Effects of Delay on Nonverbal Behavior and Interpersonal Coordination in Video Conferencing

Chenyao Diao, Stephanie Arévalo Arboleda, Alexander Raake
Audiovisual Technology Group, Technische Universität Ilmenau, Germany
chenyao.diao, stephanie.arevalo, alexander.raake@tu-ilmenau.de

***Abstract*—In this paper, we investigated the effects of transmission delay on individual nonverbal behavior and interpersonal coordination during dyadic video conferencing conversations. For that, we assessed individual-level nonverbal behaviors, including body motion and gaze patterns, and examined participants' interpersonal coordination of body movement. Our results indicate that transmission delay significantly reduces individual body motion. No significant differences in gaze behaviors were found between different delay conditions, however, a trend of participants spending more time looking at their conversational partner in the high-delay condition than in the no-delay condition was observed. Also, we found that transmission delay significantly influences interpersonal body-movement coordination, enhancing structural organization and coordination stability while showing a threshold effect on movement similarity.**

***Index Terms*—video conferencing, transmission delay, nonverbal communication, interpersonal coordination, cross-recurrence quantification analysis**

I. Introduction

Nonverbal communication is crucial in social interactions, as it can help to convey feelings and intentions and influence the responses and behaviors of a communication partner. This form of communication includes body motion, facial expressions, and eye gaze, which can enhance verbal descriptions, emotional expressiveness, and the management of turn-taking mechanisms during social interaction [1, 2, 3]. Moreover, it has been found that individuals tend to coordinate nonverbal behaviors with their conversational partners, synchronizing head- [4] and body- movements [5, 6], speech and turn-taking patterns [7, 8, 9, 10]. This coordination of nonverbal behaviors is also present in video conferencing (VC) interactions, contributing to natural and effective communication [11, 12, 13]. However, technical impairments such as transmission delays may negatively affect these nonverbal behaviors.

Delays in visual feedback can disrupt the interpretation of nonverbal cues such as gaze and facial expressions, leading to misunderstandings and frustration [14, 15, 16]. Ilomäki et al. [17] examined how transmission delay affects conversational behaviors in video-mediated counseling. They found that delays cause unintended overlapping talk, interruptions, and pauses, which hinder the natural flow of conversation and make it challenging for users to take turns smoothly. Additionally, Riedl [18] explored the fatigue and stress potential of VC interactions. Their research highlights that delays increase cognitive load and stress due to extra efforts to restore synchrony in communication processes. This makes the VC experience mentally exhausting, contributing to a higher level of "zoom fatigue" [19, 20]. These findings underscore the importance of understanding the impact of delay on nonverbal communication behaviors during VC interactions.

In this paper, we set out to assess the impact of transmission delay on nonverbal behaviors during dyadic VC conversations. For that, we carried out a user study where participants used VC to play a Celebrity Name Guessing (CNG) game across three transmission delay settings. From the collected webcam and screen recordings, we extracted body motion and gaze patterns and applied cross-recurrence quantification analysis (CRQA) and diagonal cross-recurrence profile (DCRP) to quantify the interpersonal body-movement coordination between participants. To our knowledge, the assessment of interpersonal body-movement coordination during VC conversation under transmission delay has not been performed so far. Our study is guided by the following research question:

- How does transmission delay affect individuals' body motion and gaze behavior, and the interpersonal coordination of body movements during dyadic VC conversations?

II. Related Work

A. Transmission Delay Assessment in Video Conferencing

Transmission delay is one of the most critical technical impairments in VC interactions [21]. Its impact has been evaluated within the user's Quality of Experience (QoE) framework [22], affecting various aspects such as perceived conversational quality [23], task performance [24], communication behavior [15, 23], and physiological responses [25].

