Nonverbal Dynamics in Dyadic Videoconferencing Interaction: The Role of Video Resolution and Conversational Quality

Chenyao Diao
chenyao.diao@tu-ilmenau.de
Technische Universität Ilmenau
Ilmenau, Thuringia, Germany

Stephanie Arévalo Arboleda
stephanie.arevalo@tu-ilmenau.de
Technische Universität Ilmenau
Ilmenau, Thuringia, Germany

Alexander Raake
alexander.raake@tu-ilmenau.de
Technische Universität Ilmenau
Ilmenau, Thuringia, Germany

ABSTRACT

In this paper, we investigated the influence of video resolution and perceived conversational quality on nonverbal behaviors during dyadic videoconferencing (VC) conversations. We analyzed nonverbal behaviors at the individual and interpersonal level. At the individual level, we considered body motion, facial expressions, and gaze directivity. At the interpersonal level, we considered facial expression synchrony and body movement synchrony. For the analysis, we used webcam recordings from a VC experiment, extracting the aforementioned individual nonverbal behavioral features and using windowed lagged cross-correlation (WLCC) to quantify the degree of interpersonal synchronization. Our results indicate that high video resolution significantly increased individual body movements and encouraged gaze directivity toward the conversational partner, fostering greater engagement while paradoxically reducing body movement synchrony. Higher conversational quality was associated with increased facial expression synchrony between participants. Moreover, we observed that instantaneous synchrony (as quantified with lag-zero WLCC) for both body movement and facial expressions was significantly influenced by mutual gaze-like behavior. These findings indicate a complex relationship between technical settings and nonverbal behaviors, suggesting that while higher resolution enhances some nonverbal behaviors, especially body movement and mutual gaze, it may disrupt body movement synchronization. These insights could be applied to the VC setups to achieve a high level of interpersonal coordination and engagement.

CCS CONCEPTS

• **Information systems** → **Web conferencing**; • **Human-centered computing** → **User studies**.

KEYWORDS

Videoconferencing, Video resolution, Gaze behavior, Movement coordination, Facial mimicry

ACM Reference Format:
Chenyao Diao, Stephanie Arévalo Arboleda, and Alexander Raake. 2024. Nonverbal Dynamics in Dyadic Videoconferencing Interaction: The Role of Video Resolution and Conversational Quality. In *INTERNATIONAL CONFERENCE ON MULTIMODAL INTERACTION (ICMI '24), November 04–08, 2024, San Jose, Costa Rica.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3678957.3685733

ICMI '24, November 04–08, 2024, San Jose, Costa Rica

ACM ISBN 979-8-4007-0462-8/24/11
https://doi.org/10.1145/3678957.3685733

1 INTRODUCTION

From work meetings to sports training, videoconferencing (VC) technologies provide an effective way of transmitting essential nonverbal cues in regard to body movements, facial expressions, and eye gaze that are crucial for conveying emotions and intentions in face-to-face (F2F) communication [8].

Recent research has shown the importance of interpersonal nonverbal synchrony (e.g., facial expression synchrony [32, 51] and body movement synchrony [53]) in strengthening rapport and cooperation during VC interactions [16, 33]. Moreover, it has been found that this behavior could significantly influence the subjective quality and effectiveness of communication [29, 58].

However, the quality of the VC interaction can be influenced by technical impairments such as video degradation and transmission delay [38, 46, 57]. Existing studies suggest that these impairments could significantly impact the interpretation and use of nonverbal cues, leading to difficulty in detecting facial cues [20] and inaccuracies in eye gaze interpretation [49], resulting in potential misunderstandings.

Despite extensive studies investigating the influence of technical impairments on VC interaction experience, a comprehensive understanding of how VC quality influences nonverbal behavior remains elusive. This paper investigates how variations in video resolution and the resulting overall conversational quality affect individual nonverbal behavior, including body movements, facial expressions, and gaze behavior. Additionally, we explored how video resolution and the perceived conversational quality influence interpersonal coordination in the context of facial expression synchrony and body movement synchrony. We carried out a user study in which participants engaged in a Celebrity Name Guessing (CNG) game using VC at three video resolution settings (240p, 480p, and 1080p). The perceived conversational quality during a video call was measured as the overall satisfaction with the communication experience. Following the guidelines set by ITU-T recommendations [23, 24], this subjective measurement includes several factors, such as the clarity of speech, ease of understanding, and naturalness of interaction. We extracted body motion, facial action unit activity, and gaze direction from webcam recordings. To quantify the coordination of facial expressions and body movements between participants, we employed windowed lagged cross-correlation (WLCC), a method that allows us to measure the temporal correlations of nonverbal signals in a dyad. Our main findings are:

- There is a significant effect of video resolution on nonverbal behaviors, with high resolution increasing individual body

movements and facial expressions and leading to longer gazes at conversation partners. However, in terms of interpersonal coordination, high video resolution significantly reduced body movement synchrony.
- Perceived conversational quality did not lead to significant changes in individual nonverbal behavior. However, it significantly impacted the effectiveness of interpersonal coordination, with a better conversational experience associated with quicker facial mimicry reactions and greater global synchrony of facial expressions.

2 RELATED WORK

2.1 Interpersonal Coordination

Interpersonal coordination refers to the natural tendency of participants in a conversation to match and synchronize their physical behaviors with others [5, 56]. It denotes facial expression synchrony, i.e., mirroring interlocutors' facial expressions [19, 31], and body movement synchrony, i.e., imitating the body movements of others [10, 53]. These behaviors have been well studied in F2F communication, emphasizing its role in improving social connections and interaction experiences [10, 27, 29, 31, 40, 44, 53, 54].

For instance, studies have demonstrated that facial expression synchrony during social interaction can improve social likability [27] and team cohesion [32]. Tschacher et al. [53] investigated the influence of bodily synchrony between patients and therapists during psychotherapy sessions on treatment outcomes. Their results indicated that a higher level of body movement synchrony correlates with improved session outcomes, leading to greater symptom reduction and enhanced relationship quality. Vacharkulksemsuk and Fredrickson [54] found that body movement synchrony among strangers significantly mediated the positive effects of self-disclosure on rapport formation.

Recent studies have highlighted the differences in how interpersonal synchrony is achieved across communication channels. Han et al. [17] investigated the impact of communication channels on the synchronization of physiological responses. They found that F2F interaction facilitates a higher degree of synchrony than video calls. Despite the rich opportunities for interpersonal synchrony provided by the immediacy and multi-dimensionality of physical presence [45], video calls still can facilitate a certain degree of interpersonal coordination through adapted forms of facial expressions and upper-body cues [34]. Mui et al.[33] explored smile mimicry in dyadic VC interactions and found that participants mimicked more smiles during interactions with a smiling confederate than participants who interacted with the unexpressive confederate. Their results indicate the presence of smile mimicry in video-mediation communication. Additionally, Gvirts et al. [16] investigated body movement synchrony and emotional alignment in dyadic VC, finding that higher synchrony levels were specifically linked to increased positive emotional alignment and liking.

