Effects of Sharing Real-time Stress Indicators During Video-Mediated Multilingual Collaboration

Hajin Lim, Cornell University, USA
Naomi Yamashita, NTT Communication Science Laboratories, Japan
Susan R. Fussell, Cornell University, USA

Sharing[1] biosignals of communication partners has the potential to improve communication by providing additional cues to partners' internal states. In this study, we investigated the effects of sharing real-time stress indicators in computer-mediated multilingual communication between native (NS) and non-native English speakers (NNS). We conducted a lab experiment in which 24 NS-NNS dyads collaborated on a decision-making task. Half of the dyads used a standard video conferencing system; the other half used video plus a real-time indicator of their partners' stress levels. Overall, sharing the stress indicators did not lead to a better understanding of partners' stress levels, but NS tended to accommodate NNS partners by slowing down their speech. Follow-up interviews revealed that the real-time stress indicator was interpreted differently by NS and NNS. Our findings inform the design of future tools for multilingual communication.

CCS Concepts: • **Human-centered computing** → Computer supported cooperative work

KEYWORDS

Biosignal, Biosensor, EDA, Stress indicator, Multilingual team, Multilingual communication

ACM Reference format:

1 INTRODUCTION

Multilingual teams are increasingly common in global organizations [23], and a large number of these teams are required to use a common language such as English [9][21]. However, the use of a common language creates challenges to teamwork, due to differences between native (NS) and non-native (NNS) speakers in terms of fluency and comfort levels when speaking the common language. Previous work has investigated the challenges of non-native speakers using common languages in academic [49][77] and workplace settings [21][60][63][71]. NNS often participate less actively in group discussion than NS, in part due to the additional cognitive load of processing and producing a non-native language [69] and their concerns about a lack of perfect fluency [60]. Also, when NS partners speak rapidly, NNS can lose track of the conversation and/or find it difficult to find a conversational opening to share their views [75].

A number of tools have been introduced to resolve these challenges faced by multilingual teams. For example, Yamashita and her colleagues developed voice/video chat interfaces that aid NNS comprehension of the common language by providing automatic speech transcripts [11][20], bilingual dictionaries [27], and highlights [56]. Other systems have provided language-based feedback for NS-NNS pairs to highlight underlying differences in conversation style based on written text chat logs and emails [34][35]. In general, these tools have focused on supporting better processing of the conversation content for non-native speakers, including both linguistic and visible aspects of communication.

Fewer tools have focused on providing multilingual teams with insight into partners' psychological statuses. Although most NS recognize that NNS have difficulties in communication [1][68], NS have trouble understanding the magnitude of these challenges [35][77]. While some nonverbal and vocal cues can be useful for identifying other's psychological states like anxiety and discomfort [31], they are not always a reliable source in multilingual teams due to cultural differences in nonverbal expression [6][62][66].

[1] It is a datatype.

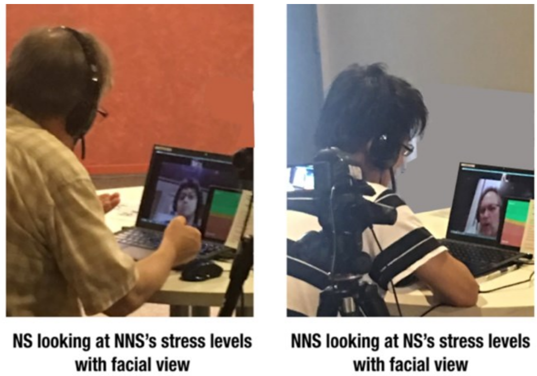


Figure 1 NS-NNS Dyad in Video-Mediated Communication with Stress Indicator

Previous work shows that sharing biosignals such as heart rate or stress level can increase emotional awareness and feelings of connectedness in computer-mediated communication (CMC), not only for close relationships [19][33] but also for task-oriented interactions [70]. Therefore, we expect that providing supplementary information about a partner's psychological status (i.e., stress level) via biosignals, can alleviate communication challenges in multilingual teams.

In this paper, we examine how sharing biosignals between NS-NNS pairs influences group collaboration processes and outcomes. We focused on investigating the effects of sharing real-time stress indicators that detect whether a person is calm or distressed using electrodermal activity (EDA). EDA is the electrical activity of the sweat glands in the skin, which can be used to measure arousal and stress levels [5][33]. In a lab setting, 24 NS-NNS dyads collaborated on a group decision making task. Half of the pairs communicated using a standard video conferencing tool, and the other half communicated using the same video conferencing tool with an added stress indicator that visualized their partner's stress levels in real-time. We measured participants' self-reported and estimated stress ratings along with their psychophysiological signals. Also, we collected quantitative measures on accommodation behaviors and perceptions, privacy concerns, cognitive load, and satisfaction with the conversation and partner. In addition, we conducted interview with participants regarding how they perceived and utilized stress indicators during the discussion.

Quantitative data analyses revealed that participants were not very accurate in their predictions of their partners' stress levels, and the presence of the sensor display did not improve their accuracy. In fact, the sensor had negative effects on NNS ability to estimate their partners' stress. Instead, participants' estimates of their partners' stress levels were strongly predicted by their native language and their own self-reported stress.

Although our results did not demonstrate the immediate benefits of stress indicators, we identified a number of ways that the stress indicators shaped the ways that NS and NNS attended to and responded to each other. We conclude with a discussion of our findings and design implications for new CMC tools that use real-time stress indicators to improve multilingual communication.

In the remainder of this paper, we first discuss related work on challenges in multilingual communication and the sharing of physiological data in interpersonal contexts. Next, we describe the design of our laboratory study and present our findings. We conclude by discussing the implications of our findings for the design of future CMC tools that incorporate real-time stress indicators for better multilingual communication.

2 BACKGROUND

In this section, we first review the literature on challenges in multilingual communication via common languages. Next, we examine previous work on the effects of sharing biosignals on social interaction. Based on this review, we present our research questions and hypotheses.

2.1 Multilingual Communication using Common Language

Previous work has emphasized the importance of communication accommodation by NS when speaking with NNS using common language (e.g., [1][60][68]). Communication accommodation is the process by which speakers adjust their communicative behavior to meet the needs of their partners [29]. Appropriate levels of accommodation can improve communication efficiency and social outcomes (e.g., [26][28][78]). In the context of multilingual communication, research on communication accommodation usually focuses on the adjustments of NS to make their messages more accessible and comprehensible for NNS, considering the degree of language proficiency of their partner [1][14][26]. Therefore, slowing down the pace, simplifying speech, avoiding difficult words, and checking the partner's comprehension frequently [39] are useful strategies for accommodating NNS.

However, NS are not always aware of NNS needs for accommodation. The challenges that NNS face in intercultural communication are often invisible due to impression management. For example, NNS's cognitive burden could evoke social discomfort and embarrassment, so they may pretend to understand [4]. Especially among those with Asian cultural backgrounds, NNS report feeling uncomfortable interrupting the flow of conversation to ask questions [59]. Thus, NS have few obvious cues to the challenges that NNS are undergoing during the conversation [35][77].

Furthermore, NS's willingness and ability to accommodate NNS during a conversation are often insufficient [68]. The lack of appropriate accommodation can make NNS more anxious and lead them to withdraw and avoid communication with their NS partners [1]. NS often misattribute the communication patterns of NNS partners (i.e., less participation) to their dispositional traits (e.g., "he is shy," "she is not knowledgeable") or intentions (e.g., "she does not want to be involved in this communication") rather than to language challenges [35].

In general, prior tool development has focused on supporting better processing of the conversation content– the linguistic and visible aspects of communication. For example, researchers have explored a variety of techniques for increasing NNS's English comprehension, including automatic speech transcripts [11][20], bilingual dictionaries [27], or highlights provided by NS [56]. Diamant et al. [16] presented relative amounts of task-related and emotion-related words in NS-NNS chat conversations.

A few systems have used group displays to raise awareness of partners' experiences. For instance, He et al. [34] used a 3D camera to provide information about non-verbal behaviors (e.g., smiling, eye-contact) to help multilingual groups reflect on individual and group communication patterns. While He et al.'s technique shows that there is promise in providing partners in multilingual groups with nonverbal feedback, their display was presented only during breaks in the conversation rather than continuously. As a result, it could not provide moment by moment understanding of partners' internal states, such as cognitive load, emotion, or stress level.

In the current work, we explore whether providing partners with better awareness of each other's psychological states, particularly stress levels, can improve multilingual CMC. We focus on biosensors because they are highly sensitive to moment by moment shifts in psychological states, and because inexpensive devices are widely available. As such, they can be easily incorporated into video conferencing set ups.

2.2 Effects of Sharing Biosignals in Social Interactions

A growing number of mobile and wearable devices are now equipped with biosensors that allow users to obtain their physiological data, including heart rate, skin conductance, or blood pressure. Such physiological data is often utilized to detect psychological states such as emotion, stress levels, or cognitive load. While the relationship between physiological signals and experienced psychological states can vary based on demographics, personality, and situational factors [3], physiological data can provide one useful cue to what others are experiencing internally [10][55]. Various classification algorithms have been introduced that can process raw biosignal data to infer and classify users' psychological states (e.g., [17][48]). Most applications monitoring psychological states via biosensors have been designed for individual well-being, for example, to help people learn how to control their own stress and emotion [24][57].

Some recent studies have explored the effects of sharing physiological and/or psychological data during social interaction. The promises of sharing biosignals were based on the assumption that sharing biosignals could work as additional social cues to communicate internal (invisible) states. Earlier studies revealed that sharing biosignals could yield positive social outcomes such as an increased feeling of connectedness [41][52], empathy [33][47], and social support [37], especially in close relationships.