Unlike other system factors such as video resolution and packet loss, transmission delay is not always directly perceived by users as a technical impairment, even when delays are substantial [26]. Comparative studies between video conferencing and audio-only communication suggest that visual information can mitigate delay effects [27]. The perception of delays can vary from imperceptible to highly frustrating, with different thresholds depending on the type of conversation [23, 28] as well as the number of participants involved [29]. Therefore, assessing the impact of delay solely on the perceived quality of the video call is quite challenging.

To better understand the impact of transmission delay on the user experience, researchers have explored several delay sensitivity measures. For example, Gergle et al. [30] found that transmission delay decreased task performance (e.g., task completion time, task errors) on a collaborative visual puzzle task. Schoenenberg et al. [23] employed a dense optical flow-based method to quantify body motion from recorded video and found that delays seemed to increase overall body movement. Uhrig et al. [25] used a novel dual-EEG setup to investigate how delay impacts the sensory and cognitive functions of participants; however, their study focused on audio-only communication. Their analysis of intra-brain data revealed significant changes in the beta and gamma frequency bands with varying delay levels, suggesting that larger delays may increase the attentional load. Several studies like [23, 25, 28] have used conversational analysis to extract turn-taking behaviors from recorded conversations, which effectively represent the influence of delay. There have been efforts to model the quality of conversations under the influence of transmission delay using turn-taking parameters (e.g., the proportion of silence, repeat, and overlaps) [31, 26, 32]. Despite these insights, the exploration of body movement and gaze behavior in video-mediated communication under transmission delay remains underdeveloped.

B. Interpersonal Coordination of Body Movements

Interpersonal coordination refers to spontaneous temporal synchronization of body movements and linguistic utterances between individuals engaged in a conversation [33]. The synchronization of body movement has been extensively confirmed in face-to-face conversations, with a positive impact on rapport building and the degree of information exchange [5, 34, 35].

Recent studies suggest that perceptual visual and auditory noise, which typically complicate communication, may require increased bodily coordination [36]. For example, Paxton et al. [37] manipulated visual stimuli using flashing screens on special glasses worn by participants. They found that these visual stimuli, perceived as noise, increased interpersonal head-movement coordination. Similarly, Miles et al. [38] reported a significant increase in interpersonal body-movement coordination between participants when auditory background noise was present. They interpreted that participants more closely coupled their movements to each other when verbal communication became more difficult. This phenomenon suggests that when visual or auditory signals are disrupted, people may subconsciously enhance their interpersonal coordination to compensate for maintaining effective communication.

As VC communication becomes increasingly prevalent, research into the impact of communication channels on interpersonal coordination has recently been of interest. Zubek et al. [39] compared interpersonal body-movement coordination between face-to-face and VC conversations from six interacting dyads. They observed exaggerated communicative gestures and decreased stability in interpersonal coordination in VC conversations, suggesting that dynamic analyses of interpersonal coordination may help explain phenomena like "zoom fatigue". Also, Gvirts et al. [13] investigated interpersonal body-movement coordination in VC interactions by comparing coordination strength in genuine VC interactions to "pseudo-interactions". Their results confirmed the existence of interpersonal coordination of body movement in VC.

Given the crucial role of interpersonal coordination during face-to-face interactions, it is important to understand how it is affected in video-mediated communication, especially in virtual environments where noise is inevitable. However, research on interpersonal coordination in VC interactions remains limited. To our knowledge, no study has explored how interpersonal body-movement coordination is affected during VC conversations in the presence of transmission delay.

III. Study Overview

This study is a secondary analysis of the data collected in [40]. The original experiment employed a within-subjects design (3*3) with three video resolutions (1080p, 480p, 240p) and three additional transmission delays (0 ms, 500 ms, 1000 ms). The current paper is self-contained and focuses on the conditions where transmission delay is the sole technical variable at a video resolution of 1080p.

A. Participants

A total of 46 subjects (23 dyads) participated, aged between 22 to 36 years ($M = 27.04, SD = 3.58$), including 27 females and 19 males. All participants reported normal vision and hearing and received €12 for their participation. We excluded recordings from three dyads because one of the participants was exposed to the study design and the other seven recordings as the data was not usable due to technical problems, leaving 113 recordings for behavioral 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.