Whereas previous works like [16, 33] have verified the presence of interpersonal coordination of body movement and facial mimicry in VC interactions and underscored its importance for the impact on the quality of social interaction, gaps remain in how technical impairments affect it.

2.2 Nonverbal Cues in Videoconferencing Communication

In VC, the dynamics of nonverbal cues undergo significant transformations due to the inherent limitations of the medium [2]. For instance, Scott [48] revealed that subtle facial expressions become less perceptible due to the reduced video frame rates. Similarly, Horn [20, 21] explored the impact of spatial and temporal video distortion on nonverbal cues and showed that such impairments can affect the precision with which viewers perceive and interpret subtle facial expression cues and body language, leading to potential misunderstandings.

Eye gaze is a significant cue of nonverbal communication that helps convey attention, however, it is often inhibited in VC. Techniques to enhance gaze perception in VC, such as image adjustments to simulate direct eye contact have been developed however their performance varies [43, 55]. Brucks and Levav [7] further observed that individuals in video calls tend to focus more narrowly on conversational partners compared to F2F interactions, influencing how attention is distributed across communication modes.

Head and body movements, such as nodding and leaning forward, indicate agreement or willingness to contribute without interruption in many cultures [25]. These movements are essential for managing conversation flow and turn-taking [18, 22]. However, Schoenenberg et al. [47] revealed that transmission delay can lead to exaggerated body motion responses. Trujillo et al. [52] observed a significant impact of video quality on body motion and reported a U-shaped relationship between video blur grade and body motion, indicating that body motion first decreased as the video blurring set in and then increased again during the stronger blur grades.

Most research has concentrated on identifying sensitive behaviors and developing enhancements without fully addressing how reductions in video quality can alter user nonverbal behavior. This study aims to fill this gap by evaluating how video resolution and the resulting conversational quality influence nonverbal cues, including facial expressions, gaze behavior, and body movements.

3 USER STUDY

This study is a secondary data analysis of a research project that explored the impact of network impairments on the user experience of VC [12]. We carried out a within-subjects experiment with three video resolutions (240p, 480p, and 1080p). In additional test conditions, pure transmission delay and combinations of resolution- and delay-specific settings were also used. The current paper is self-contained and focuses on the conditions where "resolution" was the sole technical factor being specifically set. User-perceived conversational quality was measured using a post-task questionnaire adapted from [24]. The data used for extracting nonverbal cues was the webcam recordings of the conversation sessions.

3.1 Hypotheses

This study focuses on two types of interpersonal coordination that are particularly relevant to the quality of communication and are available through VC interaction, namely, facial expression synchrony and body movement synchrony. Previous research such as Mui et al.[33] and Gvirts et al.[16] provided evidence of smile

mimicry and body movement synchrony in VC interactions. Building on this evidence, we hypothesize:

Hypothesis 1 (H1): The participant's body movements and facial expressions will lead to synchrony during VC interactions.

According to research in [20, 21], facial cues and body language become less perceptible due to video distortion. Moreover, it was shown in [52] that high video quality leads to increased body motion compared to some blurry video settings. Studies like those in [27, 53, 54] highlighted that effective conversational experience generally includes a deeper level of coordinated interactions among participants (e.g., empathetic responses). The reviewed studies indicated that both high video quality and good conversational quality are essential for enhancing the perception and interpretation of nonverbal cues. It encourages greater physical responsiveness among participants and actively contributes to a higher level of interpersonal coordination. Therefore, we hypothesize:

Hypothesis 2 (H2): Dyads engaged in VC with high video resolution and conversational quality, compared to those with low resolution and conversational quality, will exhibit more body movement, more facial expressions, and an increased gaze at the conversational partner.

Hypothesis 3 (H3): Dyads engaged in VC with high video resolution and conversational quality, compared to those with low resolution and conversational quality, will demonstrate a higher level of body movement synchrony and facial expression synchrony.

3.2 Participants

A total of 46 subjects (23 dyads) participated, with ages ranging from 22 to 36 years ($M = 27.04, SD = 3.58$). The gender distribution was 27 females and 19 males. All participants reported normal vision and hearing and were compensated with €12 for their participation. We excluded 3 dyads as one participant in the pair was exposed to the study design, leaving 20 pairs for behavior analysis. The study was pre-approved by the Ethical Committee of the university and executed following the guidelines of the national research organization and the declaration of Helsinki.

3.3 Task

Participants played a Celebrity Name Guessing (CNG) game, chosen for its robust social interaction component, making it ideal for assessing conversational quality in a controlled yet naturalistic setting [23, 24]. This task involves guessing the name of a celebrity chosen by the conversation partner by asking "yes" or "no" questions. As long as the answer is "yes," participants can continue with additional questions. If the answer is "no", it is the other participant's turn to ask the question. If the celebrity is guessed correctly, the participant chooses a new card for the partner to guess and continues playing. Each game lasted approximately three minutes.

3.4 Procedure

The experiment started with a detailed briefing on the task and procedure without revealing the specific conditions. Participants signed informed consent forms after understanding the study's scope and their rights. A training session was conducted in person to familiarize participants with the task mechanics and the VC system under a reference condition (1080p). After the training session, participants performed all trials, one for each test condition. Each trial lasted three minutes, and each participant was asked to choose up to two celebrity cards for the partner to guess. A short 50-second break after each trial was provided, allowing participants to rate the conversational quality ("How would you rate the overall conversational quality of the call?") on a 5-point Absolute Category Rating scale (1 = bad, 5 = excellent).

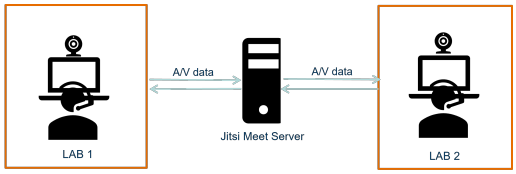


Figure 1: Example of the test deployment.

3.5 Apparatus

As shown in Figure 1, our setup included two lab rooms meeting the specifications in [24], equipped with fluorescent lighting (D65, ie. 6500K) and window shades to keep out external light. We used two identical sets of hardware components, including Linux-based laptops (Ubuntu 20.04 LTS), Logitech BRIO 4K cameras, MOTU M4 audio interfaces, LG 27" UHD 27UL850 monitors, and Beyerdynamic DT290 headsets. All the host connections were via a well-provisioned (1000 Mbps) wired LAN. A self-hosted instance of Jitsi Meet [1] was used as our VC platform. VP9 was used as the video codec, with the framerate set to 30 fps. Google Chrome was used as the browser to access Jitsi Meet and to support the VP9 video codec. Video resolution changes were implemented by modifying the Jitsi Meet media configuration file. During the experiment, participants could see the head and shoulders of their partners at all times. We gathered the real-time RTP and media stream statistic data via the WebRTC internals page [2] to verify the changes in video resolution and monitor the service (e.g., network delay). Moreover, we conducted 100 measurements of the base system latency, revealing an average latency of 14 ms, with a maximum of 25 ms and a minimum of 5 ms. For the recording, we used the V4L2 loopback kernel module to create virtual loopback devices that receive media streams directly from the webcam (1080p, 30fps) and used FFmpeg [3] to capture and encode audio in raw MP4 format with lossless compression. We truncated the video recordings to include only interactions occurring during the 3-minute conversation phase.