Further, several researchers have investigated whether sharing biosignals can also improve task-oriented CMC. For example, Tan et al. [70] discovered that a visual stress indicator as a part of a video chat interface improves situational awareness and communication grounding in remote learning environments. Goyal et al. found that the use of biosignals benefits the collaborative sensemaking of geographically dispersed teams [30]. In addition, Kim et al. [42] revealed that biofeedback on group dynamics reduces overlapping speaking time and created more equal participation in the conversation.

Overall, previous work suggests that sharing biosignal can provide useful supplementary cues for CMC, leading to positive communication processes and outcomes. Yet, it is unclear whether and how sharing biosignals can influence multilingual communication because earlier studies paid little attention to communicators' (potentially diverse) linguistic backgrounds.

2.3 Research Questions and Hypothesis

Previous empirical studies of multilingual communication using common language suggest that 1) NS sometimes cannot estimate the stress of NNS [35][77], and 2) vocal and non-verbal cues to partners' psychological states are sometimes inaccurate due to cultural differences [6][66]. Based on previous work showing that biosignals can help partners become better aware of each other's psychological states (e.g., [33][70]), we expect that subjective representation of partners' psychological states via biosignals will be a useful cue in multilingual communication.

To investigate our expectations, we focused on one psychological state: the level of stress. (Psychological) stress is defined as mental or emotional tension in demanding situations [12]. We focus on stress because stress levels are easy for people to interpret [36], so there is less chance of ambiguity and variance in interpreting the signals [8] unlike raw sensor data such as heart rate [53] or electroencephalogram (EEG) output [46]. Furthermore, we assumed that physiological stress measures can effectively capture the feelings of being overloaded and cognitively taxed that NNS often report during conversations with NS [38][69] To measure real-time stress levels, we used a commercial EDA sensor and stress-tracking applications. EDA is known for one of the reliable measures to monitor psychological stress as it shows strong correlations with the stressor stimuli such as competitive racing games [45] or challenging cognitive tasks [65].

First, we predicted that providing an indicator of real-time stress levels will increase team members' awareness of each other's physiological stress states because it can provide visible cues of partner's internal states that are otherwise difficult to detect. Therefore, we postulated that the estimation of the partner's stress levels would be more accurate with stress indicators than without stress indicators:

H1 [Estimation Accuracy] The accuracy of assessing partner's stress levels will be greater in pairs with stress indicators than pairs without stress indicators.

However, sharing private biosignals with another person might raise concerns about privacy [2], especially when communication is between teammates instead of close family or friends [67]. Therefore, we hypothesized that:

H2 [Privacy Concerns] Pairs using video conferencing with stress indicators will express greater privacy concerns than pairs without stress indicators.

Although stress indicators may increase privacy concerns, they could also provide NS with salient cues to their NNS partners' physiological states and serve as a reminder of the need to accommodate these partners. Consequently, we expected NS participants with stress indicators to show more accommodative behaviors toward their NNS partners. Specifically, we expected NS with the stress indicator to adapt their speech behaviors so that NNS can more easily understand what they are saying, by slowing down their speaking rate, or using shorter utterances [40][78]. Or, NS can do more backchanneling for NNS using more assent words (e.g., "okay', "yeah, " "uh-huh"), which could facilitate the better fluency of NNS [74] and encourage NNS participation [40][50]. Based on these rationales, we hypothesized:

H3a [NS speech speed] NS in pairs with stress indicators will speak more slowly than NS in pairs without stress indicators.

H3b [NS utterance length] The average utterance length of NS in pairs with stress indicators will be shorter than NS in pairs without stress indicators.

H3c [NS backchanneling] NS in pairs with stress indicators will use more assent words than NS in pairs without stress indicators.

While stress indicator displays may increase NS accommodation to NNS conversational needs, NNS might feel that NS adjustments are appropriate [28][29]. Or, NNS might not notice these accommodations at all [61]. Therefore, we ask:

RQ1 [NNS Accommodation Perception] How will the presence of a stress indicator during video conferencing affect NNS perceptions of NS accommodation?

In addition, we consider how the presence of a stress indicator will affect cognitive load during group discussion. It is possible that it could increase cognitive load because participants are required to monitor the stress indicators in addition to engaging with their partners. However, it is also possible that it could help people grasp each other's stress levels more easily rather than monitoring (sometimes not very obvious) nonverbal cues [31], so it could lower overall cognitive load. Thus, we pose the following research question:

RQ2 [Cognitive Load] How will the presence of a stress indicator during video conferencing affect the cognitive load of NS and NNS participants?

Finally, we explored whether having a stress indicator would influence people's satisfaction with the collaboration and their evaluations of their partners. The majority of earlier findings suggest that sharing biosignals can enhance feelings of connectedness [67] and mutual intimacy [41]. However, some other studies did not find links between sharing biosignal and positive assessment of the communication partners [46][52][53]. We therefore ask:

RQ3 [Social Outcomes] What is the effect of sharing biosignals on satisfaction with the collaboration and evaluation of partners?

3 METHOD

We conducted a laboratory experiment in which pairs consisting of one NS and one NNS performed a decision-making task via a video conferencing tool. We manipulated the existence of a stress indicator that provided a real-time visualization of partners' stress levels. Half of the pairs performed the task using video conferencing alone; the other half performed the task using video conferencing with stress indicators (see Figure 1). We used both quantitative and qualitative analyses to investigate the effects of the real-time stress indicator on participants' understanding of their partners' psychological states, NS accommodation behaviors, privacy, workload, and self-reported satisfaction with the team and partner.

3.1 Participants

A total of 48 individuals participated in the study. Participants were recruited through a temporary staffing agency. Half (N=24, seven females; mean age 46.0 [SD = 10.7]) were native English speakers who lived in Japan but grew up and received education in English speaking countries. The other half (N=24, 15 females; mean age 34.2 [SD = 14.6]) were native Japanese speakers who grew up and received education in Japan. They spoke English as a second language. NNS participants' self-reported English proficiency level was intermediate (M= 3.04, SD= 0.71 on a scale of 1: Novice – 5: Native Speaker). About half of them (13/24) had briefly experienced living in an English-speaking country. Both NS and NNS participants were compensated with $50 in Japanese currency.

Participants were randomly assigned to 24 pairs consisting of one NS and one NNS. Then the pairs were randomly assigned to one of the conditions (video conferencing tool only vs. video conferencing tool with stress indicator). There were 12 pairs in each condition.

Initial screening showed that there were no significant differences between NS-NNS pairs with or without the stress indicator in terms of their age (NS/NNS respectively), gender (NS/NNS respectively), years living in Japan (NS), years living in English-speaking countries (NNS) and self-reported English proficiency (NNS) ($p > .05$ for all variables).

3.2 Study Materials

3.2.1 Task

Participants collaborated on a modified version of the Wilderness survival task, a popular team building activity [79]. They were given a scenario in which their car crashed on a cliff in the wilderness, and they had six items with them (e.g., city map, wool blanket, a roll of duct tape, and three chocolate bars). The goal of the task was to rank the items in order of importance for their survival. After reading the scenario, they first ranked the items individually from 1 (most important for survival) to 6 (least important for survival). Then, they collaborated to come up with a group ranking.

3.2.2 Survey

Before the discussion, participants completed an online pre-experiment survey that collected their demographic information (e.g., gender, age, nationality, native language). After a discussion with their partners to decide on the group ranking, participants filled out a post-experiment questionnaire. First, they answered a manipulation check to make sure if participants perceived the treatment of experiment (non-existence vs. existence of stress indicator) correctly. Next, participants rated their overall cognitive load, privacy concerns, overall satisfaction with the collaboration, and liking of their partner. NNS participants also rated how much they felt their NS partners accommodated them. Details are provided in the Measures section.

Figure 2 Moment-by-Moment Stress Ratings

3.2.3 Moment by Moment Stress Rating

After completing the survey, participants watched ten video clips from the recordings of their group discussion. For the recorded video, we added ten random time markers over the timeline of the video. The recordings with time markers were divided and exported to mp4 video clips that were 15 – 20 seconds long, depending on when the end of an utterance occurred.

For both control and stress indicator conditions, participants watched only the clips (no stress indicator was shown) to complete moment-by-moment ratings. While watching these snippets, we asked participants to recall their stress perceptions during that time. They then marked their own (perceived) stress level and their partner's (estimated) stress level during that moment on a continuous scale from relaxed to stressed (See Figure 2). We used a single item to measure stress based on Elo et al's [22] demonstration of the utility and validity of a single-item measure of stress perception. When they are unsure of their own or partner's stress perception, they could skip it. We asked them to explain the reasons for each rating verbally (i.e., "Why did you mark your stress level so stressed?"; "What made you think your partner's stress level as relaxed?").

3.3 Software and Equipment

3.3.1 Video Conferencing Tool

Participants used 14-inch Windows laptops with 1.7 GHz CPU and 4GB memory for videoconferencing and completing questionnaires. They used Zoom (https://zoom.us/) for communication, which allowed them to talk to each other while seeing each other's faces. Participants wore headsets with microphones to communicate with their partners and receive instructions from the experimenters.

3.3.2 Real-time Stress Tracking and Visualization Tools

For tracking real-time stress levels, we used a commercial EDA sensor, the PIP (https://thepip.com), that detects electrodermal activity from fingers. EDA is the electrical activity of the sweat glands in the skin, which can be used to measure stress/arousal and classify a person as calm or distressed [5][33]. The EDA sensor and mobile app used in the experiment have been utilized for stress measurement and visualization in previous studies (e.g., [17][48]). EDA signals obtained from the sensor were transmitted via Bluetooth to a 'PIP Stress Tracker' app (see Figure 3) that processed the EDA signals simultaneously to classify one's stress levels as one of the states (Stressed vs. Steady vs. Relaxed). Then, the stress classification labels were visualized as constantly changing ripples with text labels ("Stressed", "Relaxed") throughout the session.