B. Apparatus

Participant pairs were placed in separate laboratory rooms and interacted through VC. Two laboratory rooms met the specifications outlined in the ITU recommendations [41]. Each room was equipped with identical hardware, 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. The cameras were mounted on top of the monitors so that participants could see each other's upper body (head and shoulders) when looking at the screen. We used a self-hosted instance of Jitsi Meet [1] as the VC platform. The video codec used was VP9, with a fixed video resolution of 1080p and a frame rate of 30fps. All connections were made via a well-provisioned 1000 Mbps wired LAN. The base system delay was measured at around 14 ms. We employed NetEm command [42] on the outgoing traffic of the client side to add extra transmission delays (0ms, 500ms, and 1000ms) to the base system delay. Video streams (1080p, 30 fps) from the webcam and screen were simultaneously captured and encoded in MP4 format using FFmpeg [2].

[1] https://github.com/jitsi/jitsi-meet
[2] https://ffmpeg.org/

C. Task and Procedure

The task consisted of a Celebrity Name Guessing (CNG) game [41]. It involves guessing the name of a celebrity card 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.

Participants arrived at the lab and received a detailed briefing on the study and procedures, which did not disclose the specific test conditions. After this briefing, participants reviewed and signed the consent form. An in-person training session was then conducted to familiarize participants with the CNG game. Following this, participants underwent a VC training session under a reference condition (1080p resolution, 0 ms delay). After that, participants proceeded to complete all trials, with each trial corresponding to a different test condition. Each trial lasted three minutes, during which participants selected up to two celebrity cards for their partner to guess.

D. Behavioral Measures

1) Body Motion: We employed Motion Energy Analysis (MEA) software (version 4.11b) [43] to capture body motion from both webcam and screen recordings. This method measures changes in gray-scale pixel density between consecutive video frames as the body motion intensity. We selected relevant movement areas that encompass the body movement of each participant to optimize motion capture. This approach generated a time series of 5400 data points per recording (3 minutes, 30 fps). For individual-level activity analysis, we averaged the body motion (*overall body motion*) across all frames from the webcam recording. To assess interpersonal coordination between two participants, we quantified the similarity between the two time series of the body motion: one representing the participant's body motion (as captured by webcam recording) and the other representing their conversational partner's body motion (as captured by screen recording).

2) Gaze Patterns: To understand the gaze focus distribution during interactions, we defined three RoIs: the screen (i.e., looking at the conversational partner), the celebrity card, and the room surroundings. The webcam recordings for each participant were annotated with these gaze RoIs by a coder using ELAN 6.7 [44] frame-by-frame. This approach resulted in a time series of 5400 gaze data points across the three RoIs per webcam recording. At the individual level, we analyzed the total looking time (*t*) directed at each RoI and the number (*n*) of fixations on each RoI.

3) Interpersonal Coordination Quantification: We applied cross-recurrence quantification analysis (CRQA) and analyzed the diagonal cross-recurrence profile (DCRP) to quantify interpersonal coordination using the *crqa* library [45] in R 4.3.3 [46]. CRQA evaluates the time-evolving similarity of two time series by identifying recurring behavioral states over time and quantifying the dynamic patterns of these recurrences [45]. Initially, we determined the CRQA parameters by following the steps outlined in [47]. The selected parameters are 30 for *delay*, 5 for *embedding dimensions*, and 0.3 for the *radius* within the Euclidean norm between z-normalized vectors.