3.6 Nonverbal Cues Extraction

3.6.1 Body Movement. In prior studies exploring bodily movement synchrony in F2F conversations, such as Tschacher et al. [53] and Ramseyer and Tschacher [41], the frame-differencing measurement technique has been effectively used to describe body movements. This frame-differencing method calculates changes in grayscale pixel density between consecutive video frames, which could objectively quantify the intensity of motion energy [35]. Given our controlled laboratory setting with a stable camera, background, and

[1] https://github.com/jitsi/jitsi-meet
[2] chrome://webrtc-internals
[3] https://ffmpeg.org/

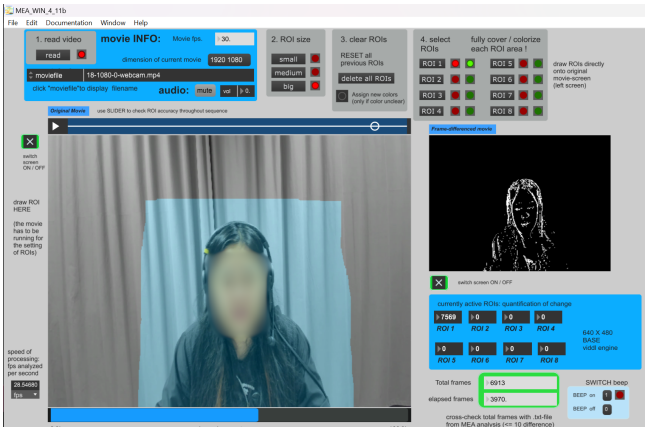


Figure 2: A screenshot of the MEA software analyzing a webcam recording of the data we collected. A predefined region of interest (ROI) captured the upper body of the participant.

lighting, changes detected between frames are primarily attributable to participant movements, we employed Motion Energy Analysis (MEA) software (version 4.11b) [42] to capture body motion from video recordings. MEA uses the frame-differencing method and features a user-friendly interface that allows for precise regions of interest (ROI) delineation to focus on relevant movement areas [42]. We selected the ROI to encompass the upper body of each participant (see Figure 2) to optimize motion capture. This approach generated a time series of 5400 data points per video, calculated based on the video's duration and frame rate (3 minutes, 30 fps). For individual-level activity analysis, the body motion energy intensities were averaged across all frames for each video to provide a consistent measure of participant body movement activity. Body movement synchrony of the dyad was measured as the temporal relationship between the body motion energy time series of each participant.

3.6.2 Facial Expression. We utilized the OpenFace 2.0 software [3] to detect the movement of facial muscles, known as Action Units (AUs), from each video frame. The AUs are based on the Facial Action Coding System (FACS) [14]. To access facial expression synchrony, we focused on AUs associated with the "Duchenne smile" [13] integral to examining smile mimicry phenomena [33]. According to FACS, the "Duchenne smile" involves AU6 (cheek raiser) and AU12 (lip corner puller). We extracted the activation intensity (on a scale from 0 to 5, with 0 indicating no activity and 5 representing maximum muscle activation) for AU06 and AU12 in each frame, resulting in a time series of 5400 points for each AU per video. Facial expression synchrony was calculated as the temporal relationship between the coded facial expression time series of each participant in a dyad. At the individual level, the intensity values for each AU were then averaged across frames for each video to create a composite score reflecting the overall activity of AU06 and Au12 during the interaction.

3.6.3 Gaze Behavior. To assess the gaze behavior of dyads during the conversation, we defined three regions of interest (ROIs) for each play trial: the screen (i.e., the conversational partner), the

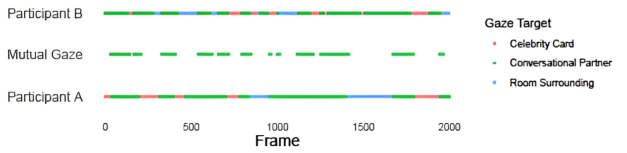


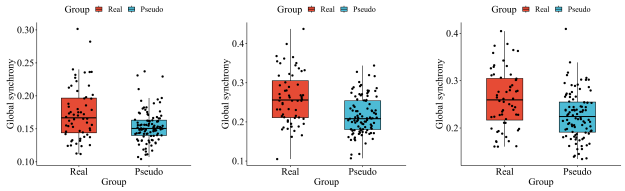
Figure 3: Example of a coupled gaze ROI stream (showing only 2000 data points) from a dyad, each color representing a different ROI. Mutual gaze was measured when two participants looked at the conversational partner simultaneously.

celebrity card, and the room surroundings. These ROIs, used similarly in the study by Brucks and Levav [7], help us understand focus distribution during interactions. Gaze data were annotated using ELAN 6.7 [15], a tool supporting frame-by-frame annotation of video. A coder blinded to our hypotheses manually coded these ROIs frame-by-frame. Each dyad's interaction resulted in two streams of gaze data across the three ROIs. Figure 3 shows a representative example of the raw fixation ROI streams from one dyad.

At the individual level, we analyzed: the total looking time directed at each ROI; the number of fixations (frequency) on each ROI; and the average duration of each fixation on an ROI. At the interpersonal level, we focused on mutual gaze-like behavior, which we defined as occurring when both participants look at the screen simultaneously. We quantified this behavior by measuring: the total duration of mutual gaze; the count of mutual gaze; and the average duration of each mutual gaze.

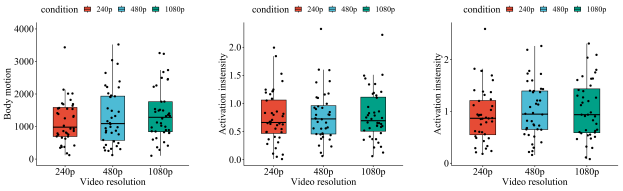
3.6.4 Synchrony Quantification. In our work, we defined interpersonal coordination as the temporal relationship between the behaviors of each participant in a dyad. To quantify it, we employed windowed and lagged cross-correlation (WLCC) [6], using the rMEA package [42] to calculate the cross-correlation table for the body movement and facial expression time series. Initially, the raw data were pre-processed following the steps in [42]. Firstly, the data were averaged over 0.5-second windows to reduce fluctuations from signal distortion across the videos. Subsequently, the data were standardized using the standard deviation as the scaling factor. Artifacts and outlier data points—missing values or values exceeding 10 times the standard deviation of each time series—were removed. For the WLCC settings, we selected a 30-second window based on the shorter turn-taking duration observed in our task compared to the longer duration in psychotherapy contexts [30]. We used a maximum lag of ±5 seconds to capture synchronous movements with delayed onset following the setting in [16] and calculated the absolute values with Fisher's z transform of the WLCC.