Pairs in both conditions were asked to put their fingers on metal discs on PIP during the discussion. However, only the pairs in stress-indicator condition were able to see the real-time stress visualizations of their partners via an 8-inch Android Tablet, which was located right next

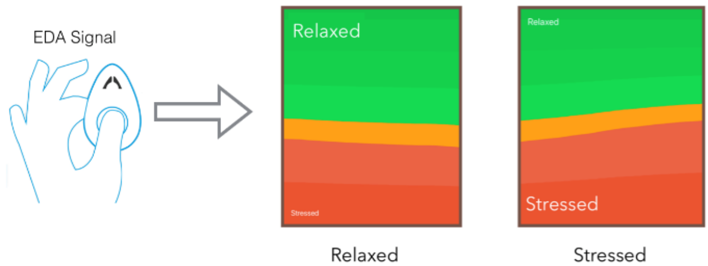


Figure 3 User Interfaces for the Stress Tracker Application and Visualization of Relaxed and Stressed States

to the laptop screen (see Figure 4). For the pairs in the control condition, they could only see the partner's face via the video conferencing tool, so we explained the reason for EDA measuring to them was just to measure their stress levels for the purpose of data collection.

3.4 Procedure

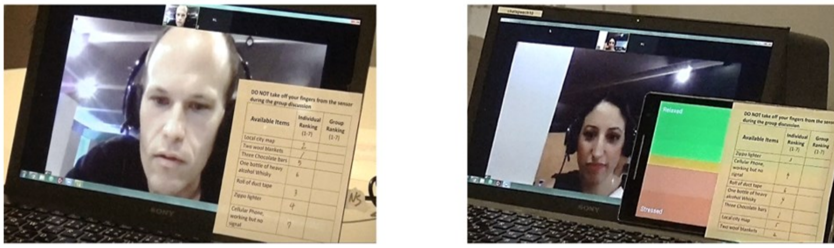
Partners were seated in separate rooms. They first read and signed consent forms. Then, the experimenters explained the study procedure. After participants completed the pre-experiment survey, they were asked to read the task scenario carefully and individually rank the six items. We prepared the small piece of the paper that participants could write down their individual rankings. This paper sheet was attached either to the laptop or tablet screen that they could easily recall the items and individual rankings during the discussion (See Figure 4). After both NS and NNS participants finished their individual rankings, the experimenters introduced the PIP device and explained how it worked. Then, we calibrated the PIP device by asking participants to put their fingers on the sensor for 2-3 minutes. For the participants in the stress-indicator condition, we demonstrated the 'PIP stress tracker' interface and explained how the visualization reflected their partner's real-time stress levels. Also, we asked them to pay attention to the stress indicator display as much as possible. Then, we placed the tablet at the side of the laptop screen (See Figure 4) to ensure the consistent attention to the stress indicator during the discussion without covering up the facial view.

After calibrating, participants put on the headsets to start video conferencing. The experimenter let them know that the discussion duration was fixed at 15 minutes and asked them not to take their fingers off the sensor during the discussion. Lastly, we told participants there would be extra compensation if their group rankings were identical to expert rankings, in order to motivate active discussion. After this, the participants engaged in their 15-minute discussion. The experimenter notified them of the remaining time at five and two minutes before the discussion ended.

After the group discussion, we asked participants to take off the headset and remove their fingers from the sensor. Then, they filled out the post-experiment survey. While they were completing it, we added the ten random time markers over the recording of the group discussion and exported them as ten short video snippets for the moment-by-moment stress rating session.

After participants completed the post-experiment questionnaire, the researchers showed the ten video clips and asked participants to rate their own and (perceived) partner's stress levels at that point on a continuous scale from relaxed to stressed (Figure 2). Participants were able to repeat each clip as many times as they wanted.

Finally, we conducted semi-structured interviews in the participants' native languages, which took 10 -20 minutes. We asked all participants about their overall experience working with their partners. Then, we inquired whether and how they had accommodated their partner.



Setup for Video Only Condition **Setup for Video with Stress Indicator**

Figure 4 Configuration of the Displays.
Left: Video Only Condition with Facial View + Item Ranking Sheet.
Right: Video with Stress Indicator Condition with Facial View + Stress Indicator + Item Ranking Sheet.

For participants in stress indicator condition, we further asked how they attended to, interpreted, and made use of the indicator during the discussion. Overall, each session lasted about one and a half hours.

3.5 Quantitative Measures

3.5.1 Manipulation Check

To ensure if participants perceived the treatment of experiment (non-existence vs. existence of stress indicator) correctly, we presented a multiple-select question. This question asked them to indicate all the displays they saw during the discussion (facial view, stress indicator). The manipulation check result showed that all participants correctly perceived the presence or absence of the stress indicator, so we do not consider this measure further.

3.5.2 Self-reported Stress Ratings and Estimated Stress Ratings of Partner's

We converted participants' markings on their own and their partner's (estimated) stress ratings obtained from moment-by-moment stress ratings (see Figure 2) into numeric values on a scale of 1 (relaxed) - 10 (stressed) based on the absolute positions of markings on the continuous scale. For each participant, there were total 20 stress ratings: 10 for self-reported stress and another 10 for estimated partner stress.

3.5.3 Accuracy of Estimating Partner's Stress Levels

To measure how (in)accurately participants were able to estimate their partner's stress levels, we investigated how closely the self-report stress ratings and the estimation made by their partners were correlated depending on the stress indicator condition (with vs. without) and participant's language background (NS vs. NNS). Stronger relationships reflect better estimation ability.

3.5.4 Privacy Concerns

To measure privacy concerns, we adapted the Affective Benefits and Costs of Communication Technologies (ABCCT) questionnaire [76]. Participants responded to two statements ("I was worried about my privacy while my partner and I were using this technology," "I was worried that my partner might have learned something that I wanted to keep secret") on a 5-point Likert scale (1 = Disagree to 5 = Agree). The two items formed a reliable scale (Cronbach's α = .69) and were averaged to measure 'privacy concern'.

3.5.5 NS's Accommodative Behaviors

To measure how accommodative NS participants were toward their NNS partner during the discussion, we calculated the average speech rate, average utterance length, and percentage of assent words in uses of NS participants.

NS Speech Rate We used a speech recognition tool to obtain the total number of words and total speech duration for NS participants. The speech rate was calculated by dividing the total number of words by the total duration of speaking.

NS Utterance Length Based on the transcripts, we calculated the average utterance length of utterances of NS participants by dividing the total number of words by the number of speaking turns that NS participants took.

Use of Assent Words We used the Text Analysis and Word Count (TAWC) tool [44], a Unix adaptation of LIWC [58] to count the number of assent words (e.g., “okay,” “yeah”). Then, we computed percentages of assent words by dividing the counts of assent words by the total number of words produced by NS participants.

3.5.6 NNS’s Perceived Accommodation of NS

NNS’s perceived accommodation of NS partners was measured by two statements (“My partner was trying to help me to feel comfortable during the conversation” and “My partner was trying to help me get more actively involved in the conversation”) on a 5-point Likert scale (1 = Disagree to 5 = Agree). The items formed a reliable scale (Cronbach’s α = .80) and were averaged to measure ‘Perceived accommodation of NNS’.

3.5.7 Cognitive Load

Cognitive load during the discussion was measured using 5-point Likert scales adapted from the NASA Task Load Index [32] (1 = Low to 5 = High). Four items (Mental load, Time pressure, Effort, Frustration) formed a reliable scale (Cronbach’s α = .68) and were averaged to create a measure of ‘cognitive load’.

3.5.8 Collaboration Satisfaction

To measure the overall satisfaction of the collaboration, participants responded to eight statements on a 5-point Likert scale (1 = Disagree to 5 = Agree) that were subjected to factor analysis with Varimax rotation. The results indicated the presence of two factors, one corresponding to how well the pair communicated to each other (e.g., “I felt our conversation went very well,” “It was easy to keep track of my partner’s ideas”), and the other corresponding to how well the pair performed the task (e.g., “I am satisfied with the group ranking,” “We built upon each other’s opinions”). These factors accounted for 38% and 15% of the variance, respectively. By averaging their responses to questions loading on each factor, we created the measure of ‘communication satisfaction’ (Cronbach’s α = .74) and ‘performance satisfaction’ (Cronbach’s α = .73).

3.5.9 Evaluation of Discussion Partner

To measure the social perception about the conversational partner, participants responded to ten statements on a 5-point Likert scale (1 = Disagree to 5 = Agree) based on interpersonal attraction scales [51]. Factor analysis with Varimax rotation revealed only one factor, so we averaged the responses and created our measure of ‘social attraction toward partner’ (Cronbach’s α = .93).

3.6 Analysis of Interview Data

We fully transcribed interviews; all the interviews with NS were transcribed by native English speakers, and the interviews with NNS were transcribed by native Japanese speakers. All the Japanese transcripts were translated into English by professional translators.

Based on an inductive analysis approach [72], we open-coded all the interview transcripts line by line to identify interesting quotes and assign tentative codes [54]. After all, we generated 57 codes at this stage. Next, we iteratively developed the major themes by sorting and synthesizing the frequently occurring codes. As a result, we identified three high level categories (i.e., asymmetric attention, different conception/interpretation of stress signals, additional function of stress indicators).

4 RESULTS

In this section, we first assess the accuracy of the real-time stress indicators by exploring how the magnitude of stress identified via EDA sensors and self-report stress ratings are related to each other. Next, we examine our hypotheses and research questions.

4.1 Accuracy of Stress Indicators

We first explored how accurately the stress indicator was able to detect participants' stress levels throughout the session. We utilized the "session report" (see Figure 5) that the PIP Stress Tracker app automatically generated for each completed session. This report included a PIP score ranging from 0% to 100% which represented how relaxed (low stress) the participant was during the session (Figure 5-A). Lower PIP scores indicate higher stress. This score is calculated based on the proportions of Stressed, Steady and Relaxed events captured throughout the session (Figure5-B). In addition, the report provided an Event Timeline (5-C) showing how the participant's stress levels changed throughout the session. However, the report did not include the raw EDA data. The average PIP score of all participants was 56.5 (SD = 4.97) and the average percentage of "Stressed" status was 29.1% (SD = 8.2%), which meant about 30% of the time, pairs with stress indicator saw their partner was stressed via visualization on the displays.