From CRQA, we extracted three primary metrics: the percentage of recurrence (%REC), the percentage of determinism (%DET), and the length of the maximum diagonal line (maxLine). %REC quantifies the total number of instances in two time series visit coinciding regions to the total possible recurrent points, reflecting how frequently similar states recur. It is a measure of global movement similarity that does not take into account the temporal patterns of the recurrent points. %DET captures the total number of recurrent points that are located along diagonal lines, indicating deterministic or structured interactions between individuals. Higher %DET values imply more predictable and structured behavior, whereas lower %DET values suggest more random and unstructured behavior. MaxLine measures the longest diagonal line in the recurrence matrix, reflecting the temporal stability of the coordination. A longer maxLine indicates that the participants can maintain synchronized movements for extended periods, reflecting strong and stable coordination.

DCRP is a detailed analysis within the CRQA framework that captures the recurrence rate (RR) at different time lags, providing a time-dependent measure of how synchronized the states of two interacting individuals are over time [48]. It allows us to explore time-lagged coordination patterns such as leading-following and turn-taking [49]. For calculating DCRP, we used a maximum lag of ±5 seconds (± 150 samples) to capture synchronous movements with delayed onset following the lag setting in [13]. We extracted the RR value at each lag to analyze the synchronization patterns.

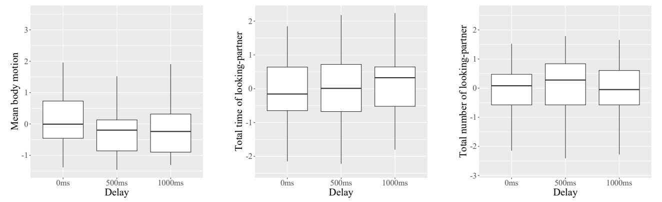
IV. Results

A. Individual Behavior

We applied separate Linear Mixed-effects Models (LMMs) to quantify the effects of transmission delay on individual indicators of nonverbal behavior. The models treated nonverbal behavior measures as the dependent variable and included the transmission delay as fixed effects and participants as random effects. The LMMs were fitted using the *lmer* library [50] in R 4.3.3 [46]. All nonverbal behavior measures were z-normalized prior to analysis.

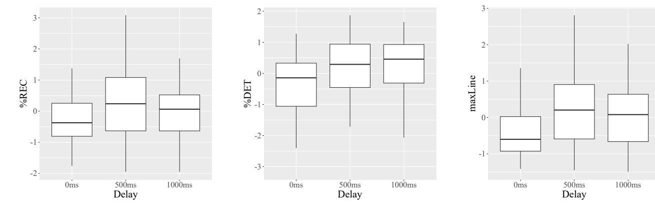
1) Body Motion: The *overall body motion* grouped by transmission delay is shown in Fig. 1a. The analysis indicated a significant main effect of transmission delay on *overall body motion* ($F(2, 75.57) = 3.32, p = 0.041$). In particular, there were significant reductions in *overall body motion* in both the 500ms delay condition ($\beta = -0.34, SE = 0.16, t(75.36) = -2.17, p = 0.034$) and the 1000ms delay condition ($\beta = -0.37, SE = 0.16, t(75.68) = -2.29, p = 0.025$), compared to the baseline condition with no delay.

2) Gaze Patterns: The results showed no significant effect of transmission delay on gaze behavior overall. However, there was a noticeable trend where participants spent more time looking at their conversational partner in the 1000 ms delay condition compared to the no-delay condition ($\beta = 0.25, SE = 0.14, t(74.37) = 1.82, p = 0.073$), approaching



(a) Overall body motion. (b) Looking-partner (t). (c) Looking-partner (n).

Fig. 1: Box plots of (a) overall body motion, (b) total time and (c) total number of times participants looked at their conversational partner.



(a) %REC (b) %DET. (c) maxLine.

Fig. 2: Box plots of (a) the global movement similarity (%REC), (b) the structural organization (%DET), and (c) coordination stability (maxLine) of interpersonal body-movement coordination.

statistical significance. Fig 1b and Fig. 1c illustrate the total time and occurrences that participants look at their conversational partners, grouped by delay.