We validated the nonverbal synchrony measured in our study by employing a pseudo-synchrony approach [39], where N=100 pseudo dyads were generated for each nonverbal behavior by randomly matching the time series of Participant A with Participant B from a different conversation session. This approach assesses the likelihood that observed synchrony is due to chance rather than genuine interaction effects. Then we used the global synchrony



(a) Body movement. (b) AU06. (c) AU12.

Figure 4: Box plot of the distribution of the global synchrony grouped by real and pseudo dyads.



(a) Body movement. (b) AU06. (c) AU12.

Figure 5: Box plot of the distribution of the individual-level nonverbal behaviors grouped by video resolution.

score, quantified as the grand average of WLCC, to compare the strength of synchronization between the real dyads and pseudo dyads.

To analyze the dynamics of interpersonal synchrony, we extracted various measures from the cross-correlation table. We calculated the global synchrony (the grand average of the entire cross-correlation table), representing the average strength of synchronized movement throughout the interaction. Lag zero synchrony, representing near-simultaneous shared activity, was measured from non-lagged cross-correlation values. Additionally, we extracted peak synchrony, i.e., the maximum correlation value from the cross-correlation table, representing the highest degree of synchronization between participants, and peak lag, i.e., the time delay at which peak synchrony occurs.

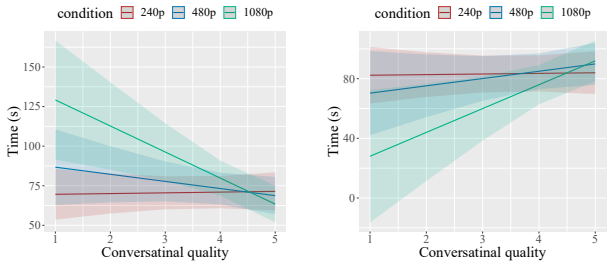
4 RESULTS

4.1 Comparing with Pseudo Dyads

The Welch Two Sample t-test was performed to determine whether there were significant differences in the global synchrony between real and pseudo dyads. Cohen's d was calculated to quantify the effect size of the differences and interpreted according to Cohen's (1988) benchmarks [9].

4.1.1 Body Movement Synchrony. The distribution of the global synchrony for the real dyads versus the pseudo dyads is presented in Figure 4a. A significant difference was found ($t(85.1) = 3.48, p < .001, d = .64$), with real dyads showing higher global synchrony ($M = .17, SD = .04$) than pseudo dyads ($M = .15, SD = .02$).

4.1.2 Facial Expression Synchrony. Figure 4b and Figure 4c illustrate the global synchrony level of AU06 and AU12 activity for real dyads and pseudo dyads. Our results yielded a significant difference in the global synchrony level between real dyads and pseudo dyads with a large effect size ($t(96.55) = 4.59, p < .0001, d = .81$). This suggests that the global synchrony of the AU06 activity was significantly higher in real dyads ($M = .26, SD = .07$) compared to pseudo dyads ($M = .22, SD = .05$). For the AU12 activity, the global synchrony in real dyads ($M = .26, SD = .06$) was significantly higher than in pseudo dyads ($M = .23, SD = .05$; $t(103.71) = 3.69, p < .001, d = .64$).



(a) Time looking-partner. (b) Time looking-card.

Figure 6: Estimated regression line plots from the LMMs, with mean and standard error of the time participants spend a) looking at their conversational partner and b) celebrity cards grouped by video resolution and conversational quality.

4.2 Individual-level Characteristics

We conducted separate Linear Mixed-effects Models (LMMs) to quantify the effects of video resolution and conversational quality on nonverbal behavior at the individual level. Each model treated the nonverbal cue measures as the dependent variable and included perceived conversational quality (treated as a continuous variable) and video resolution (treated as a categorical variable with three levels: 240p, 480p, and 1080p) as fixed effects and participants as a random intercept. The evaluation of model significance was calculated using Satterthwaite's method to estimate degrees of freedom and p-values. All the analysis was performed in R 4.3.3 [37] using *lme4* [4] and *lmerTest* [28].

4.2.1 Body Movement. Figure 5a shows the average body motion grouped by the video resolution. The results indicated significant main effects of video resolution on body movement patterns. There was a significant increase in body movement for the 1080p condition compared to the baseline (240p condition) ($\beta = 1425.10, SE = 663.85, t(86.89) = 2.15, p = .034$). The increase body movement observed in 480p did not reach statistical significance ($\beta = 202.34, SE = 443.65, t(82.59) = .46, p = .65$). Regarding conversational quality, we did not find a significant independent effect on body movement ($p = .544$). However, a decreasing

trend was observed at the 1080p condition ($\beta = -273.96, SE = 149.85, t(85.43) = -1.83, p = .071$).

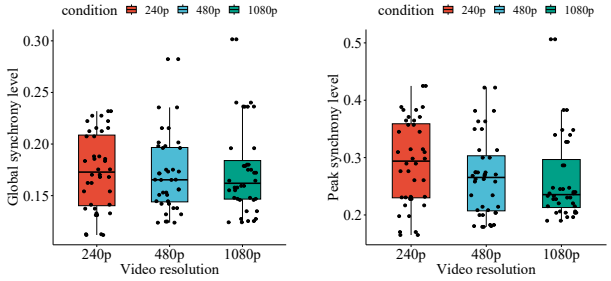
4.2.2 Facial Expression. Figure 5b and Figure 5c illustrate the activation intensity of AU06 and AU12 grouped by video resolution. Our findings revealed no significant main effects of video resolution or conversational quality on AU06 and AU12 activity. However, there was a trend towards increased AU06 activity in the 1080p condition compared to the baseline 240p condition ($\beta = .54, SE = .29, t(82.5) = 1.88, p = .064$). The 480p condition did not significantly differ from the baseline ($p = .705$). The interaction between video resolution (1080p) and conversational quality also showed a marginal negative effect on AU06 activity ($\beta = -.11, SE = .07, t(81.61) = -1.73, p = .087$).