We found a moderate negative correlation between participants' PIP scores and the mean of their 10 self-report stress ratings (r= -.321, p = .03). There was also a positive correlation between the percentage of "Stressed" labels and averaged self-report stress ratings (r = .32, p = .03). However, there was no significant correlation between moment by moment self-reported stress levels and moment by moment stress readings from the sensor (r= -.03, p = .48).

Based on these results, we concluded that the stress indicator we used in our experiment provided reasonably accurate information about participants' stress levels across the overall session, but it did not provide reliable information about momentary stress levels. Thus, in interpreting and discussing the findings below, we address the fact that the stress indicator was not able to provide reliable momentary stress levels. Also, we limit our analysis of how the "presence" of the stress indicator influenced NS-NNS collaboration rather than examining how the stress indicator's real-time representation of stress levels affected our quantitative measures.

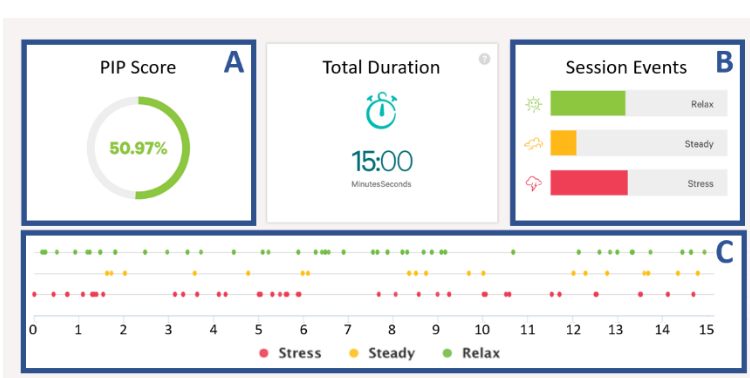


Figure 5 Session Report Generated by 'PIP Stress Tracker'
(A: PIP Score, B: Session Events, C: Event Timeline)

4.2 Accuracy in Estimating Partner's Stress Level (H1)

H1 hypothesized that accuracy of assessing partner stress levels will be greater in pairs with stress indicators than pairs without stress indicators. To examine H1, we ran a mixed models ANOVA examining estimated partner stress as a function of display condition (with or without stress indicator), language background (NS/NNS), the partner's self-reported stress levels, and the rater's own self-reported stress. The latter was included in the model based on literature that shows that people often use their own feelings as a guide for their estimates of others feelings (e.g., [25][73]). Participants were nested in dyads and treated as random.

In this model, the effect of partner's self-reported stress level on estimates is a measure of accuracy: higher positive correlations between participants' estimates and partners' self-reported stress levels indicate greater accuracy. The interaction between display condition and partner self-reported stress tests whether the display improved accuracy. The effect of a rater's own self-reported stress is a measure of bias, which represents how much the estimated stress level of their partners' were based on their own (rater's) personal stress levels.

Table 1 Model Summary

Parameters	Estimate	SE	df	t	Sig.
Intercept	3.37	0.44	396.83	7.67	0.000
Display Condition	-0.78	0.59	371.63	-1.33	0.185
Language	**-2.69**	**0.68**	**409.02**	**-3.97**	**0.000**
Display Condition * Language	0.75	0.88	400.90	0.85	0.396
Rater's Self Rating	**0.46**	**0.04**	**402.45**	**12.10**	**0.000**
Partner's Self Rating	-0.04	0.07	419.30	-0.67	0.506
Display * Partner's Self Rating	0.05	0.10	397.72	0.53	0.599
Language * Partner's Self Rating	**0.28**	**0.12**	**424.56**	**2.29**	**0.022**
Display * Language * Partner's Self Rating	0.08	0.16	413.71	0.48	0.629

Table 1 shows the output from this model. Overall, there was no main effect for partner's self-rating (t [1, 419.30] < 1, ns), suggesting that participants were not skilled at estimating partner's stress levels in general. To test H1, we examined the interaction effect between display condition and partner's self-rating on estimated partners' stress levels. However, this interaction was not significant (t [1, 397.72] = .53, p. = .60). Thus, H1 was not supported.

Instead, we found a significant effect of participant language, such that NS estimates of NNS stress were higher than NNS estimates of NS stress (t [1, 409] = -.3.97, p < .0001). There was also an interaction between participant language and participant self-rating [t [1, 424.56] p = .02). As shown in Figure 6, NNS were fairly accurate in estimating NS partners' stress levels (r=.28, p < .001) but NS were not accurate at assessing of NNS partners' stress levels (r=-.04, p = .48). Correlational analyses for each language condition separately indicated that NS estimates of their partners' stress levels were uncorrelated with their NNS partner's self-ratings both with (r= -.11, p = .28) and without (r=-.01, p= 94) the stress indicator. Furthermore, NNS estimates of their partners' stress levels were more accurate when there was no stress indicator (r= .41, p < .001) than there was stress indicator (r= .17, p = .07; z= 1.88, p = .06). (See Figure 6.)

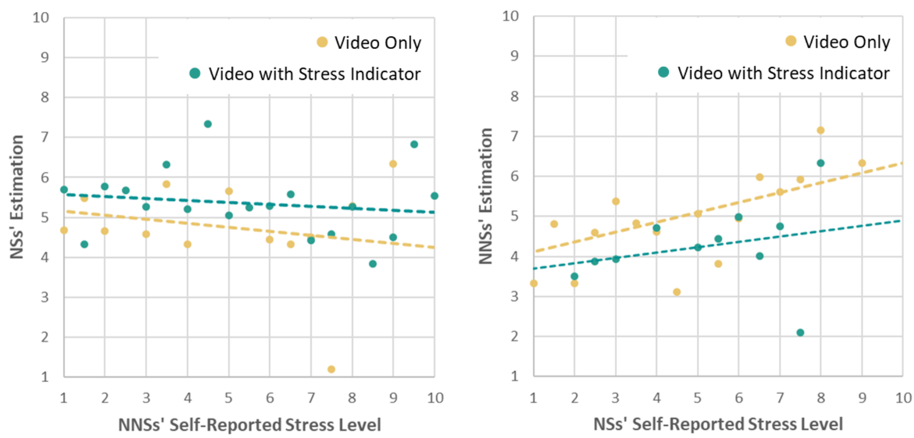


Figure 6 Correlations between self-reported vs. estimated stress ratings
Left: NS's Estimation Accuracy (NS Estimation vs. NNS Self-reported stress levels)
Right: NNS's Estimation Accuracy (NNS Estimation vs. NS Self-reported stress levels)

Finally, there was a strong effect of participants' own self ratings on their estimates of their partners' stress ($t [1, 402.45] = 12.10$, $p < .001$). These results suggest that people's estimates were heavily influenced by their own feelings of stress.

4.3 Privacy Concerns (H2)

H2 hypothesized that participants in pairs with stress indicators would report greater privacy concerns than those without stress indicators. We tested H2 using a two-way ANOVA that examined the effect of stress indicator (with/without) and language background (NS/NNS) on privacy concerns.

Overall, participants' self-reported privacy concerns were low, averaging 1.4 (SE = .69) on a scale of 1 (low) to 5 (High). We found no significant effects of the stress indicator ($p = .09$) or participants' language backgrounds ($p = .20$). However, there was a significant interaction effect between stress indicator and language background ($F [1, 44] = 3.80$, $p = .05$) (see Figure 7). Post hoc simple effects analysis indicated that NNS were more concerned about their privacy with the stress indicator (M = 1.91, SD = .90) than without the indicator (M = 1.20, SD = .39) ($F [1, 22] = 3.01$, $p = .02$). However, the stress indicator did not influence the privacy concerns of NS participants ($p = .89$).

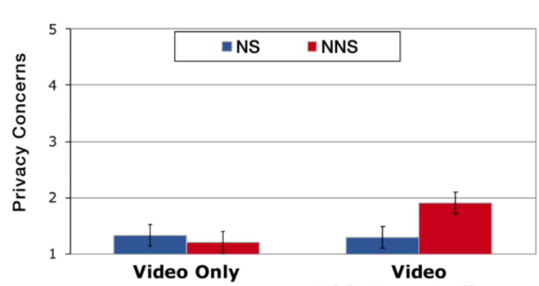


Figure 7 Self-reported concerns about privacy on scale of 1 (Low) to 5 (High) as a function of condition and language background. Error bars represent the standard errors of the means.

4.4 NS Accommodation to NNS Partners (H3)

H3 (a-c) predicted that NS in pairs with stress indicators would exhibit accommodative behaviors more frequently than NS in pairs without stress indicators. For testing these hypotheses, we ran one-way ANOVAs to examine the effect of stress indicators on accommodation behaviors of NS.

H3a hypothesized NS in pairs with stress indicators will speak more slowly than NS in pairs without stress indicators. The analysis result indicated marginal support for H3a. NS in pairs with stress indicators spoke marginally more slowly than NS in pairs without stress indicators (F [1,22] = 3.24, p = .08). The WPM (word per minute) of NS in the stress indicator condition (M = 120.4, SE = 4.8) was lower than that of NS in the no stress indicator condition (M= 132.89, SE= 4.8).

H3b hypothesized that average utterance lengths of NS in pairs with stress indicators would be shorter than NS in pairs without stress indicators. H3b was not supported (F [1, 22] = 1.65, p = .21). Although the average utterance length of NS in stress indicator condition (M = 17.6, SE = 2.9) was shorter than that of NS in no stress indicator condition (M = 22.9, SE = 2.9), it was not significantly different.

