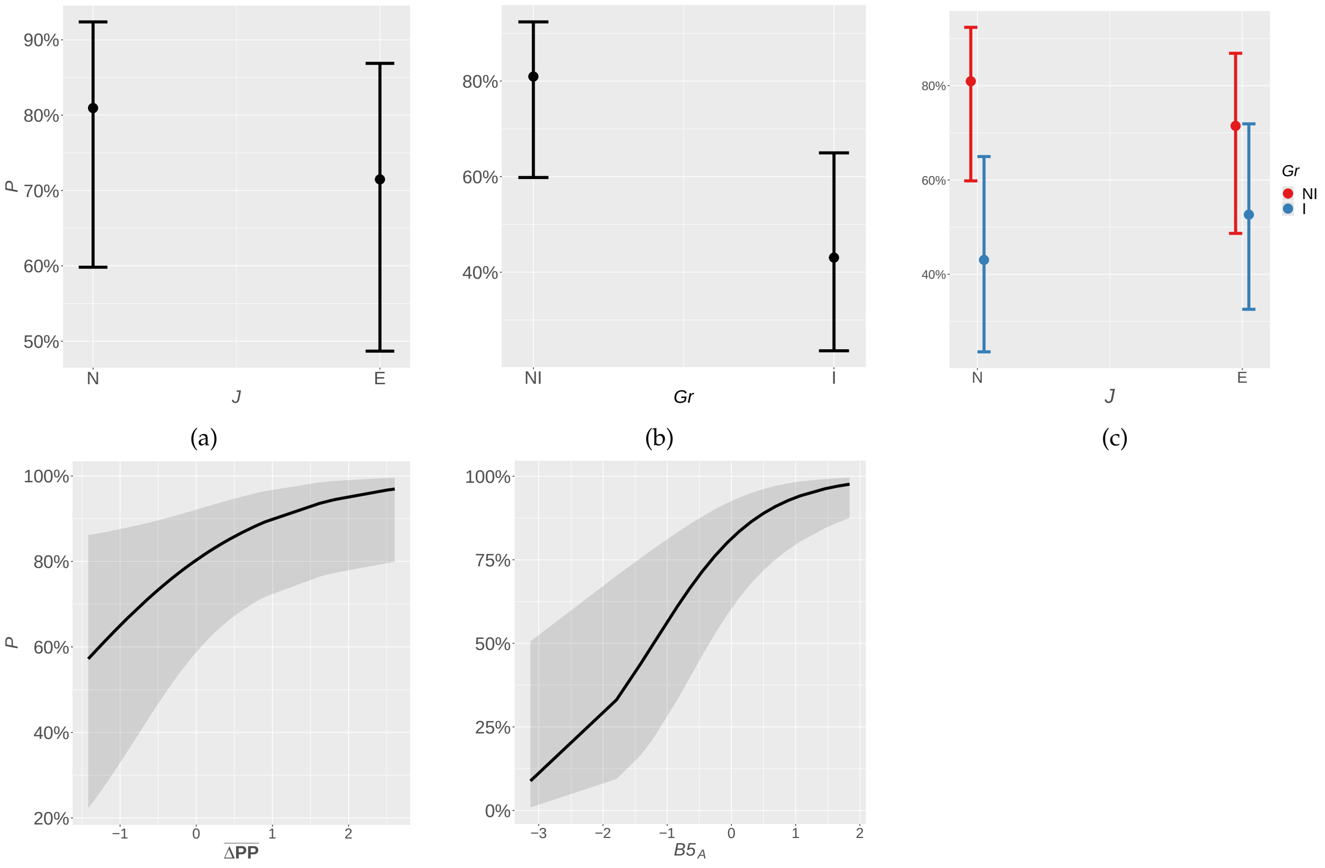
H1b: Remote speakers with higher Agreeableness scores should show higher mimicry

The agreeableness score of presenting participants does not significantly influence their emotional mimicry (row $J[E] \times B5_A$ in Table 1: $p = 0.095$). Based on the p value, however, it is worth noting that this is trending, and conceivably could become significant if the pool of participants expands. Thus, we find no support for H1b.



(a) (b)

Fig. 5: 95% confidence intervals for the coefficients of model (5). a. Intervals for coefficients of random effects (i.e., individual participants) expressed as odds ratios. b. Intervals for coefficients of fixed effects expressed as odds ratios.



(a) (b) (c)

(d) (e)

Fig. 6: Significant main and interaction effects of model (5). a. Main effect of judges' emotional display levels on the emotionality of participants. b. Main effect of group levels on the emotionality of participants. c. Effect of interaction between emotional status of judges and group membership on the emotionality of participants. d. Main effect of presentation stress levels on the emotionality of participants. e. Main effect of agreeableness predisposition on the emotionality of participants.

RQ2a: Do knowledgeable speakers exhibit more emotions when they deliver virtual talks upon unexpected requests?