4.2.3 Gaze Behavior. Figure 6a and Figure 6b show the estimated values of the time participants spend looking at their conversational partner and celebrity card as a function of conversational quality and video resolution. Our results revealed significant main effects of video resolution on gaze behavior. Participants in the 1080p condition spent significantly more time looking at their conversational partner ($\beta = 76.51, SE = 23.44, t(85.26) = 3.26, p = .002$) compared to the baseline 240p condition. Conversely, they spent significantly less time looking at the celebrity card ($\beta = -69.87, SE = 27.62, t(85) = -2.53, p = .013$). While conversational quality alone did not significantly affect gaze behaviors towards the partner or the card, a notable interaction effect was observed with video resolution. At the 1080p condition, higher conversational quality was associated with an increase in the time spent gazing at celebrity cards ($\beta = 15.57, SE = 6.23, t(83.83) = 2.5, p = .014$) and a corresponding decrease in time spent gazing at the conversational partner ($\beta = -16.91, SE = 5.29, t(84.08) = -3.20, p = .002$).

4.3 Dyadic-level Characteristics

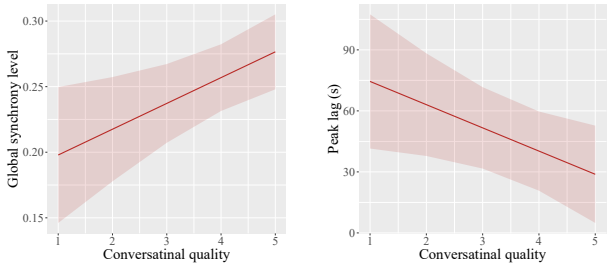
The primary goal of this analysis was to investigate how video resolution and conversational quality influence mutual gaze behavior, body movement synchrony, and facial expression synchrony. We used separate LMMs to quantify the effects of video resolution and conversational quality on mutual gaze and interpersonal synchrony. In these LMMs, mutual gaze behavior and interpersonal synchrony were treated as dependent variables, and conversational quality and video resolution were treated as fixed effects. Additionally, we conducted an exploratory analysis using separate LMMs to investigate the relationship between mutual gaze behavior and interpersonal synchrony. In these models, mutual gaze behavior was treated as a fixed effect to determine if it mediated changes in interpersonal synchrony. All LMMs included the dyad as a random effect.

4.3.1 Body Movement Synchrony. Figure 7 demonstrates the global and peak synchrony of body movements grouped by the video resolution. We found significant main effects of video resolution on global and peak synchrony. A significant negative effect was found on global synchrony in the 480p condition compared with the baseline (240p) condition ($\beta = -.04, SE = .02, t(96.91) = -2.21, p = .029$). Peak synchrony also showed significant reductions for both 480p and 1080p conditions compared to 240p condition ($\beta = -.03, SE = .01, t(97.61) = -2.04, p = .045$ for 480p; $\beta = -.03, SE = .01, t(98.85) = -2.19, p = .031$ for 1080p).



(a) Global body movement synchrony. (b) Peak body movement synchrony.

Figure 7: Box plot of a) global synchrony and b) peak synchrony levels of body movement grouped by video resolution.

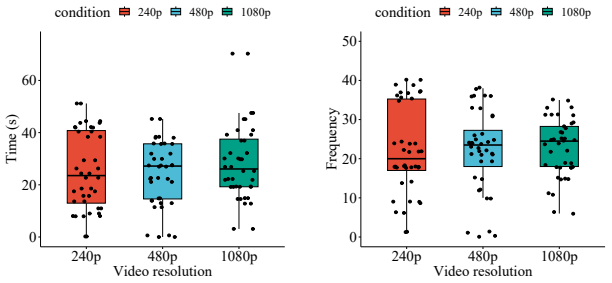


(a) Global synchrony of AU12. **(b) Peak lag of AU06.**

Figure 8: Estimated regression line plots from the LMMs, with mean and standard error of the a) global synchrony of AU12 and b) peak lag of AU06 grouped by conversational quality.

4.3.2 Facial Expression Synchrony. We found no significant main effects of video resolution on facial expression synchrony. Conversational quality showed a significant impact on the synchrony and timing of facial expression reactions (see Figure 8). Specifically, higher conversational quality significantly increased the global synchrony of AU12 ($\beta = .02, SE = .01, t(101.12) = 2.67, p = .009$) and reduced the peak lag of AU06 ($\beta = -11.41, SE = 5.23, t(103.95) = -2.18, p = .031$).

4.3.3 Mutual Gaze. Figure 9a and Figure 9b show the time and frequency of mutual gaze grouped by video resolution. Significant main effects of video resolution on mutual gaze-like behaviors were observed. In the 1080p condition, participants engaged in mutual gaze for a longer time ($\beta = 34.85, SE = 16.46, t(98.3) = 2.12, p = .037$) and with greater frequency ($\beta = 16.88, SE = 7.54, t(97.76) = 2.24, p = .028$) than in the 240p condition. However, high video resolution (1080p) coupled with greater conversational quality led to a significant reduction in both mutual gaze time ($\beta = -7.55, SE =$



(a) Time of mutual gaze. **(b) Frequency of mutual gaze.**

Figure 9: Box plot of a) time of mutual gaze and b) frequency of mutual gaze grouped by video resolution.

$3.74, t(98.03) = -2.02, p = .046$) and frequency ($\beta = -3.77, SE = 1.71, t(95.53) = -2.2, p = .03$).

Additionally, our exploratory analysis results show that the overall time and frequency of mutual gaze positively influenced the synchrony of body movements, although the effect sizes are small. Significant positive effects were noted on lag zero synchrony of body movement for overall mutual gaze time ($p = .048$) and frequency ($p = .003$). Furthermore, mutual gaze significantly enhanced the synchrony for AU06 and AU12. Specifically, the positive effects of mutual gaze time on lag zero synchrony were statistically significant ($p = .004$ for AU06; $p = .014$ for AU12). More time of mutual gaze significantly reduced peak lags for both AU06 ($\beta = -.84, SE = .26, t(98.1) = -3.2, p = .002$) and AU12 ($\beta = -0.76, SE = 0.25, t(112.31) = -3.01, p = .003$). Additionally, longer fixation duration during mutual gaze was associated with reduced peak lag for AU06 ($\beta = -5.24, SE = 2.72, t(116.98) = -1.93, p = .056$) and AU12 ($\beta = -6.68, SE = 2.5, t(114.435) = -2.67, p = .009$).

5 DISCUSSION

5.1 Interpersonal Coordination in Video-mediated Communication

Body movement synchrony as measured using WLCC, has been described and validated in the literature for F2F [30, 53] and VC [16] interactions. For example, the work in [53] has reported that body movement synchrony in F2F interactions can vary significantly depending on the nature of the conversational task, which ranged from .17 in cooperative tasks to .2 in more engaging and enjoyable tasks like fun activities. Lin et al. [30] explored body movement synchrony among friend dyads as they engaged in problem-focused F2F discussions and reported a mean global synchrony value of .16 ($SD = .02$). Gvirts et al. [16] reported a lower body movement synchrony level ($M = .14, SD = .04$) in VC interaction between strangers when sharing personal difficulties.