H3c predicted that NS in pairs with stress indicators would use more assent words than NS in pairs without stress indicators. H3c was not supported (F [1, 22] = .20, p = .65). There was no significant difference in the percentage of assent words between NS with the stress indicator (M = .041, SE= .006) vs. without the stress indicator (M = .037, SE = .006).

4.5 Perceived Accommodation of NNS (RQ1)

RQ1 inquired how stress indicators would affect NNS's perceptions of their NS partners' accommodations. We ran a one-way ANOVA to examine the effect of stress indicators on NNS's perceptions of their NS partners' accommodation. NNS in dyads with the stress indicator (M = 4.33, SE = .25) rated their NS partner's accommodation as higher than NNS in dyads without stress indicator (M = 3.95, SE = .25), but the difference was not significant (F [1, 21] = 1.30, p = .25).

4.6 Cognitive Load (RQ2)

RQ2 investigated whether stress indicators would increase or decrease participants' cognitive load. A two-way ANOVA examined the effect of stress indicator (with/without) and language background (NS/NNS) on perceived cognitive load.

The main effect of stress indicator condition was marginally significant (F [1, 44] = 3.45, p = .07). Participants with stress indicators (M = 3.16, SD= .49) reported higher cognitive load than those without stress indicators (M = 2.84, SD= .49). We also found a significant main effect of

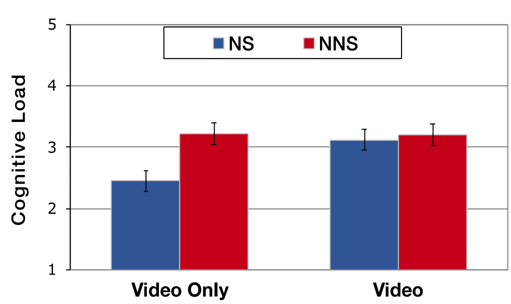


Figure 8 Cognitive load on a scale of 1 (low) to 5 (high) as a function of condition and language background. Errors bars represent the standard errors of the mean.

language background (F[1, 44] = 6.03, p = .01) such that NNS participants (M = 3.21, SD=.54) reported higher cognitive load than NS participants (M = 2.79, SD = .71).

Interestingly, as shown in Figure 8, there was a significant interaction between stress indicator condition and language background (F [1, 44] = 3.91, p = .05). Simple effects analysis showed that NS with stress indicators reported greater cognitive load than NS without stress indicators (p = .01), but the cognitive load of NNS was not influenced by the stress indicator (p= .98).

4.5 Collaboration Satisfaction and Evaluation of the Partner (RQ3)

RQ3 explored the influence of the stress indicator on participants' satisfaction with the collaboration and liking for their partners. We ran two-way ANOVAs that examined the effect of stress indicator (with/without) and language background (NS/NNS) on satisfaction with communication and collaboration. In terms of group communication satisfaction, there was a main effect of language background (F (1, 44) = 7.78, p = .008) such that NNS (M= 4.60, SE = .127) were more satisfied than NS (M = 4.10, SE = .127). However, there was no main effect of stress indicators (p = .48) and interaction effect (p = .35). With respect to performance satisfaction, there was a main effect of language background (F (1, 44) = 3.96, p = .05) such that NS (M = 4.37, SE = .128) were more satisfied than NNS (M = 4.01, SE = .128), but there was no main effect of stress indicators (p = .54) and interaction (p = .44). Lastly, there was no significant effect of stress indicator condition (p = .96) and language background (p = .22), and no interaction effect (p = .42) on evaluation of the partners.

5 INSIGHTS FROM INTERVIEWS

We further deepened our understanding of the effects of stress indicators with the insights through interviews with participants who had stress indicators in their group discussion. We identified interesting themes related to how NS and NNS attended to, perceived, and utilized the stress indicator during the discussion.

5.1 Asymmetric Attention Resources between NS and NNS

When we asked how frequently they paid attention to the stress indicators, most NS answered they paid attention to the stress monitor from time to time during the conversation.

> *Unless I was fully involved in the conversation, I kept checking [stress indicator].* (G14, NS)

However, most NNS responded that they did not have much attentional capacity to monitor the display because they were already overwhelmed by using a second language in the conversation.

> *I could not pay much attention to the stress monitor because I was already overwhelmed using English.* (G14, NNS)

However, when NNS got used to understanding and speaking with their partner, they gradually got to have room to see NS's stress level.

> *As I got gradually accustomed to speaking with my partner, I got the room to look at NS's stress level occasionally.* (G22, NNS)

5.2 Interpreting and Coping with Partner's Stress

NNS participants expressed a strong assumption that their NS partner's stress levels would much lower than their own due to a lack of language issues. Therefore, some NNSs felt it was unnecessary to monitor NSs' stress levels as it the sensor would always show relaxed states.

[Stress indicator] was not necessary for me. Because he [NS partner] is a native English speaker, it would be always green [relaxed]. I was busy dealing with my [language] issue, so I had no room to worry about his stress levels. (G2, NNS)

Since they strongly believed that their NS's partners stress levels would be always stable and relaxed, they became quite surprised or even panicked when the sensor indicated that their NS partner's stress was high.

Against my expectation, my [NS] partner's stress level had ups and downs. I felt bad when my partner's stress went up because I didn't want to make her feel uncomfortable. (G9, NNS)

As stated in the above quote, NNS participants usually attributed the cause of NS's stress to their own lack of language skills. NNS thought that their NS partners must be having a difficult time understanding what they said or accommodating to their needs, which made them feel pressured to lower the NS partner's stress. To reduce this stress, some NNS chose to speak less and agree with their partner's opinions.

When I found his stress was high, I felt panicked, and I thought it was because he could not understand what I was saying. Afterwards, I just kept saying, "I think so" to him to move on. (G16, NNS)

I noticed that her stress level was high when she was speaking. I thought it was because she was paying extra effort - being selective of which words to use so that I could understand her. I felt bad for her and felt like I had to concentrate more. (G13, NNS)

However, none of the NS participants mentioned that the English fluency of their NNS partners made them stressed. Rather, they reported that task-related issues such as time pressure and cognitive activities were the cause of their stress.

I was a little stressed because I could not remember the details of the scenario. (G8, NS)

[My main stressors were] the initial disagreement between [individual] rankings and time pressure. (G14, NS)

NS assumed that the overall stress level of NNS would be moderately higher than themselves due to language fluency. However, they did not expect that NNS's stress would fluctuate much. When they saw NNS's stressed signals, they tried to understand what was causing this stress and how they could make it go down. In many cases, NS participants thought the cause of NNS's stress was related to language issues, especially if there were also signs of misunderstanding, hesitation, or disfluencies. Therefore, they tried out various accommodation strategies to reduce NNS stress:

When I noticed the red levels growing up, I slowed down or asked questions or give her a chance to speak. I think I was trying to respond to it as much as I could." (G16, NS)

When I saw her stress level went up, I tried to say, "I think so too", "good point" more often than usual. I would not have said something if I did not see the signal. (G5, NS)

However, it was not easy for them to figure out what aspects needed to be addressed to let NNS partner's stress down.

Whenever I saw her stress going up, I tried to think about what I can do to make them go down. But it was so complicated. There were too many possibilities. Should I talk slower? Should I be nicer in tone? Or, should I compromise more? (G5, NS)

Some other NS participants attributed the high stress of NNS partners more to the nature of the group decision-making task, which required deeper thinking and complicated negotiation processes. Therefore, it did not trigger any specific accommodation for NNS:

> *I noticed every time she talked for a long time, her stress went up, which would be normal. I think anyone, regardless of language fluency, will be stressed more than usual when you need to express your opinion and think about how other people think.* (G8, NS)

While partner's high stress signals triggered some sorts of social pressure and accommodative behaviors, partner's low stress signals worked as an indication showed that the conversation was going well. Both NS and NNS reported their partner's relaxed signals allowed them to communicate with their partner more comfortably:

> *Her stress levels were pretty much steady, so I was more relaxed to discuss things with her.* (G22, NS)

> *When I saw his stress level went down, I felt relieved and felt encouraged to talk.* (G15, NNS)

5.3 Stress Indicator for Reflection, Understanding and Communication

Overall, NS participants reported that the stress indicators worked as a constant reminder to keep reflecting on their conversation styles and group dynamics. They reported that having stress indicators made them more conscious about themselves and their partners rather than focusing on proceeding with the discussion:

> *Having it [stress indicator] helped me give a track of the direction of the conversation. No matter how accurate it is, thinking about the stress level around you, so it is something useful.* (G16, NS)

Also, some NS found the stress indicator was useful to understand NNS's thoughts and feeling that they found it quite challenging due to the cultural difference:

> *Based on my experience, Japanese people don't show the emotion on the outside, so it is really challenging for me to figure out what's in their mind. So that [stress indicator] could be really useful.* (G13, NS)

Similarly, some NNS mentioned the usefulness of stress indicators in understanding different conversational norms across cultures. For example, as the silence could mean either negative or positive depending on the context and cultures, NNS in Group18 was not sure what the silence of her NS partner meant. However, based on the relaxed status of NS, she assumed it was okay to move on:

> *When there was silence, I took a look at [stress indicator] to guess what the silence meant. When I saw his stress level, it was low so that I could concentrate on thinking.* (G18, NNS)

While NS found the stress indicator helpful for self-reflection, NNS participants saw sharing their stress levels as an effective medium to communicate their (language-related) struggles to tiger NS's accommodation:

> *I felt the stress indicator was useful in giving him some clue for accommodation.* (G16, NNS)

> *One time, I felt stressed when I could not come up the English words. I thought my partner saw my stress level rising and he tried to use easier words and shorter sentences.* (G7, NNS)

By showing their (high) stress levels via stress indicators, they felt safe because they believed their challenges in using second language would be conveyed to their NS partner more naturally without needs to ask for more accommodation or express their struggles explicitly:

> *Because I could not express my thoughts fully as much as I wished, giving my partner additional cue that I had trouble speaking in English (via stress indicator) was helpful. By*

doing so, my partner would not assume I was just blunt or not competent but just having issues with speaking English. (G7, NNS)

6 DISCUSSION

In this study, we investigated the effects of the presence of stress indicators during the group conversations between NS and NNS speakers. Overall, our quantitative analysis of estimation accuracy [H1], collaboration satisfaction [RQ3] and evaluation of partners [RQ1, RQ3] did not demonstrate any significant effects of stress indicators. While there was a trend for NS with the stress indicator to accommodate their partners by speaking more slowly [H3a], our results indicated that the mere presence of stress indicators did not lead to better multilingual communication processes and outcomes. However, our interview data and quantitative results on privacy concerns [H2] and cognitive load [RQ2] revealed that additional dimensions of how differently NS and NNS attended to, perceived, and responded to partners' stress indicators. Below, we first discuss these results in more detail and examine how real-time stress indicators might be incorporated into tools to improve multilingual communication.