The informed vs. not informed status of presenting participants significantly influences their expressions (row $Gr[I]$ in Table 1: $\beta_{Gr} = -1.726$, $p = 0.015$). Given that the coefficient estimate for Gr is $\beta_{Gr} = -1.726$, the odds ratio is $e^{-1.726} = 0.18$, that is, for participants who belong to the informed group I, the odds to be emotional are 0.18 to 1 (Fig. 5b). To compute the exact effect of informed vs. not informed status on participants when all other factors are held constant, we use the reduced model $\text{logit}(P) = \beta_0 + \beta_{Gr}Gr$. For the case where participants did not know of the upcoming presentation $Gr[NI] = 0$, this reduced model yields $\text{logit}(P) = \beta_0 = 1.434$; expressed as probability this becomes $P = \frac{e^{1.434}}{e^{1.434}+1} = 0.807$ or 80.7%. Similarly, for the case where participants knew of the upcoming presentation $Gr[I] = 1$, we estimate $\text{logit}(P) = \beta_0 + \beta_{Gr}Gr = 1.434 - 1.726 * 1 = -0.292$; expressed as probability this becomes $P = \frac{e^{-0.292}}{e^{-0.292}+1} = 0.429$ or 42.9%. Hence, when participants know in advance about the presentation, the probability of emotionality during the delivery drops from 80.7% to 42.9%. Figure 6b shows the 95% confidence interval plots for the emotionality of participants, depending on their informed vs. not informed status Gr. The expected probability estimates 80.7% and 42.9% for participants appear as dots in the left ($Gr[NI]$) and right ($Gr[I]$) confidence intervals, respectively. Thus, people who were told in advance they would give a presentation exhibited less emotion in their faces.

RQ2b: Do knowledgeable speakers show less mimicry when they deliver virtual talks upon unexpected requests?

The momentary expressions of judges affect differently the momentary expressions of participants, depending on whether participants were informed in advance if they were presenting or not (row $J[E] \times Gr[I]$ in Table 1: $\beta_{J\times Gr} = 0.914$, $p = 0.003$). Given that the coefficient estimate for $J \times Gr$ is $\beta_{J\times Gr} = 0.914$, the odds ratio is $e^{0.914} = 2.49$, that is, as the judges become emotional, the odds that the participants who belong to the I group become emotional are 2.49 to 1 (Fig. 5b). To compute the exact effect of this interaction on participants when all other factors are held constant, we use the reduced model $P = \beta_0 + \beta_J J + \beta_{Gr}Gr + \beta_{J\times Gr} J \times Gr$. For the case of participants belonging to group $Gr[I] = 1$, when judges are neutral $J[N] = 0$, this reduced model yields $P = \beta_0 + \beta_{Gr}Gr = 1.434 - 1.726 * 1 = -0.292$; expressed as probability this becomes $P = \frac{e^{-0.292}}{e^{-0.292}+1} = 0.428$ or 42.8%. Similarly, for the case where judges are emotional $J[E] = 1$, we estimate for the I group $P = \beta_0 + \beta_J J + \beta_{Gr}Gr + \beta_{J\times Gr} J \times Gr = 1.434 - 0.520 * 1 - 1.726 * 1 + 0.914 * 1 = 0.102$; expressed as probability this becomes $P = \frac{e^{0.102}}{e^{0.102}+1} = 0.526$ or 52.6%.

Hence, when judges become emotional, the probability that participants in the I group, who knew in advance of the upcoming presentation, are emotional, increases from 42.8% to 52.6%. Figure 6c shows the 95% confidence interval plots for the emotionality of participants, depending on their informed vs. not informed status Gr and the emotionality of judges J. The expected probability estimates 42.8% and 52.6% for the I group of participants appear as dots in the blue left and right confidence intervals, respectively. The red confidence intervals show what happens to the emotionality of the NI group of participants, depending on the emotionality of judges. The picture is antithetical to that of the I group. Hence, participants I, who knew in advance, tend to respond to judges' emotional status changes in ways that increases emotional mimicry. Exactly the opposite happens with participants NI who did not know in advance about the upcoming presentation. Specifically, NI group participants tend to respond to judges' emotional status changes in ways that reduces emotional mimicry.

6.3 Overall mimicry of speakers

The momentary expressions of judges significantly influence the momentary expressions of presenting participants (row $J[E]$ in Table 1: $\beta_J = -0.520$, $p = 0.029$). Given that the coefficient estimate for J is $\beta_J = -0.520$, the odds ratio is $e^{-0.520} = 0.59$, that is, as the judges become emotional, the odds that the participants become emotional are 0.59 to 1 (Fig. 5b). In other words, as the judges' facial expressions move from neutral N to emotional E, the facial expressions of participants move in the opposite direction, that is, they trend neutral, suggesting less mimicry. To compute the exact effect of judges J on participants S when all other factors are held constant, we use the reduced model $\text{logit}(P) = \beta_0 + \beta_J J$. For the case where judges are neutral $J[N] = 0$, this reduced model yields $\text{logit}(P) = \beta_0 = 1.434$; expressed as probability this becomes $P = \frac{e^{1.434}}{e^{1.434}+1} = 0.807$ or 80.7%. Similarly, for the case where judges are emotional $J[E] = 1$, we estimate $\text{logit}(P) = \beta_0 + \beta_J J = 1.434 - 0.520 * 1 = 0.914$; expressed as probability this becomes $P = \frac{e^{0.914}}{e^{0.914}+1} = 0.714$ or 71.4%. Hence, when judges become emotional, the probability participants are emotional drops from 80.7% to 71.4% . Figure 6a shows the 95% confidence interval plots for the emotionality of participants, depending on the emotionality of judges J. The expected probability estimates 80.7% and 71.4% for participants appear as dots in the left ($J[N]$) and right ($J[E]$) confidence intervals, respectively. This result identifies a general pattern in the sample without taking into account grouping; thus, in the context of our group design has limited utility, but is an inescapable term in model construction.