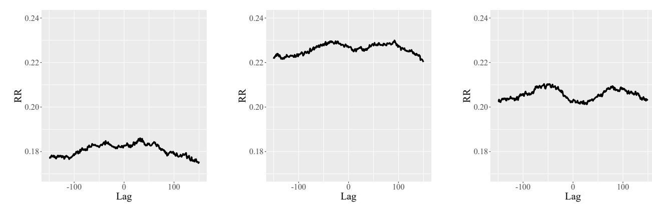
B. Interpersonal Coordination

For the CRQA measures, we conducted a series of LMMs to quantify the effects of transmission delay on %REC, DET, and maxLine. The models treated CRQA measures as dependent variables and included transmission delay as a fixed effect. For the DCRP measures, we performed a growth curve analysis (GCA) [51] to estimate the effects of the lag and transmission delay on RR. Our analysis focused on the first-order (linear lag) and second-order (quadratic lag) orthogonal polynomial forms of the lag. The linear lag allows us to explore the leading-following patterns, while the quadratic lag allows us to investigate concurrent or time-lagged coordination patterns (e.g., turn-taking) [37]. The models treated RR as the dependent variable and included polynomial lag and transmission delay as fixed effects. Participants were included as random effects in all these models to account for individual differences. All measures were z-normalized before analysis.

1) CRQA Measures: %REC of body-movement coordination between participants is illustrated in Fig 2a. The model revealed a significant main effect of transmission delay on %REC ($F(2, 73.73) = 4.42, p = 0.015$). Specifically, the 500 ms delay condition was associated with a significant increase in %REC ($\beta = 0.46, SE = 0.15, t(73.53) = 2.97, p = 0.004$), while the 1000 ms delay condition was not significantly different from the baseline (0 ms delay) ($\beta = 0.21, SE = 0.16, t(73.83) = 1.37, p = 0.176$).

%DET of body-movement coordination between participants is illustrated in Fig. 2b. The model revealed a significant main effect of transmission delay on %DET ($F(2, 74.66) = 3.69, p = 0.030$). Specifically, both the 500 ms delay condition ($\beta = 0.36, SE = 0.16, t(74.44) = 2.24, p = 0.028$) and the 1000 ms delay condition ($\beta = 0.40, SE = 0.16, t(74.78) = 2.45, p = 0.017$) were associated with significant increases in %DET compared to the 0ms delay condition.

MaxLine of body-movement coordination between participants is illustrated in Fig. 2c. The model revealed a significant main effect of transmission delay on maxLine ($F(2, 75.86) = 5.73, p = 0.005$). Specifically, both the 500 ms delay condition



(a) 0 ms. (b) 500 ms. (c) 1000 ms.

Fig. 3: Line plots of the mean diagonal cross-recurrence profiles (in RR) across ± 5 seconds (± 150 samples in 30 fps) lags.

($\beta = 0.62, SE = 0.19, t(75.54) = 3.31, p = 0.001$) and the 1000 ms delay condition ($\beta = 0.42, SE = 0.19, t(76.02) = 2.24, p = 0.028$) were associated with significant increases in maxLine compared to the 0 ms delay condition.

2) DCRP Measures: Fig. 3 shows the DCRP of the mean RR at each lag grouped by delay. The analysis revealed significant main effects of transmission delay on RR ($F(2, 33975) = 1274.03, p < 0.001$). Specifically, both the 500 ms delay ($\beta = 0.37, SE = 0.01, t(33970) = 49.71, p < 0.001$) and the 1000 ms delay ($\beta = 0.13, SE = 0.01, t(33970) = 17, p < 0.001$) were associated with significant increases in RR compared to the 0ms delay condition. In addition, there were significant main effects of polynomial lag on RR ($F(2, 34274) = 15.23, p < 0.001$). Specifically, all conversations showed significant synchronous movement dynamics, particularly as described by the quadratic lag ($\beta = -4.65, SE = 0.96, t(33970) = -4.83, p < 0.001$). No significant leading-following dynamics were observed as captured by the linear lag ($p = 0.862$). Furthermore, there was a significant positive effect of quadratic lag on RR at the 1000 ms delay condition ($\beta = 3.87, SE = 1.37, t(33970) = 2.83, p = 0.005$).