First, although we expected that the stress indicators would increase awareness of partners' psychological states [H1], the accuracy in estimating partner's stress levels did not significantly differ between conditions.

Most of all, this result could be explained by the inaccuracy of the stress sensor as we discussed in 4.1. Although the stress indicator was reasonably accurate at the session level (15 minutes), its accuracy in the shorter term (15-20 seconds) was not reliable, perhaps due to the limited capacity (e.g., low sampling rate) of the consumer sensor that we utilized for the study. Furthermore, for the most accurate detection of stress, it is important to limit body movements, irregular breathing and speech [64]. However, in our experiment task, we could not impose such restrictions on participants. Due to the unstable accuracy of the stress level visualization, we limited our findings to the effects of the 'presence' of the stress indicators, while future research is needed using more precise sensor equipment.

Another potential account for the H1 [Estimation accuracy] result was that participants relied on their own perceived stress levels when estimating their partner's stress levels as shown in the strong correlation between estimation vs. rater's self-stress levels. Therefore, the magnitude of effects that partner's stress signals might have been minimal. Similarly, people had preconceived ideas about their partners' likely stress levels (e.g., NNS assumed NS would always have low stress). Given that NNS were not able to pay close attention to the stress indicator signals, we suspect they may have used the indicator as a way to confirm their preconceptions about their NS partner (i.e., low stress).

Yet, NNS were more accurate in estimating their partner's status than NS, regardless of the presence of stress indicators. This result is consistent with previous findings that NNS in multilingual communication are more accurate in understanding the reasons for NS behavior than NS are at understanding NNS behavior [35], for a variety of reasons including NNS's tendency to conceal their feelings to maintain face [4][59] and cultural differences in emotional sensitivity (e.g., [43]).

The results on privacy concerns [H2] and cognitive load [RQ2] revealed that additional dimensions of how differently NS and NNS perceived and attended to the stress indicators. First, the presence of stress indicators increased privacy concerns for NNS but not NS participants. We speculate that it might be caused by NNS's reluctance to share their inner state in general [4][59].While the presence of stress indicators increased the privacy concerns of NNS to some extent, our interview data showed that NNS thought the stress indicator was useful as a way to expressing their language-related struggles to their NS partners. Conversely, the presence of stress indicators imposed additional cognitive load only on NS, not NNS, although NNS participants reported higher cognitive load than NS regardless of the presence of stress indicators. The additional cognitive load of NS might be because NS had to process the stress signals of their NNS partner and accommodate NNS as reported in the interview.

As such, our interview data suggested how NS and NNS found stress indicators useful for better multilingual communication: NS found stress indicators useful as a self-reflection tools to monitor their conversation style in accommodating their NS partner. NNS found it useful as a communication medium to express language-related struggles to give the cues to NNS for accommodation. However, the quantitative result of the effects of stress indicators on NS accommodation [H3] showed marginal support for only reduced speech rate [H3a]. Also, NNS's perceived accommodation of NS [RQ1] was not increased with the presence of stress indicators. Furthermore, RQ3 [Social Outcomes] results did not demonstrate the direct benefits of sharing stress indicators in terms of group satisfaction and liking for partners.

These differences between the quantitative and qualitative results could be explained in part by the inaccuracy of the moment-by-moment stress signals. Assuming that NSs were motivated to accommodate only when they saw NNS's high stress, NS's accommodation may not have happened at the right moment when NNS felt stressed and needed further accommodation. Further, even if the presence of stress indicator made NS more motivated to attempt more accommodation toward their NNS partner, it is possible that NS's accommodation might not have been effective or sufficient. Interview data showed that NSs tried out multiple different ways to accommodate NNS partners, such as slowing down, being nice in tone or compromising, and agreeing more. However, they were sometimes unsure of which aspects need to be accommodated exactly. Therefore, even if the stress signals might have triggered more accommodation attempts by NS, this might not lead to better actual accommodation.

6.1 Design Implications

Based on the results and discussion, we discuss how the results from our study can be used to design a better application of stress indicators for multilingual communication contexts.

6.1.1 Meta Talk about the Concept of Stress

Our results indicated that participants' previous conception of their own and partner's stress influences how they interpreted, utilized and responded to the stress indicators. Therefore, it is important to provide shared knowledge about each other's beliefs about stress before incorporating stress indicators into the multilingual communication settings. One way to ensure better awareness of stress is by facilitating "meta talk" about what makes them stressed and how they think about each other's stress [34][35]. In the meta talk, NNS could make it clear that their stress mostly stems from language issues, while NS clarify that NNS language was not the source of their stress. If they have sufficient understanding of each other's concepts of stress and primary stressors, it might help partners interpret and respond to the partner's real-time stress signals more effectively. It may also be useful to provide participants with a display that shows both their own and their partners real time stress levels, to help them see that their own stress levels are not always good predictors of their partner's stress.

6.1.2 Customizing Sensor Displays

Our interview data showed that NS and NNS differed in how they reported using and reacting to the stress indicators. NS found stress indicators helpful as a cue to reflect their own communication styles. NNS utilized stress indicators as a communication medium to express their stress to NS. Based on these findings, we suggest using different ways of displaying the sensor data. For example, NS's stress displays could be optimized for comprehensive monitoring of the trends and changes of NNS's stress. Also, the interface could incorporate other kinds of feedback that could be helpful for accommodating NNS such as a speech speedometer [18], or language-based feedback [16].

For NNS, in contrast, it might be desirable to show NS partners' sensor data only on demand, since NNS often misinterpreted the reasons for NS stress levels (e.g., attributing them to their own language instead of task difficulty or time pressure). Also, the interface could provide a mechanism for NNS to indicate needs for greater accommodation, for example by magnifying their stress signals to make them more salient to NS.

Another idea would be to manipulate the display for better communication processes, which is a similar design approach to distorted social feedback of conversation [7] or false psychological feedback systems [13]. For example, to enhance NS accommodation behaviors, the display for an NNS partner might be made artificially high, and to reduce NNS's social pressure to reduce NS's stress, the display for an NS partner might be made artificially low.

6.1.3 Scaffolding the Accommodation of NS

Another issue was that NS was sometimes unable to adjust their behaviors and accommodate NNS as a response to the NNS's high stress signals. One way to address this is to give NS knowledge about available and effective accommodation strategies, perhaps within the system. Also, a more advanced system such as augmented social coaching systems (e.g., [15]) could suggest the exact strategies for them to use to accommodate NNS appropriately.

6.2 Limitations and Future Directions

The findings of the current study need to be interpreted with several limitations in mind, and these limitations also lay out directions for future work to explore.

First, our approach (lab experiment) trades ecological validity for control while we acknowledge examining the effects of stress indicators in a natural setting is crucial. We made that tradeoff because we thought control was important for getting an initial understanding of how people perceive and utilize stress indicators in multilingual communication contexts. We plan to follow up with a field study in which we can examine the role of other factors such as participants' cultural backgrounds and interaction contexts (e.g., task-oriented vs. relationship building) that could affect the outcomes.

Second, the stress measure and equipment for tracking one's stress level is a consumer level device, so the accuracy of information was found to be unreliable. Therefore, we need to test our research questions and hypotheses with more precise measurements and equipment.

Third, the study design could be improved with additional control conditions (e.g., secondary display showing irrelevant information), to tease apart the effects of secondary display vs. the efforts in interpreting and responding to stress signals. Also, it would be useful to implement eye-tracking to keep track of people's attention to the stress indicator display.

Fourth, since different methods for the visualizations of the same biosignal could evoke different interpretations [46], we further need to investigate how other kinds of visualizations influence the ways people make sense of stress indicators.

Last, the population of the study was native Japanese and native English speakers who had a lot of experience interacting with NNS. Therefore, our NS participants might already have stronger skills for accommodating NNS. Also, different cultural backgrounds of NNS could lead to different results. Therefore, future research needs to include other populations, such as NS who have little experience interacting with NNS and NNS who are from different cultural backgrounds (e.g., Individualistic culture) from NNS in Japan.

7 CONCLUSION

In this paper, we presented a study that systematically explored how sharing the real-time stress signals of communication partners affects multilingual video-mediated communication. The presence of real-time stress indicators did not lead to a better understanding of partners' stress levels or better satisfaction on communication while NS tended to accommodate NNS partners by slowing down their speech. The presence of stress indicators increased the cognitive load of NS but not NNS participants, and it increased the privacy concerns of NNS but not NS participants. Our qualitative analysis of interview data showed that NS and NNS had different conceptualizations of sources of stress, which led them to misinterpret their partners' stress signals. Our findings suggest several ways that future CMC tools might better support multilingual communication with the potential application of real-time stress indicator displays.

REFERENCES

[1] Nathalie Aichhorn and Jonas Puck. 2017. "I just don't feel comfortable speaking English": Foreign language anxiety as a catalyst for spoken-language barriers in MNCs. *International Business Review* 26, 4: 749–763.