6.4 Synopsis of key results

- Participants in groups I and NI showed no differences in amount of stress.
- Participants who scored higher in the agreeableness personality trait displayed more emotion when presenting, but no relationship with emotional mimicry was found.
- Participants I, who received advance notice they would be presenting, showed less emotion throughout their presentation. When they did show emotion, however, it tended to match that of the judges. In other words, group I was more responsive to the

judges' facial expressions, exhibiting higher mimicry. For an example from that group see Fig. 7.

- Participants NI, who did not receive advance notice they would be presenting, showed more emotion throughout their presentation. This plethora of displayed emotions, however, was distributed in a way that was less responsive to the judges' facial expressions, exhibiting less mimicry. For an example from that group see Fig. 8.

7 Discussion

As online and hybrid operations are here to stay, there is an urgent need to better understand this relatively new work environment and the implications it bears on human actors. In this context, virtual talks are an important class of actions that are worth investigating because of their utility in career advancement and education. Extensive evidence in the literature suggests that speakers' emotional expressiveness and mimicry play an important role in their effectiveness [3], [13], [65]. This must also be true for online speakers, although the said issues are less well researched. In fact, speakers' face becomes more central in online talks, as it is the only visible part of their body. Moreover, virtual talks carry higher stakes, as they are the only impression speakers leave, due to the lack of follow-up social interactions afforded after the end of conventional talks. This work contributes a state-of-the-art method to quantify the physiological, psychometric, and observational features that contribute to the emotional expressiveness and mimicry exhibited by online speakers. The method uses a clinically validated thermal imaging algorithm to quantify electrodermal activity on the face - a proxy of stress. It uses the Big-Five questionnaire to quantify personality traits and Agreeableness in particular. And, it uses a validated CNN algorithm to quantify valence for the speaker as well as the audience, thus estimating the level of mimicry between the two parties. Importantly, the method uses a logistic regression model that predicts the speaker's valence and mimicry from the speaker's stress and personality, as well as the audience's emotional displays. Affective researchers can readily use our integrated set of scripts that realize this methodology (Tsiamyrtzis, P., Wesley, A., and Pavlidis, I. Online-Mimicry-Methods. *GiHub* https://github.com/UH-CPL/Online-Mimicry-Methods) to collect and analyze data from studies on online talks, operationalizing various situational scenarios of interest.

To demonstrate the utility of our method and investigate a scenario that often arises in practice, we ran a parallel group study largely modeled after the Leiden-PST design. Two groups were given ample time to write an essay about the issue of technological singularity, consulting online sources if they wish, and thus educate themselves on the matter. One of the groups was made aware that after the essay they will also need to deliver an online talk to an evaluative committee. The other group was made aware of the speech requirement, after they finished the essay and only a couple of minutes before they needed to deliver the talk. The audience was a pre-recorded Skype video, but timed and played in such a way that all speakers thought it was real.

Online speakers who are either more stressed during their talk or more agreeable in nature tend to be more emotionally expressive. This result is largely in agreement with prior findings on conventional talks. The most interesting and novel result of our example study, however, is the antithesis between the virtual speaker groups I and NI with respect to their emotional expressiveness and mimicry. Group I participants who were forewarned about giving a talk after the essay, appear less emotionally expressive but attain higher mimicry during their speech. In contrast, group NI participants who were not forewarned about giving a talk after the essay, appear more emotionally expressive but attain lower mimicry during their speech. To explain the underlying mechanism of this phenomenon we need to take into account the following: First, in the exit interviews nearly all NI participants stated that they were taken aback by the talk request, and although knowledgeable about the topic, it took them some time to find their bearings. Second, because of the topic, the tone was somber and emotional displays in both the speakers and the evaluative audience were largely fluctuating between neutral and sad. Neutral in this and other circumstances is a resting place for the facial muscles [66], and the question is how long and how often the facial muscles go back to this baseline position; the longer and the more often, the less emotional the appearance of the subject.

With this context in mind, virtual speakers in group I are not looking for words as much, and thus are more assertive and less emotional. Importantly, they can spare more of their heightened resources to properly manage the visual communication channel. In this context, when I presenters meet emotional displays from the audience, they are more ready to constructively engage through emotional mimicry. Virtual speakers in group NI likely engage in frantic efforts to 'find their words', a precarious and potentially embarrassing condition that generates excessive emotional valence. Interestingly, this emotional valence is not uniform; it adjusts toward neutrality when it is met with emotional displays from the pre-recorded audience. We theorize that this is a freeze response. Participants of the NI group are dazed from the unexpected request and are trying to regain their footing by articulating a speech on the fly. In this vulnerable position, they tend to perceive emotional expressions from the audience as social threats. There is support in the literature that perceived social threats may activate the freeze response [67], which in this case translates to assuming a neutral face. Therefore, this systematic shift toward neutrality, in an otherwise highly emotional performance, can be explained in a bio evolutionary framework.