V. Discussion

A. Impact of Delay on Body Motion and Gaze Behavior

Our results indicate that delays significantly affect body motion; however, they do not significantly affect gaze behavior.

Specifically, *overall body motion* was significantly reduced in the delayed conditions compared to the no-delay condition (see Fig. 1a), suggesting an inhibition effect of delay on user body movements. Our findings contrast with the study in [23], which found that delays (400 ms and 800 ms) increase body movements. There are two key differences between the studies that may account for these contrasting results. On the one hand, in our study, most participants were not familiar with their conversational partners well, only nine pairs reported being friends with their conversational partners. Conversely, Schoenenberg et al. [23] carried out their tests with well-familiar pairs, being either good friends or partners. The informal pre-test in their study showed that familiarity was highly important to create a relaxed atmosphere for the participants by not acting restrained in their social behavior and freely using the visual channel for interaction. On the other hand, the video resolution setting in our experiment was fixed at 1080p, while in their study, the resolution was much lower (480p). Higher video quality may mitigate some of the adverse effects of transmission delays [27], thereby reducing compensatory body movements. These differences suggest that the effect of transmission delay on body motion may be influenced by relationship familiarity and video quality. Further analytical exploration is needed to understand these discrepancies fully.

Regarding gaze patterns, we only found a trend suggesting that participants spent more time looking at their conversation partners under the high delay condition (1000 ms) compared with the no-delay condition (see Fig.1b), approaching statistical significance. This aligns with the idea that participants may compensate for communication difficulty by increasing gaze engagement (e.g., regulating turn-taking and conveying attention) [52].

B. Impact of Delay on Body-movement Coordination

We investigated the impact of transmission delay on three primary measures from CRQA: %REC (*movement similarity*), %DET (*structural organization*), and maxLine (*coordination stability*). The *structural organization* and *coordination stability* were significantly greater under delayed conditions compared to the no-delay condition (see Fig. 2b and Fig. 2c). This result suggests that delays can increase certain aspects of interpersonal body-movement coordination. We relate these findings to those that found other factors such as audio and visual noise to enhance interpersonal coordination [37, 38]. One possible explanation for this is that participants may simplify their movement and adopt more predictable and stable movement patterns to compensate for the communication difficulty caused by delay. Our finding is consistent with research presented in [53], which investigated feedback delay using a beat-tapping task and rhythm-clapping task. Their study demonstrated that delay can lead participants to rely heavily on anticipatory behaviors and adapt their movements to maintain stable patterns of interpersonal coordination.

The impact of delay on *movement similarity* showed a threshold effect. At a medium delay condition (500 ms), *movement similarity* was significantly greater than in the no-delay condition, indicating that a moderate delay might encourage participants to synchronize their movements more frequently. In contrast, at a high delay condition (1000 ms), *movement similarity* was not significantly different from the no-delay condition (see Fig. 2a). This could imply that beyond a certain threshold, the delay becomes too disruptive, preventing participants from effectively adjusting their movements to maintain similarity. Supporting this, research by Washburn et al. [54] on a joint movement task found that small to moderate feedback delays (200–400 ms) can lead to more deliberate synchronization (i.e., more effective in facilitating anticipation), while larger delays (over 600 ms) may exceed the window for effective compensation.

Our analysis of the DCRP measures revealed significant positive effects of transmission delay on RR. This finding aligns with the results from the CRQA measures, demonstrating that delays significantly enhance the recurrence of interaction patterns. Participants are likely to repeat certain body movements to ensure that their actions are understood. Moreover, we observed significantly more pronounced time-lagged coordination patterns in the high (1000 ms) delay condition compared to the no-delay condition (see Fig. 3c). This suggests that with a larger delay, participants rely more on time-lagged coordination to synchronize their body movements, i.e., participants adjust their body movements over a longer period (the lag needed to reach max RR) to align with their partner, highlighting the adaptive mechanisms they employ to cope with larger delays.