[2] Moshaddique Al Ameen, Jingwei Liu, and Kyungsup Kwak. 2012. Security and privacy issues in wireless sensor networks for healthcare applications. Journal of medical systems 36, 1: 93–101.

[3] Mortimer H. Appley and Richard A. Trumbull eds. 2012. *Dynamics of Stress Physiological, Psychological and Social Perspectives*, Springer Verlag.

[4] Terry Kit Fong Au, Annie Fong Pui Kwok, Lester Chun Pong Tong, Liao Cheng, Hannah Man Yan Tse, and Sun Ah Jun. 2017. The social costs in communication hiccups between native and nonnative speakers. *Journal of Cross-Cultural Psychology* 48, 3: 369–383. https://doi.org/10.1177/0022022116687852

[5] Jorn Bakker, Mykola Pechenizkiy, and Natalia Sidorova. 2011. What's your current stress level? Detection of stress patterns from GSR sensor data. In *Proceedings - IEEE International Conference on Data Mining, ICDM*, 573–580. https://doi.org/10.1109/ICDMW.2011.178

[6] Christian Becker-Asano and Hiroshi Ishiguro. 2011. Intercultural differences in decoding facial expressions of the android robot Geminoid F. Journal of Artificial Intelligence and Soft Computing Research 1.

[7] Tony Bergstrom and Karrie Karahalios. 2012. Distorting Social Feedback in Visualizations of Conversation. *2012 45th Hawaii International Conference on System Sciences* (2012). DOI:http://dx.doi.org/10.1109/hicss.2012.222

[8] Kirsten Boehner, Rogério Depaula, Paul Dourish, and Phoebe Sengers. 2007. How emotion is made and measured. *International Journal of Human-Computer Studies* 65, 4 (2007), 275–291. DOI:http://dx.doi.org/10.1016/j.ijhcs.2006.11.016

[9] Angelika Breiteneder. 2009. English as a lingua franca in Europe: An empirical perspective. World Englishes 28, 2: 256–269.

[10] John T. Cacioppo and Louis G. Tassinary. 1990. Inferring psychological significance from physiological signals. *American Psychologist* 45, 1 (1990), 16–28. DOI:http://dx.doi.org/10.1037/0003-066x.45.1.16

[11] Xun Cao, Naomi Yamashita, and Toru Ishida. 2016. Investigating the impact of automated transcripts on non-native speakers' listening comprehension. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction*, 121–128.

[12] Sheldon Cohen, Denise Janicki-Deverts, and Gregory E. Miller. 2007. Psychological Stress and Disease. *Jama* 298, 14 (October 2007), 1685. DOI:http://dx.doi.org/10.1001/jama.298.14.1685

[13] Jean Costa, Alexander T. Adams, Malte F. Jung, François Guimbretière, and Tanzeem Choudhury. 2016. EmotionCheck: leveraging bodily signals and false feedback to regulate our emotions. In Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing(UbiComp '16). ACM, New York, NY, USA, 758-769. DOI: https://doi.org/10.1145/2971648.2971752

[14] Nikolas Coupland, Justine Coupland, Howard Giles, and Karen Henwood. 1988. Accommodating the elderly: Invoking and extending a theory. *Language in society* 17, 1: 1–41.

[15] Ionut Damian, Chiew Seng Sean Tan, Tobias Baur, Johannes Schöning, Kris Luyten, and Elisabeth André. 2015. Augmenting social interactions: Realtime behavioural feedback using social signal processing techniques. In *Proceedings of the 33rd annual ACM conference on Human factors in computing systems*, 565–574.

[16] E Ilana Diamant, Brian Y Lim, Andy Echenique, Gilly Leshed, and Susan R Fussell. 2009. Supporting intercultural collaboration with dynamic feedback systems: preliminary evidence from a creative design task. In *CHI'09 Extended Abstracts on Human Factors in Computing Systems*, 3997–4002.

[17] Alison Dillon, Mark Kelly, Ian H. Robertson, and Deirdre A. Robertson. 2016. Smartphone Applications Utilizing Biofeedback Can Aid Stress Reduction. *Frontiers in Psychology* 7 (2016). DOI:http://dx.doi.org/10.3389/fpsyg.2016.00832

[18] Wen Duan, Naomi Yamashita, and Susan R. Fussell. 2019. Increasing Native Speakers Awareness of the Need to Slow Down in Multilingual Conversations Using a Real-Time Speech Speedometer. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (July 2019), 1–25. DOI:http://dx.doi.org/10.1145/3359273

[19] Joan Morris DiMicco, Vidya Lakshmipathy, and Andrew Tresolini Fiore. 2002. Conductive Chat: Instant messaging with a skin conductivity channel. In *Proceedings of Conference on Computer Supported Cooperative Work.*

[20] Andy Echenique, Naomi Yamashita, Hideaki Kuzuoka, and Ari Hautasaari. 2014. Effects of video and text support on grounding in multilingual multiparty audio conferencing. In *Proceedings of the 5th ACM international conference on Collaboration across boundaries: culture, distance & technology*, 73–81.

[21] Susanne Ehrenreich. 2010. English as a business lingua franca in a German multinational corporation: Meeting the challenge. The Journal of Business Communication (1973) 47, 4: 408–431.

[22] Anna-Liisa Elo, Anneli Leppänen, and Antti Jahkola. 2003. Validity of a single-item measure of stress symptoms. *Scandinavian Journal of Work, Environment & Health* 29, 6 (2003), 444–451. DOI:http://dx.doi.org/10.5271/sjweh.752

[23] Alan J. Feely and Anne-Wil Harzing. 2003. Language management in multinational companies. *Cross Cultural Management: An International Journal* 10, 2: 37–52. https://doi.org/10.1108/13527600310797586

[24] Jérémy Frey and Jessica R. Cauchard. 2018. Remote Biofeedback Sharing, Opportunities and Challenges. Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers - UbiComp 18 (2018). DOI:http://dx.doi.org/10.1145/3267305.3267701

[25] Susan R. Fussell and Robert M. Krauss. 1992. Coordination of knowledge in communication: Effects of speakers assumptions about what others know. *Journal of Personality and Social Psychology* 62, 3 (1992), 378–391. DOI:http://dx.doi.org/10.1037/0022-3514.62.3.378

[26] Cindy Gallois, Tania Ogay, and Howard Giles. 2005. Communication Accommodation Theory: A Look Back and a Look Ahead. *W.B. Gudykunst (ed) Theorizing About Communication and Culture*, January: 121–148. https://doi.org/10.1002/9781118611463.wbielsi066

[27] Ge Gao, Naomi Yamashita, Ari M.J. Hautasaari, and Susan R. Fussell. 2015. Improving Multilingual Collaboration by Displaying How Non-native Speakers Use Automated Transcripts and Bilingual Dictionaries. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*, 3463–3472. https://doi.org/10.1145/2702123.2702498

[28] Jessica Gasiorek and Howard Giles. 2012. Effects of Inferred Motive on Evaluations of Nonaccommodative Communication. *Human Communication Research* 38, 3: 309–331. https://doi.org/10.1111/j.1468-2958.2012.01426.x

[29] Howard Giles, Anthony Mulac, James J. Bradac, and Patricia Johnson. 1987. Speech Accommodation Theory: The First Decade and Beyond. *Annals of the International Communication Association* 10, 1 (1987), 13–48. DOI:http://dx.doi.org/10.1080/23808985.1987.11678638

[30] Nitesh Goyal and Susan R. Fussell. 2017. Intelligent Interruption Management using Electro Dermal Activity based Physiological Sensor for Collaborative Sensemaking. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 1, 3, Article 52

[31] Tammy S Gregersen. 2005. Nonverbal cues: Clues to the detection of foreign language anxiety. Foreign Language Annals 38, 3: 388–400.

[32] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in psychology* 52: 139–183.

[33] Mariam Hassib, Daniel Buschek, Paweł W Wozniak, and Florian Alt. 2017. HeartChat: Heart rate augmented mobile chat to support empathy and awareness. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 2239–2251.

[34] Helen Ai He, Naomi Yamashita, Ari Hautasaari, Xun Cao, and Elaine M Huang. 2017. Why Did They Do That? Exploring Attribution Mismatches Between Native and Non-Native Speakers Using Videoconferencing. In *CSCW 2017*, 297–309. https://doi.org/10.1145/2998181.2998205

[35] Helen Ai He, Naomi Yamashita, Elaine M Huang, Chat Wacharamanotham, Andrea B Horn, and Jenny Schmid. 2017. Two Sides to Every Story: Mitigating Intercultural Conflict through Automated Feedback and Shared Self-Reflections in Global Virtual Teams. *Proc. ACM Hum.-Comput. Interact* 1, 21. https://doi.org/10.1145/3134686

[36] Stevan E. Hobfoll. 2004. Stress, Culture, And Community: the Psychology And Philosophy Of Stress, Plenum Pub Corp.

[37] Noura Howell, Laura Devendorf, Rundong Kevin Tian, Tomás Vega Galvez, Nan-Wei Gong, Ivan Poupyrev, Eric Paulos, and Kimiko Ryokai. 2016. Biosignals as social cues: Ambiguity and emotional interpretation in social displays of skin conductance. In Proceedings of the 2016 ACM Conference on Designing Interactive Systems, 865–870.

[38] Priska Imberti. 2007. Who Resides behind the Words? Exploring and Understanding the Language Experience of the Non-English-Speaking Immigrant. *Families in Society: The Journal of Contemporary Social Services* 88, 1 (2007), 67–73. DOI:http://dx.doi.org/10.1606/1044-3894.3593

[39] International Actuarial Association. 2014. *Participation of Non-Native English Speakers and New Delegates in the Work of the IAA.* Retrieved September 16, 2019 from http://www.actuaries.org/ABOUT/Documents/Participation_NonNative_English_Speakers_EN.pdf

[40] Elizabeth Jones, Cynthia Gallois, Victor Callan, and Michelle Barker. 1999. Strategies of accommocation: Development of a coding system for conversational interaction. Journal of Language and Social Psychology 18, 2: 123–151.