Consistent with our hypothesis (H1), the global synchrony of facial expression and body movement for real dyads was significantly higher than in pseudo-interactions. This is in line with the findings in previous research by Gvirts et al. [16]. Our finding suggests the existence of interpersonal synchrony during a video call, with the value of global synchrony being .17 for body movements and .26 for facial expressions. This global synchrony value of body movements was comparable to F2F settings reported in [53] and [30], indicating that VC can achieve a similar level of interpersonal synchrony compared to F2F. Moreover, we found that both AU06 and AU12 showed a higher average global synchrony than body movement, indicating a stronger presence of synchrony through facial expressions compared to body movements. Additionally, we observed medium to large effect sizes from the analysis, with the effect size for body movement and AU12 was $d = .64$, and for AU06, it was $d = .81$. These effect sizes are comparable to those reported by Tschacher et al. [53], ranging from $d = .56$ to $d = 1.11$ in unacquainted dyads engaged in F2F social tasks. These findings emphasize the practical significance of our results and provide objective evidence of genuine interpersonal coordination during dyadic VC interactions when playing a CNG game.

5.2 The Role of Video Resolution

In this work, we investigated the effect of video resolution (with three levels: 240p, 480p, and 1080p) on nonverbal behaviors during VC interactions.

Our results support the hypothesis (**H2**) that high video resolution enhances individual nonverbal responsiveness. This was evidenced by increased body movement and AU06 activity at 1080p compared to 240p, indicating improved visibility and responsiveness of body movement and facial expressions at high video resolution. Our results align with the findings in [52], suggesting high video quality would lead to greater body motion. Moreover, we found that participants may spend more time looking at their conversational partner and less time checking their celebrity card at high (1080p) resolution than at low (240p) resolution. Consistently, participants in the 1080p setting engaged more in mutual gaze. These findings suggest a more engaged interaction experience between participants at high video resolution than at low video resolution during video calls. Interestingly, these nonverbal behaviors in 480p did not show significant changes compared with the 240p condition, suggesting that intermediate improvements in video resolution may not substantially influence the user's nonverbal behavior.

Contrary to our expectations (**H3**), higher video resolution did not significantly improve interpersonal coordination. Instead, bodily synchronization decreased at higher video resolutions compared to the low (240p) resolution. This may be due to increased visual noise at low resolution enhancing bodily synchronization, similar to findings in F2F interactions [26, 36], where visual occlusion can inadvertently enhance interpersonal coordination. Paxton and Dale [36] manipulated visual occlusion stimuli by asking participants to wear special glasses and adapting flashing screens on the glass during the dyadic conversation. Their results suggest that perceiving visual noise makes communication more difficult but may enhance physical coordination. Kodama et al. [26] found that occluding visual information increased head-movement coordination between participants by comparing visible (both visual and auditory information were available) and invisible (only auditory information) conditions. In contrast to physical visual occlusion, in a VC setting, low video resolution presented more visual noise on the display (e.g., jagged and blurred details) compared to high video resolution. Extending the findings in [26, 36], our results suggest

that a VC system with low video resolution may inadvertently boost bodily coordination.

5.3 The Role of Conversational Quality

We initially hypothesized that greater conversational quality would enhance individual nonverbal activity and interpersonal synchrony in VC interactions.

Contrary to our hypothesis (**H2**), we did not find significant main effects of conversational quality on individual behaviors. However, in the high (1080p) video resolution condition, high conversational quality was correlated with reduced body movement, AU06 activity, and decreased eye gaze at the conversational partner. This result suggests that high conversational quality may lead to reduced nonverbal expressiveness in high-resolution VC. A possible explanation would be that high-quality conversations are typically characterized by increased verbal fluency and interpersonal understanding, which might lessen the need for compensatory nonverbal cues such as exaggerated body movements or facial expressions. Daly-Jones et al. [11] investigated the impact of visual cues on conversational fluency and interpersonal awareness by comparing VC and audio-only conferencing. They found that incorporating visual cues had the effect of leading to more fluent conversations especially with larger group numbers, however, this was not so pronounced in dyadic conversation where auditory cues to turn-taking were generally sufficient. This aligns with our observation of reduced nonverbal expressiveness with higher conversational quality in dyadic VC interaction.

Partially confirming our hypothesis (**H3**), we noted a significant enhancement in the global synchrony of AU12 activity and quicker responsiveness of AU06 activity with higher conversational quality. This suggested that higher conversational quality likely foments a more emotionally congruent environment, enabling synchronized facial expressions of positive emotion, like those observed in our study. Studies that approached results like ours include Suzuki et al. [50]. They developed a VC system that artificially generates facial expressions so that participants seem to mimic smiling, which improves both the quality and emotional satisfaction of remote conversations.

6 LIMITATIONS AND FUTURE WORK

Firstly, our result was limited to the conversational task we employed. The CNG game is a well-structured and task-oriented conversational scenario that may not fully capture the nonverbal behavior dynamics as those in more natural and complex conversations. Future research should employ a variety of free conversational tasks (see Tschacher et al. [53]) to validate and possibly extend our findings. Secondly, we only focused on Duchenne smile-related facial action units and did not evaluate other facial action units that related to negative expressions or low-occurrence expressions (see Tomprou et al. [51]). Future studies could test our findings on these facial expressions. Thirdly, the eye gaze behavior was mutually annotated by a single annotator in our work, and more advanced techniques such as eye trackers can be used in future studies to enhance the accuracy. Moreover, in this work, participants' personality traits were not collected and assessed. However, some key personality traits such as agreeableness and extroversion, according to the work by Arellano-Véliz et al. [1], could significantly influence bodily synchrony. Moreover, we did not consider the impact of base system latency on the measurement of interpersonal coordination. Because the measured maximum base system latency was 25ms, which is less than one frame, and our measurements of user behavior were frame-based, therefore we did not make adjustments for this factor.

7 CONCLUSION

In this work, we explored nonverbal dynamics among dyads as they played a CNG game using a VC platform, focusing on investigating how video resolution and conversational quality impact the user's nonverbal behaviors at the individual and interpersonal levels. We explored several nonverbal cues that can be extracted from webcam recordings, including individual nonverbal activities like body motion, facial expressions, gaze directions, and interpersonal coordination such as facial expression and body movement synchrony. Our analysis provides evidence of the presence of interpersonal coordination during VC interactions. Our results indicate that high video resolution enhances body movements and facial expressions, leading to richer interactions. We found that high video resolution can promote the gaze behavior of looking at the conversational partner and reduce the distraction of attention elsewhere, thereby creating more gaze behaviors similar to mutual gaze. Our results suggest that high video resolution significantly decrease the body movement synchrony, however, participants still exhibited strong facial expression synchrony with different levels of video resolution. Additionally, the result shows that the conversational quality and mutual gaze significantly impacted interpersonal synchrony. Higher conversational quality was associated with shorter peak lags of facial expression synchrony and higher global synchrony of facial expressions. Furthermore, we found that increased mutual gaze was associated with higher lag-zero synchrony of body movement and facial expressions. Our findings could be applied to the VC setups to enhance interpersonal coordination and engagement.