7.1 Limitations

Our sample had an uneven gender balance with more females than males. To our knowledge, it is not clear if there are gender differences in emotional mimicry and further research is needed to test this. Furthermore, unlike neutrality, which is only one type, there is more than one type of emotionality. In fact, there are six basic emotions in the probability vectors $\overrightarrow{V}_{S,t}$ produced by the CNN (section 5.1). As can be seen in Fig. 1, disgust and surprise are almost non-existent in the data, which practically leaves four emo-

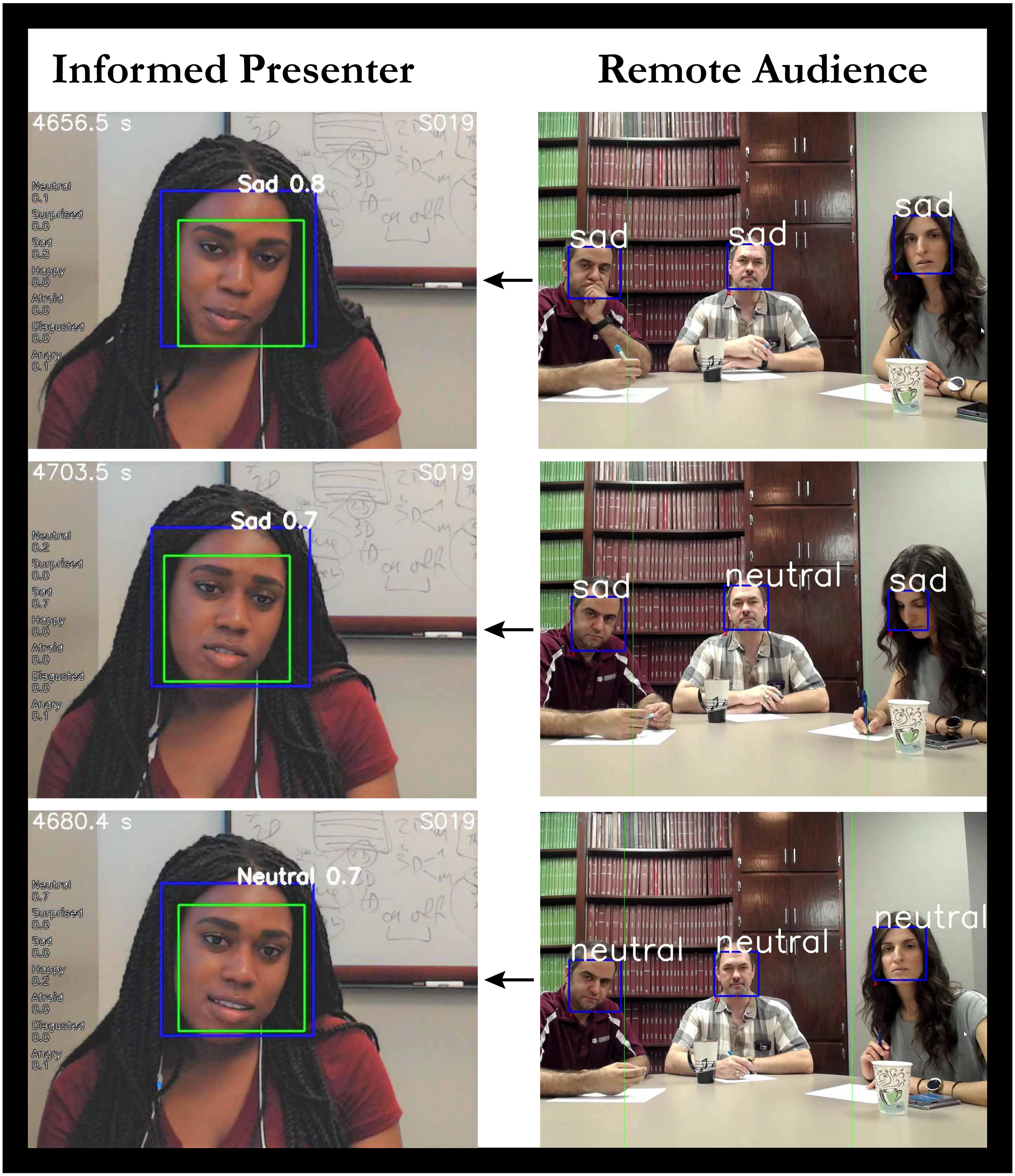


Fig. 7: Examples of emotional mimicry from participant S019, belonging to the I group. The left column features snapshots of the participant, while the right column shows the corresponding snapshots of the remote judge panel.