[41] Jina Kim, Young-Woo Park, and Tek-Jin Nam. 2015. BreathingFrame: An inflatable frame for remote breath signal sharing. In *Proceedings of the Ninth International Conference on Tangible, Embedded, and Embodied Interaction*, 109–112.

[42] Taemie Kim, Agnes Chang, Lindsey Holland, and Alex Sandy Pentland. 2008. Meeting mediator. *Proceedings of the ACM 2008 conference on Computer supported cooperative work - CSCW 08* (2008). DOI:http://dx.doi.org/10.1145/1460563.1460636

[43] Naz Kaya and Margaret J Weber. 2003. Cross-cultural differences in the perception of crowding and privacy regulation: American and Turkish students. *Journal of environmental psychology* 23, 3: 301–309.

[44] Adam D.I. Kramer, Susan R. Fussell, and Leslie D. Setlock. 2004. Text analysis as a tool for analyzing conversation in online support groups. In *Conference on Human Factors in Computing Systems - Proceedings.* https://doi.org/10.1145/985921.986096

[45] Yi Li, Adel S. Elmaghraby, and Estate M. Sokhadze. 2015. Designing immersive affective environments with biofeedback. *2015 Computer Games: AI, Animation, Mobile, Multimedia, Educational and Serious Games (CGAMES)* (2015). DOI:http://dx.doi.org/10.1109/cgames.2015.7272965

[46] Fannie Liu, Laura Dabbish, and Geoff Kaufman. 2017. Can Biosignals be Expressive? *Proceedings of the ACM on Human-Computer Interaction* 1, CSCW (June 2017), 1–21. DOI:http://dx.doi.org/10.1145/3134706

[47] Fannie Liu, Geoff Kaufman, and Laura Dabbish. 2019. The Effect of Expressive Biosignals on Empathy and Closeness for a Stigmatized Group Member. Proc. ACM Hum.-Comput. Interact. 3, CSCW, Article 201 (November 2019), 17 pages. DOI:https://doi.org/10.1145/3359303

[48] Mark Matthews et al. 2015. Real-Time Representation Versus Response Elicitation in Biosensor Data. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI 15* (2015). DOI:http://dx.doi.org/10.1145/2702123.2702485

[49] Anna Mauranen. 2003. The Corpus of English as Lingua Franca in Academic Settings. *TESOL Quarterly* 37, 3 (January 2003), 513. DOI:http://dx.doi.org/10.2307/3588402

[50] Senko K. Maynard. 1986. On back-channel behavior in Japanese and English casual conversation. *Linguistics* 24, 6 (1986). DOI:http://dx.doi.org/10.1515/ling.1986.24.6.1079

[51] James C. McCroskey and Thomas A. McCain. 1974. The measurement of interpersonal attraction. Speech Monographs 41, 3: 261–266. https://doi.org/10.1080/03637757409375845

[52] Nick Merrill and Coye Cheshire. 2016. Habits of the Heart (rate): Social Interpretation of Biosignals in Two Interaction Contexts. In Proceedings of the 19th international conference on supporting group work, 31–38.

[53] Nick Merrill and Coye Cheshire. 2017. Trust Your Heart: Assessing Cooperation and Trust with Biosignals in Computer-Mediated Interactions. *CSCW '17 Proceedings of the ACM 2017 conference on Computer Supported Cooperative Work*: 2–12. https://doi.org/10.1145/2998181.2998286

[54] Alireza Moghaddam. 2006. Coding issues in grounded theory. Issues in educational research 16, 1: 52-66

[55] Gary R. Morrow and Anthony H. Labrum. 1978. The relationship between psychological and physiological measures of anxiety. *Psychological Medicine* 8, 1 (1978), 95–101. DOI:http://dx.doi.org/10.1017/s0033291700006668

[56] Mei-Hua Pan, Naomi Yamashita, and Hao-Chuan Wang. 2017. Task Rebalancing: Improving Multilingual Communication with Native Speakers-Generated Highlights on Automated Transcripts. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, 310–321.

[57] Alexandros Pantelopoulos and Nikolaos G Bourbakis. 2010. A survey on wearable sensor-based systems for health monitoring and prognosis. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 40, 1: 1–12.

[58] James W Pennebaker, Martha E Francis, and Roger J Booth. 1999. Linguistic Inquiry and Word Count. *Word Journal Of The International Linguistic Association.* https://doi.org/10.4018/978-1-60960-741-8.ch012

[59] Pamela Rogerson-Revell. 2008. Participation and performance in international business meetings. *English for Specific Purposes* 27, 3: 338–360. https://doi.org/10.1016/j.esp.2008.02.003

[60] Pamela Rogerson-Revell. 2010. "CAN YOU SPELL THAT FOR US NONNATIVE SPEAKERS?". *Journal of Business Communication* 47, 4: 431–454. https://doi.org/DOI: 10.1177/0021943610377304

[61] Ellen Bouchard Ryan, Mary Lee Hummert, and Linda H Boich. 1995. Communication predicaments of aging: Patronizing behavior toward older adults. Journal of Language and Social Psychology 14, 1–2: 144–166.

[62] Klaus R Scherer, Rainer Banse, and Harald G Wallbott. 2001. Emotion inferences from vocal expression correlate across languages and cultures. Journal of Cross-cultural psychology 32, 1: 76–92.

[63] Barbara Seidlhofer, Angelika Breiteneder, and Marie-Luise Pitzl. 2006. English As A Lingua Franca In Europe: Challenges For Applied Linguistics. *Annual Review of Applied Linguistics* 26 (2006). DOI:http://dx.doi.org/10.1017/s026719050600002x

[64] Society for Psychophysiological Research Ad Hoc Committee on Electrodermal Measures, Wolfram Boucsein, Don C Fowles, Sverre Grimnes, Gershon Ben-Shakhar, Walton T Roth, Michael E Dawson, and Diane L Filion. 2012. Publication recommendations for electrodermal measurements. *Psychophysiology* 49, 8 (August 2012), 1017–1034. DOI:http://dx.doi.org/10.1111/j.1469-8986.2012.01384.x

[65] Yu Shi, Natalie Ruiz, Ronnie Taib, Eric Choi, and Fang Chen. 2007. Galvanic skin response (GSR) as an index of cognitive load. *CHI 07 extended abstracts on Human factors in computing systems - CHI 07* (2007). DOI:http://dx.doi.org/10.1145/1240866.1241057

[66] Kimiko Shimoda, Michael Argyle, and Pio Ricci Bitti. 1978. The intercultural recognition of emotional expressions by three national racial groups: English, Italian and Japanese. European Journal of Social Psychology 8, 2: 169–179.

[67] Petr Slovák, Joris Janssen, and Geraldine Fitzpatrick. 2012. Understanding heart rate sharing: towards unpacking physiosocial space. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 859–868.

[68] Emma Sweeney and Zhu Hua. 2010. Accommodating toward your audience: Do native speakers of english know how to accommodate their communication strategies toward nonnative speakers of english? *Journal of Business Communication* 47, 4: 477–504. https://doi.org/10.1177/0021943610377308

[69] Yohtaro Takano and Akiko Noda. 1993. A temporary decline of thinking ability during foreign language processing. *Journal of Cross-Cultural Psychology* 24, 4: 445–462.

[70] Chiew Seng Sean Tan, Johannes Schöning, Kris Luyten, and Karin Coninx. 2014. Investigating the effects of using biofeedback as visual stress indicator during video-mediated collaboration. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14*, 71–80. https://doi.org/10.1145/2556288.2557038

[71] Hanne Tange and Jakob Lauring. 2009. Language management and social interaction within the multilingual workplace. *Journal of Communication Management* 13, 3: 218–232.

[72] David R. Thomas. 2006. A General Inductive Approach for Analyzing Qualitative Evaluation Data. *American Journal of Evaluation* 27, 2 (2006), 237–246. DOI:http://dx.doi.org/10.1177/1098214005283748

[73] Amos Tversky and Daniel Kahneman. 1982. Judgment under uncertainty: Heuristics and biases. *Judgment under Uncertainty* (1982), 3–20. DOI:http://dx.doi.org/10.1017/cbo9780511809477.002

[74] James P. Wolf. 2008. The effects of backchannels on fluency in L2 oral task production. *System* 36, 2 (2008), 279–294. DOI:http://dx.doi.org/10.1016/j.system.2007.11.007

[75] Naomi Yamashita, Andy Echenique, Toru Ishida, and Ari Hautasaari. 2013. Lost in Transmittance: How Transmission Lag Enhances and Deteriorates Multilingual Collaboration. In *ACM Conference on Computer Supported Cooperative Work Social Computing Social Computing*, 923–934. https://doi.org/10.1145/2441776.2441881

[76] Svetlana Yarosh, Panos Markopoulos, and Gregory D. Abowd. 2014. Towards a questionnaire for measuring affective benefits and costs of communication technologies. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW*. https://doi.org/10.1145/2531602.2531634

[77] Chien Wen Yuan, Leslie D Setlock, Dan Cosley, and Susan R Fussell. 2013. Understanding informal communication in multilingual contexts. In *Proceedings of the 2013 conference on Computer supported cooperative work*, 909–922.

[78] Jane Zuengler. 1991. Accommodation in native-nonnative interactions: Going beyond the "what" to the "why" in second language research. Contexts of accommodation: Developments in applied sociolinguistics: 223–244

[79] Human Synergistics. Survival Simulation Series - Team building activity. Retrieved January 9, 2020 from https://www.humansynergistics.com/change-solutions/change-solutions-for-groups-and-teams/team-building-simulations/survival-series