ACKNOWLEDGMENTS

This research is funded by the Carl-Zeiss-Foundation ("Breakthroughs 2020" program, https://www.carl-zeiss-stiftung.de/programm/czs-durchbrueche), in the CO-HUMANICS project. The author acknowledges that ChatGPT (https://chat.openai.com) was used in the drafting of this paper for grammatical correctness and clarity. We ensured that the manuscript thoroughly underwent human review and revision.

REFERENCES

[1] Nicol A Arellano-Véliz, Bertus F Jeronimus, E Saskia Kunnen, and Ralf FA Cox. 2024. The interacting partner as the immediate environment: Personality, interpersonal dynamics, and bodily synchronization. *Journal of Personality* 92, 1 (2024), 180–201.

[2] Jeremy N Bailenson. 2021. Nonverbal overload: A theoretical argument for the causes of Zoom fatigue. (2021).

[3] Tadas Baltrusaitis, Amir Zadeh, Yao Chong Lim, and Louis-Philippe Morency. 2018. OpenFace 2.0: Facial Behavior Analysis Toolkit. In *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*. 59–66. https://doi.org/10.1109/FG.2018.00019

[4] Douglas Bates, Martin Maechler, Ben Bolker, Steven Walker, Rune Haubo Bojesen Christensen, Henrik Singmann, Bin Dai, Gabor Grothendieck, Peter Green, and M Ben Bolker. 2015. Package 'lme4'. *convergence* 12, 1 (2015), 2.

[5] Frank J Bernieri and Robert Rosenthal. 1991. Interpersonal coordination: Behavior matching and interactional synchrony. (1991).

[6] Steven M Boker, Jennifer L Rotondo, Minquan Xu, and Kadijah King. 2002. Windowed cross-correlation and peak picking for the analysis of variability in the association between behavioral time series. *Psychological methods* 7, 3 (2002), 338.

[7] Melanie S Brucks and Jonathan Levav. 2022. Virtual communication curbs creative idea generation. *Nature* 605, 7908 (2022), 108–112.

[8] Judee K Burgoon, Laura K Guerrero, and Valerie Manusov. 2011. Nonverbal signals. *Handbook of interpersonal communication* (2011), 239–280.

[9] Jacob Cohen. 2013. *Statistical power analysis for the behavioral sciences*. Routledge.

[10] Rick Dale, Gregory A Bryant, Joseph H Manson, and Matthew M Gervais. 2020. Body synchrony in triadic interaction. *Royal Society open science* 7, 9 (2020), 200095.

[11] Owen Daly-Jones, Andrew Monk, and Leon Watts. 1998. Some advantages of video conferencing over high-quality audio conferencing: fluency and awareness of attentional focus. *International Journal of Human-Computer Studies* 49, 1 (1998), 21–58.

[12] Chenyao Diao, Luljeta Sinani, Rakesh Rao Ramachandra Rao, and Alexander Raake. 2023. Revisiting videoconferencing qoe: Impact of network delay and resolution as factors for social cue perceptibility. In *2023 15th International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 240–243.

[13] Paul Ekman, Richard J Davidson, and Wallace V Friesen. 1990. The Duchenne smile: Emotional expression and brain physiology: II. *Journal of personality and social psychology* 58, 2 (1990), 342.

[14] Paul Ekman and Wallace V Friesen. 1978. Facial action coding system. *Environmental Psychology & Nonverbal Behavior* (1978).

[15] Version ELAN. 2023. 5.9 [Computer Software] Nijmegen: Max Planck Institute for Psycholinguistics. *The Language Archive* (2023).

[16] Hila Gvirts, Lya Ehrenfeld, Mini Sharma, and Moran Mizrahi. 2023. Virtual social interactions during the COVID-19 pandemic: the effect of interpersonal motor synchrony on social interactions in the virtual space. *Scientific Reports* 13, 1 (2023), 10481.

[17] Jingjing Han, Lucía Cores-Sarría, and Han Zhou. 2024. In-person, video conference, or audio conference? Examining individual and dyadic information processing as a function of communication system. *Journal of Communication* (2024), jqae003.

[18] Jinni A Harrigan. 1985. Listeners' body movements and speaking turns. *Communication Research* 12, 2 (1985), 233–250.

[19] Ursula Hess and Sylvie Blairy. 2001. Facial mimicry and emotional contagion to dynamic emotional facial expressions and their influence on decoding accuracy. *International journal of psychophysiology* 40, 2 (2001), 129–141.

[20] Daniel B Horn. 2001. Is seeing believing? Detecting deception in technologically mediated communication. In *CHI'01 Extended Abstracts on Human Factors in Computing Systems*. 297–298.

[21] Daniel B Horn, Lana Karasik, and Judith S Olsen. 2002. The effects of spatial and temporal video distortion on lie detection performance. In *CHI'02 Extended Abstracts on Human Factors in Computing Systems*. 714–715.

[22] Carlos Toshinori Ishi, Hiroshi Ishiguro, and Norihiro Hagita. 2014. Analysis of relationship between head motion events and speech in dialogue conversations. *Speech Communication* 57 (2014), 233–243.

[23] ITU-T Recommendation. P.1305. 2016. Effect of delays on telemeeting quality. https://www.itu.int/rec/T-REC-P.1305-201607-I/en

[24] ITU-T Recommendation. P.920. 2000. Interactive test methods for audiovisual communications. https://www.itu.int/rec/T-REC-P.920/en

[25] Richard G Jones. 2013. *Communication in the real world: An introduction to communication studies*. Flat World Knowledge.

[26] Kentaro Kodama, Daichi Shimizu, and Ken Fujiwara. 2022. Influence of Visual Information on Interpersonal Coordination of Head-and Body-Movement During Dyad Conversations. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, Vol. 44.

[27] Wojciech Marek Kulesza, Aleksandra Cisłak, Robin R Vallacher, Andrzej Nowak, Martyna Czekiel, and Sylwia Bedynska. 2015. The face of the chameleon: The experience of facial mimicry for the mimicker and the mimickee. *The Journal of social psychology* 155, 6 (2015), 590–604.

[28] Alexandra Kuznetsova, Per B Brockhoff, and Rune Haubo Bojesen Christensen. 2017. lmerTest package: tests in linear mixed effects models. *Journal of statistical software* 82, 13 (2017).