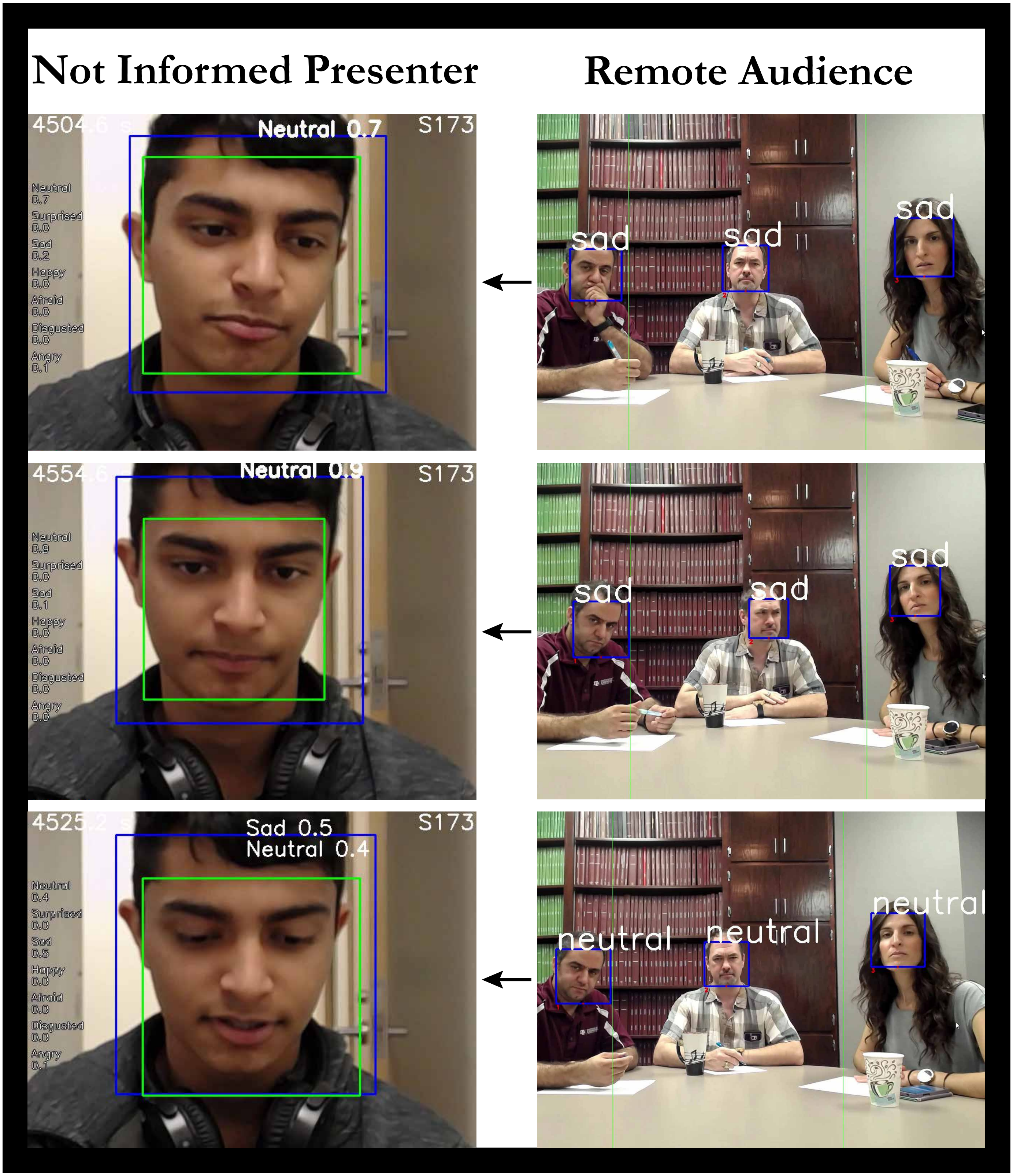


Fig. 8: Examples of emotional incongruity from participant S173, belonging to the NI group. The left column features snapshots of the participant, while the right column shows the corresponding snapshots of the remote judge panel.

tions for consideration. Because we binarized the emotional vectors, however, we cannot see the nuances of the mimicry. Due to the overwhelming frequency of the sad expressions (Fig. 1), we have to assume that in most cases sadness is met with sadness. However, almost certainly there are other less frequent combinations (e.g., anger is met with fear), which would be worth investigating in future studies. Such future studies should feature multi-dimensional topics, which can naturally accommodate a range of emotions during speech delivery. Additionally, the said studies should be powered with more subjects, so that rarer emotional combinations have enough data mass to be properly analyzed. In this case, it would be relatively easy to transform the binomial logistic regression model we use in the analysis to a multinomial logistic regression model.

Although quantification of facial electrodermal activity (EDA) via thermal imaging is an excellent method to unobtrusively quantify stress in experiments and field studies [68], thermal imaging sensors are still not widely available. HCI researchers who use our methodology to study emotionality in virtual talks, can substitute the thermal imaging channel with a heart function channel, realized via smartwatches. Literature reports suggest that heart function proxies of stress, although not as accurate as facial EDA, offer an acceptable alternative [68].

Acknowledgments

The authors would like to thank...

References

[1] D. Autor and E. Reynolds, "The nature of work after the COVID crisis: Too few low-wage jobs," The Hamilton Project, Essay 14, 2020.

[2] C. Heath and P. Luff, "Disembodied conduct: Communication through video in a multi-media office environment," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1991, pp. 99–103.

[3] R. E. Riggio and H. S. Friedman, "Impression formation: The role of expressive behavior," *Journal of Personality and Social Psychology*, vol. 50, no. 2, p. 421, 1986.

[4] U. Hess, P. Philippot, and S. Blairy, "Facial reactions to emotional facial expressions: Affect or cognition?" *Cognition & Emotion*, vol. 12, no. 4, pp. 509–531, 1998.