Furthermore, we found that the baseline condition (0 ms delay) generally resulted in lower %REC, %DET, maxLine, and RR. This may indicate that coordination patterns in delay-free VC interactions rely more on real-time feedback (see Fig. 3a). This can sometimes lead to less stable and structured coordination as interlocutors constantly adapt to immediate changes.

VI. Conclusion

We investigated how transmission delay influences nonverbal behaviors among dyads as they played a CNG game using VC. We explored several nonverbal cues that can be extracted from webcam and screen recordings, including individual nonverbal activities like body motion, gaze directions, and interpersonal coordination of body movement. Our analysis indicates that transmission delays significantly inhibit body movement and increase interpersonal body-movement coordination. In terms of gaze behavior, our results showed a trend of participants spending more time looking at their conversational partner in the high-delay condition than in the non-delay condition. Our main contribution is an evaluation of how transmission delay affects interpersonal body-movement coordination. This could inform future research on developing objective VC QoE assessment methods using these nonverbal cues. Our data and analysis materials are freely available in public repositories (see https://osf.io/xgq7d/).

Acknowledgment

This research is part of the CO-HUMANICS (Co-Presence of Humans and Interactive Companions for Seniors) project and is funded by the Carl-Zeiss-Foundation ("Breakthroughs 2020" program, https://www.carl-zeiss-stiftung.de/programm/czs-durchbrueche). 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 this manuscript thoroughly underwent human review and revision.