[29] Juha M Lahnakoski, Paul AG Forbes, Cade McCall, and Leonhard Schilbach. 2020. Unobtrusive tracking of interpersonal orienting and distance predicts the subjective quality of social interactions. *Royal Society Open Science* 7, 8 (2020), 191815.

[30] Lisa Lin, Mallory J Feldman, Ashley Tudder, Abriana M Gresham, Brett J Peters, and David Dodell-Feder. 2023. Friends in Sync? Examining the Relationship Between the Degree of Nonverbal Synchrony, Friendship Satisfaction and Support. *Journal of Nonverbal Behavior* 47, 3 (2023), 361–384.

[31] Daniel N McIntosh. 2006. Spontaneous facial mimicry, liking and emotional contagion. *Polish Psychological Bulletin* 37, 1 (2006), 31.

[32] Dan Mønster, Dorthe Døjbak Håkonsson, Jacob Kjær Eskildsen, and Sebastian Wallot. 2016. Physiological evidence of interpersonal dynamics in a cooperative production task. *Physiology & behavior* 156 (2016), 24–34.

[33] Phoebe HC Mui, Martijn B Goudbeek, Camiel Roex, Wout Spierts, and Marc GJ Swerts. 2018. Smile mimicry and emotional contagion in audio-visual computer-mediated communication. *Frontiers in Psychology* 9 (2018), 411451.

[34] David T Nguyen and John Canny. 2009. More than face-to-face: empathy effects of video framing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 423–432.

[35] Alexandra Paxton and Rick Dale. 2013. Frame-differencing methods for measuring bodily synchrony in conversation. *Behavior research methods* 45 (2013), 329–343.

[36] Alexandra Paxton and Rick Dale. 2017. Interpersonal movement synchrony responds to high-and low-level conversational constraints. *Frontiers in psychology* 8 (2017), 235014.

[37] R R Core Team et al. 2013. R: A language and environment for statistical computing. (2013).

[38] Alexander Raake, Markus Fiedler, Katrin Schoenenberg, Katrien De Moor, and Nicola Döring. 2022. Technological factors influencing videoconferencing and zoom fatigue. *arXiv preprint arXiv:2202.01740* (2022).

[39] Fabian Ramseyer and Wolfgang Tschacher. 2010. Nonverbal synchrony or random coincidence? How to tell the difference. *Development of Multimodal Interfaces: Active Listening and Synchrony: Second COST 2102 International Training School, Dublin, Ireland, March 23-27, 2009, Revised Selected Papers* (2010), 182–196.

[40] Fabian Ramseyer and Wolfgang Tschacher. 2011. Nonverbal synchrony in psychotherapy: coordinated body movement reflects relationship quality and outcome. *Journal of consulting and clinical psychology* 79, 3 (2011), 284.

[41] Fabian Ramseyer and Wolfgang Tschacher. 2014. Nonverbal synchrony of head- and body-movement in psychotherapy: different signals have different associations with outcome. *Frontiers in psychology* 5 (2014), 979.

[42] Fabian T Ramseyer. 2020. Motion energy analysis (MEA): A primer on the assessment of motion from video. *Journal of counseling psychology* 67, 4 (2020), 536.

[43] Holger Regenbrecht and Tobias Langlotz. 2015. Mutual gaze support in videoconferencing reviewed. *Communications of the Association for Information Systems* 37, 1 (2015), 45.

[44] Wataru Sato and Sakiko Yoshikawa. 2007. Spontaneous facial mimicry in response to dynamic facial expressions. *Cognition* 104, 1 (2007), 1–18.

[45] Richard C Schmidt and Michael J Richardson. 2008. Dynamics of interpersonal coordination. In *Coordination: Neural, behavioral and social dynamics*. Springer, 281–308.

[46] Katrin Schoenenberg. 2016. The quality of mediated-conversations under transmission delay. (2016).

[47] Katrin Schoenenberg, Alexander Raake, and P Lebreton. 2014. Conversational quality and visual interaction of video-telephony under synchronous and asynchronous transmission delay. In *2014 Sixth International Workshop on Quality of Multimedia Experience (QoMEX)*. IEEE, 31–36.

[48] Derek Scott. 1996. A Pilot Study of Nonverbal Cues in Videotelecommunicatton. *Psychological reports* 78, 2 (1996), 555–561.

[49] Lucas M Seuren, Joseph Wherton, Trisha Greenhalgh, and Sara E Shaw. 2021. Whose turn is it anyway? Latency and the organization of turn-taking in video-mediated interaction. *Journal of pragmatics* 172 (2021), 63–78.

[50] Keita Suzuki, Masanori Yokoyama, Shigeo Yoshida, Takayoshi Mochizuki, Tomohiro Yamada, Takuji Narumi, Tomohiro Tanikawa, and Michitaka Hirose. 2017. Faceshare: Mirroring with pseudo-smile enriches video chat communications. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 5313–5317.

[51] Maria Tomprou, Young Ji Kim, Prerna Chikersal, Anita Williams Woolley, and Laura A Dabbish. 2021. Speaking out of turn: How video conferencing reduces vocal synchrony and collective intelligence. *PLoS One* 16, 3 (2021), e0247655.

[52] James P Trujillo, Stephen C Levinson, and Judith Holler. 2022. A multi-scale investigation of the human communication system's response to visual disruption. *Royal Society Open Science* 9, 4 (2022), 211489.

[53] Wolfgang Tschacher, Georg M Rees, and Fabian Ramseyer. 2014. Nonverbal synchrony and affect in dyadic interactions. *Frontiers in psychology* 5 (2014), 117886.

[54] Tanya Vacharkulksemsuk and Barbara L Fredrickson. 2012. Strangers in sync: Achieving embodied rapport through shared movements. *Journal of experimental social psychology* 48, 1 (2012), 399–402.

[55] Roel Vertegaal, Ivo Weevers, Changuk Sohn, and Chris Cheung. 2003. Gaze-2: conveying eye contact in group video conferencing using eye-controlled camera direction. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 521–528.

[56] Ishabel M Vicaria and Leah Dickens. 2016. Meta-analyses of the intra-and interpersonal outcomes of interpersonal coordination. *Journal of Nonverbal Behavior* 40 (2016), 335–361.

[57] Chi-Lan Yang, Xiaotong Li, Takuji Narumi, and Hideaki Kuzuoka. 2022. Understanding the impact of technical issues on people's perception and attribution of responsibility in videoconferencing. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. 1–6.

[58] Julian Zubek, Ewa Nagórska, Joanna Komorowska-Mach, Katarzyna Skowrońska, Konrad Zieliński, and Joanna Rączaszek-Leonardi. 2022. Dynamics of remote communication: Movement coordination in video-mediated and face-to-face conversations. *Entropy* 24, 4 (2022), 559.

A DIGITAL APPENDIX

This paper's digital appendix contains the collected study data and the scripts used for the statistical analysis, which are available at https://osf.io/5tpmf/.