[5] J. L. Lakin, V. E. Jefferis, C. M. Cheng, and T. L. Chartrand, "The chameleon effect as social glue: Evidence for the evolutionary significance of nonconscious mimicry," *Journal of Nonverbal Behavior*, vol. 27, no. 3, pp. 145–162, 2003.

[6] R. R. McCrae and P. T. Costa Jr, "A five-factor theory of personality," in *Handbook of Personality: Theory and Research*, O. P. John, R. W. Robins, and L. A. Pervin, Eds. New York, NY: The Guilford Press, 2008, ch. 5, pp. 159–181.

[7] C. Kirschbaum, K.-M. Pirke, and D. H. Hellhammer, "The 'Trier Social Stress Test'–a tool for investigating psychobiological stress responses in a laboratory setting," *Neuropsychobiology*, vol. 28, no. 1-2, pp. 76–81, 1993.

[8] C. Ingraham, "America's top fears: Public speaking, heights and bugs," https://www.washingtonpost.com/news/wonk/wp/2014/10/30/clowns-are-twice-as-scary-to-democrats-as-they-are-to-republicans/, 2014, retrieved Mar. 16, 2022.

[9] S. Sheth, "America's Top Fears 2019," https://www.chapman.edu/wilkinson/research-centers/babbie-center/_files/americas-top-fears-2019.pdf, 2019, retrieved Mar. 16, 2022.

[10] P. Tsimayrtzis, A. Wesley, and I. Pavlidis, "Online-Mimicry-Methods," 2022. [Online]. Available: https://github.com/UH-CPL/Online-Mimicry-Methods

[11] A. Wesley and I. Pavlidis, "Office Tasks 2019 - Presenter/Audience Emotions," 2022. [Online]. Available: https://osf.io/ytb26/

[12] P. Vuilleumier and S. Schwartz, "Emotional facial expressions capture attention," *Neurology*, vol. 56, no. 2, pp. 153–158, 2001.

[13] M. N. Carminati and P. Knoeferle, "Effects of speaker emotional facial expression and listener age on incremental sentence processing," *PloS One*, vol. 8, no. 9, p. e72559, 2013.

[14] P. Lewinski, "Don't look blank, happy, or sad: Patterns of facial expressions of speakers in banks' YouTube videos predict video's popularity over time," *Journal of Neuroscience, Psychology, and Economics*, vol. 8, no. 4, p. 241, 2015.

[15] E. A. Butler, B. Egloff, F. H. Wlhelm, N. C. Smith, E. A. Erickson, and J. J. Gross, "The social consequences of expressive suppression," *Emotion*, vol. 3, no. 1, p. 48, 2003.

[16] U. Dimberg, M. Thunberg, and K. Elmehed, "Unconscious facial reactions to emotional facial expressions," *Psychological Science*, vol. 11, no. 1, pp. 86–89, 2000.

[17] S. Blairy, P. Herrera, and U. Hess, "Mimicry and the judgment of emotional facial expressions," *Journal of Nonverbal Behavior*, vol. 23, no. 1, pp. 5–41, 1999.

[18] U. Hess and A. Fischer, "Emotional mimicry: Why and when we mimic emotions," *Social and Personality Psychology Compass*, vol. 8, no. 2, pp. 45–57, 2014.

[19] U. Dimberg and L. Christmanson, "Facial reactions to facial expressions in subjects high and low in public speaking fear," *Scandinavian Journal of Psychology*, vol. 32, no. 3, pp. 246–253, 1991.

[20] E. Hatfield, L. Bensman, P. D. Thornton, and R. L. Rapson, "New perspectives on emotional contagion: A review of classic and recent research on facial mimicry and contagion," *Interpersona*, vol. 8, no. 2, p. 159, 2014.

[21] J. B. Bavelas, A. Black, C. R. Lemery, and J. Mullett, ""I show how you feel": Motor mimicry as a communicative act." *Journal of Personality and Social Psychology*, vol. 50, no. 2, p. 322, 1986.

[22] E. Hatfield, J. T. Cacioppo, and R. L. Rapson, "Emotional contagion," *Current Directions in Psychological Science*, vol. 2, no. 3, pp. 96–100, 1993.

[23] U. Hess and A. Fischer, "Emotional mimicry as social regulation," *Personality and Social Psychology Review*, vol. 17, no. 2, pp. 142–157, 2013.

[24] F. B. de Waal and S. D. Preston, "Mammalian empathy: Behavioural manifestations and neural basis," *Nature Reviews Neuroscience*, vol. 18, no. 8, pp. 498–509, 2017.

[25] J. Borup, R. E. West, and C. R. Graham, "Improving online social presence through asynchronous video," *The Internet and Higher Education*, vol. 15, no. 3, pp. 195–203, 2012.

[26] T. Aitamurto, S. Zhou, S. Sakshuwong, J. Saldivar, Y. Sadeghi, and A. Tran, "Sense of presence, attitude change, perspective-taking and usability in first-person split-sphere 360 video," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 2018, pp. 1–12.

[27] T. Chanwimalueang, L. Aufegger, W. von Rosenberg, and D. P. Mandic, "Modelling stress in public speaking: evolution of stress levels during conference presentations," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2016, pp. 814–818.

[28] D. S. Berry and J. W. Pennebaker, "Nonverbal and verbal emotional expression and health," *Psychotherapy and Psychosomatics*, vol. 59, no. 1, pp. 11–19, 1993.

[29] K. Button, G. Lewis, I. Penton-Voak, and M. Munafò, "Social anxiety is associated with general but not specific biases in emotion recognition," *Psychiatry Research*, vol. 210, no. 1, pp. 199–207, 2013.

[30] S. R. Vrana and D. Gross, "Reactions to facial expressions: effects of social context and speech anxiety on responses to neutral, anger, and joy expressions," *Biological Psychology*, vol. 66, no. 1, pp. 63–78, 2004.

[31] C. Dijk, A. H. Fischer, N. Morina, C. Van Eeuwijk, and G. A. Van Kleef, "Effects of social anxiety on emotional mimicry and contagion: Feeling negative, but smiling politely," *Journal of Nonverbal Behavior*, vol. 42, no. 1, pp. 81–99, 2018.

[32] J. P. Nitschke, C. S. Sunahara, E. W. Carr, P. Winkielman, J. C. Pruessner, and J. A. Bartz, "Stressed connections: Cortisol levels following acute psychosocial stress disrupt affiliative mimicry in humans," *Proceedings of the Royal Society B*, vol. 287, no. 1927, p. 20192941, 2020.

[33] D. Keltner, "Facial expressions of emotion and personality," in *Handbook of emotion, adult development, and aging*, C. Magai and S. H. McFadden, Eds. Elsevier, 1996, pp. 385–401.

[34] G. Perugia, M. Paetzel, and G. Castellano, "On the role of personality and empathy in human-human, human-agent, and human-

robot mimicry," in *International Conference on Social Robotics*. Springer, 2020, pp. 120–131.

[35] E. Kurzius and P. Borkenau, "Antecedents and consequences of mimicry: A naturalistic interaction approach," *European Journal of Personality*, vol. 29, no. 2, pp. 107–124, 2015.

[36] H. Mauersberger, C. Blaison, K. Kafetsios, C.-L. Kessler, and U. Hess, "Individual differences in emotional mimicry: Underlying traits and social consequences," *European Journal of Personality*, vol. 29, no. 5, pp. 512–529, 2015.

[37] W. G. Graziano, L. A. Jensen-Campbell, and E. C. Hair, "Perceiving interpersonal conflict and reacting to it: the case for agreeableness," *Journal of Personality and Social Psychology*, vol. 70, no. 4, p. 820, 1996.

[38] M. J. Beatty and M. H. Friedland, "Public speaking state anxiety as a function of selected situational and predispositional variables," *Communication Education*, vol. 39, no. 2, pp. 142–147, 1990.

[39] C. Greive and K. de Berg, "The use of surprise and sequential questioning as a teaching technique," in *Proceedings of the Fifth International Conference on Science, Mathematics and Technology Education*, 2008, pp. 152–161.

[40] W. K. Goodman, J. Janson, and J. M. Wolf, "Meta-analytical assessment of the effects of protocol variations on cortisol responses to the Trier Social Stress Test," *Psychoneuroendocrinology*, vol. 80, pp. 26–35, 2017.

[41] M. T. Banich, "Executive function: The search for an integrated account," *Current Directions in Psychological Science*, vol. 18, no. 2, pp. 89–94, 2009.

[42] P. R. Cannon, A. E. Hayes, and S. P. Tipper, "An electromyographic investigation of the impact of task relevance on facial mimicry," *Cognition and Emotion*, vol. 23, no. 5, pp. 918–929, 2009.

[43] A. N. Dalton, T. L. Chartrand, and E. J. Finkel, "The schema-driven chameleon: How mimicry affects executive and self-regulatory resources." *Journal of Personality and Social Psychology*, vol. 98, no. 4, p. 605, 2010.

[44] L. Pessoa, M. McKenna, E. Gutierrez, and L. G. Ungerleider, "Neural processing of emotional faces requires attention," *Proceedings of the National Academy of Sciences*, vol. 99, no. 17, pp. 11 458–11 463, 2002.

[45] G. Hajcak, A. MacNamara, and D. M. Olvet, "Event-related potentials, emotion, and emotion regulation: An integrative review," *Developmental Neuropsychology*, vol. 35, no. 2, pp. 129–155, 2010.

[46] S. Zaman, A. Wesley, D. R. D. C. Silva, P. Buddharaju, F. Akbar, G. Gao, G. Mark, R. Gutierrez-Osuna, and I. Pavlidis, "Stress and productivity patterns of interrupted, synergistic, and antagonistic office activities," *Scientific Data*, vol. 6, no. 1, pp. 1–18, 2019.

[47] S. Perowne and W. Mansell, "Social anxiety, self-focused attention, and the discrimination of negative, neutral and positive audience members by their non-verbal behaviours," *Behavioural and Cognitive Psychotherapy*, vol. 30, no. 1, pp. 11–23, 2002.

[48] P. M. Westenberg, C. L. Bokhorst, A. C. Miers, S. R. Sumter, V. L. Kallen, J. van Pelt, and A. W. Blöte, "A prepared speech in front of a pre-recorded audience: Subjective, physiological, and neuroendocrine responses to the Leiden Public Speaking Task," *Biological Psychology*, vol. 82, no. 2, pp. 116–124, 2009.

[49] B. M. Kudielka, D. H. Hellhammer, and C. Kirschbaum, "Ten years of research with the Trier Social Stress Test—revisited," in *Social Neuroscience: Integrating Biological and Psychological Explanations of Social Behavior*, E. Harmon-Jones and P. Winkielman, Eds. New York, NY: The Guilford Press, 2007, ch. 4, pp. 56–83.

[50] D. Shastri, M. Papadakis, P. Tsiamyrtzis, B. Bass, and I. Pavlidis, "Perinasal imaging of physiological stress and its affective potential," *IEEE Transactions on Affective Computing*, vol. 3, no. 3, pp. 366–378, 2012.

[51] W. Boucsein, *Electrodermal Activity*. Springer Science & Business Media, 2012.

[52] M. Garbarino, M. Lai, D. Bender, R. W. Picard, and S. Tognetti, "Empatica E3—A wearable wireless multi-sensor device for real-time computerized biofeedback and data acquisition," in *2014 4th International Conference on Wireless Mobile Communication and Healthcare-Transforming Healthcare Through Innovations in Mobile and Wireless Technologies (MOBIHEALTH)*. IEEE, 2014, pp. 39–42.

[53] P. Tsiamyrtzis, M. Dcosta, D. Shastri, E. Prasad, and I. Pavlidis, "Delineating the operational envelope of mobile and conventional EDA sensing on key body locations," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 2016, pp. 5665–5674.

[54] I. Pavlidis, P. Tsiamyrtzis, D. Shastri, A. Wesley, Y. Zhou, P. Lindner, P. Buddharaju, R. Joseph, A. Mandapati, B. Dunkin *et al.*, "Fast by nature-how stress patterns define human experience and performance in dexterous tasks," *Scientific Reports*, vol. 2, p. 305, 2012.

[55] I. Pavlidis, M. Dcosta, S. Taamneh, M. Manser, T. Ferris, R. Wunderlich, E. Akleman, and P. Tsiamyrtzis, "Dissecting driver behaviors under cognitive, emotional, sensorimotor, and mixed stressors," *Scientific Reports*, vol. 6, p. 25651, 2016.

[56] F. Akbar, A. E. Bayraktaroglu, P. Buddharaju, D. R. Da Cunha Silva, G. Gao, T. Grover, R. Gutierrez-Osuna, N. C. Jones, G. Mark, I. Pavlidis *et al.*, "Email makes you sweat: Examining email interruptions and stress using thermal imaging," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 2019, pp. 1–14.

[57] A. T. Lopes, E. de Aguiar, A. F. De Souza, and T. Oliveira-Santos, "Facial expression recognition with convolutional neural networks: Coping with few data and the training sample order," *Pattern Recognition*, vol. 61, pp. 610–628, 2017.

[58] P.-L. Carrier and A. Courville, "Facial Expression Recognition (FER) Dataset," https://www.kaggle.com/c/challenges-in-representation-learning-facial-expression-recognition-challenge/data, 2013.

[59] Y. Zhou, P. Tsiamyrtzis, P. Lindner, I. Timofeyev, and I. Pavlidis, "Spatiotemporal smoothing as a basis for facial tissue tracking in thermal imaging," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 5, pp. 1280–1289, 2013.

[60] S. R. Livingstone and F. A. Russo, "The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English," *PloS ONE*, vol. 13, no. 5, p. e0196391, 2018.

[61] S. Minaee and A. Abdolrashidi, "Deep-emotion: Facial expression recognition using attentional convolutional network," *arXiv e-prints*, p. arXiv:1902.01019, 02 2019.

[62] D. Shastri, A. Merla, P. Tsiamyrtzis, and I. Pavlidis, "Imaging facial signs of neurophysiological responses," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 2, pp. 477–484, 2009.

[63] J. Burstein, J. Tetreault, and N. Madnani, "The E-rater® automated essay scoring system," in *Handbook of Automated Essay Evaluation*, M. D. Shermis and J. Burstein, Eds. New York, NY: Routledge, 2013, ch. 4, pp. 77–89.

[64] ETS, "Score Level Descriptions for the Analytical Writing Measure," https://www.ets.org/gre/revised_general/prepare/analytical_writing/score_level_descriptions/, 2020.

[65] K. A. Duffy and T. L. Chartrand, "Mimicry: causes and consequences," *Current Opinion in Behavioral Sciences*, vol. 3, pp. 112–116, 2015.

[66] D. Terzopoulos and K. Waters, "Analysis and synthesis of facial image sequences using physical and anatomical models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 6, pp. 569–579, 1993.

[67] K. Roelofs, M. A. Hagenaars, and J. Stins, "Facing freeze: Social threat induces bodily freeze in humans," *Psychological Science*, vol. 21, no. 11, pp. 1575–1581, 2010.

[68] S. Taamneh, P. Tsiamyrtzis, M. Dcosta, P. Buddharaju, A. Khatri, M. Manser, T. Ferris, R. Wunderlich, and I. Pavlidis, "A multimodal dataset for various forms of distracted driving," *Scientific Data*, vol. 4, no. 1, pp. 1–21, 2017.

PLACE
PHOTO
HERE

Michael Shell Biography text here.

John Doe Biography text here.

Jane Doe Biography text here.