References

[1] S. Duncan. "Some signals and rules for taking speaking turns in conversations." In: *Journal of personality and social psychology* 23.2 (1972), p. 283.
[2] B. M. DePaulo et al. "Nonverbal communication." In: (1998).
[3] K. Jokinen. "Gaze and gesture activity in communication". In: *Universal Access in Human-Computer Interaction. Intelligent and Ubiquitous Interaction Environments: 5th International Conference, UAHCI 2009, Held as Part of HCI International 2009, San Diego, CA, USA, July 19-24, 2009. Proceedings, Part II 5*. Springer. 2009, pp. 537–546.
[4] T. Yokozuka et al. "The relationship between head motion synchronization and empathy in unidirectional face-to-face communication". In: *Frontiers in psychology* 9 (2018), p. 1622.
[5] W. Tschacher et al. "Nonverbal synchrony and affect in dyadic interactions". In: *Frontiers in psychology* 5 (2014), p. 117886.
[6] N. Higo et al. "Interpersonal similarity between body movements in face-to-face communication in daily life". In: *PLoS One* 9.7 (2014), e102019.
[7] A. R. McGarva et al. "Attraction and social coordination: Mutual entrainment of vocal activity rhythms". In: *Journal of Psycholinguistic Research* 32 (2003), pp. 335–354.
[8] C. Gallois et al. "Communication accommodation theory". In: *Theorizing about intercultural communication* (2005), pp. 121–148.
[9] K. Shockley et al. "Conversation and coordinative structures". In: *Topics in Cognitive Science* 1.2 (2009), pp. 305–319.
[10] J. H. Manson et al. "Convergence of speech rate in conversation predicts cooperation". In: *Evolution and Human Behavior* 34.6 (2013), pp. 419–426.
[11] J. Campbell et al. "The perpetuation of entrained behavior during computer-mediated communication". In: *Social Infulence* 5.1 (2010), pp. 59–73.
[12] M. Tomprou et al. "Speaking out of turn: How video conferencing reduces vocal synchrony and collective intelligence". In: *PLoS One* 16.3 (2021), e0247655.
[13] H. Gvirts et al. "Virtual social interactions during the COVID-19 pandemic: the effect of interpersonal motor synchrony on social interactions in the virtual space". In: *Scientific Reports* 13.1 (2023), p. 10481.
[14] S. M. Ehrlich et al. "Communicating facial affect: It's not the realism, it's the motion". In: *CHI'00 extended abstracts on human factors in computing systems*. 2000, pp. 251–252.
[15] S. R. Powers et al. "The effect of video feedback delay on frustration and emotion communication accuracy". In: *Computers in Human Behavior* 27.5 (2011), pp. 1651–1657.
[16] L. M. Seuren et al. "Whose turn is it anyway? Latency and the organization of turn-taking in video-mediated interaction". In: *Journal of pragmatics* 172 (2021), pp. 63–78.
[17] S. Ilomäki et al. "Effects of transmission delay on client participation in video-mediated group health counseling". In: *Qualitative health research* 31.12 (2021), pp. 2328–2339.
[18] R. Riedl. "On the stress potential of videoconferencing: definition and root causes of Zoom fatigue". In: *Electronic Markets* 32.1 (2022), pp. 153–177.
[19] B. K. Wiederhold. *Connecting through technology during the coronavirus disease 2019 pandemic: Avoiding "Zoom Fatigue"*. 2020.
[20] N. Döring et al. "Videoconference fatigue: A conceptual analysis". In: *International journal of environmental research and public health* 19.4 (2022), p. 2061.
[21] D. P. Hove et al. "The Shortcomings of Video Conferencing Technology, Methods for Revealing Them, and Emerging XR Solutions". In: *PRESENCE: Virtual and Augmented Reality* 31 (2022), pp. 283–305.
[22] J. Skowronek et al. "Quality of experience in telemeetings and videoconferencing: a comprehensive survey". In: *IEEE Access* 10 (2022), pp. 63885–63931.
[23] K. Schoenenberg et al. "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. 2014, pp. 31–36.
[24] S. Garg et al. "A study of the effects of network latency on visual task performance in video conferencing". In: *CHI Conference on human factors in computing systems extended abstracts*. 2022, pp. 1–7.
[25] S. Uhrig et al. "Effects of delay on perceived quality, behavior and oscillatory brain activity in dyadic telephone conversations". In: *2018 Tenth International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE. 2018, pp. 1–6.
[26] G. Berndtsson et al. "Methods for human-centered evaluation of MediaSync in real-time communication". In: *MediaSync: Handbook on Multimedia Synchronization* (2018), pp. 229–270.
[27] J. Tam et al. "Video increases the perception of naturalness during remote interactions with latency". In: *CHI'12 Extended Abstracts on Human Factors in Computing Systems*. 2012, pp. 2045–2050.
[28] T. Michael et al. "Interactivity-Based quality prediction of conversations with transmission delay". In: *Speech and Computer: 22nd International Conference, SPECOM 2020, St. Petersburg, Russia, October 7–9, 2020, Proceedings 22*. Springer. 2020, pp. 336–345.
[29] M. Schmitt et al. "The influence of interactivity patterns on the quality of experience in multi-party video-mediated conversations under symmetric delay conditions". In: *Proceedings of the 3rd International Workshop on Socially-aware Multimedia*. 2014, pp. 13–16.
[30] D. Gergle et al. "The impact of delayed visual feedback on collaborative performance". In: *Proceedings of the SIGCHI conference on Human Factors in computing systems*. 2006, pp. 1303–1312.
[31] A. Raake et al. "Predicting speech quality based on interactivity and delay." In: *INTERSPEECH*. 2013, pp. 1384–1388.
[32] J. He et al. "QoE Metrics for Interactivity in Video Conferencing Applications: Definition and Evaluation Methodology". In: *Proceedings of the 15th ACM Multimedia Systems Conference*. 2024, pp. 178–189.
[33] F. J. Bernieri et al. "Synchrony, pseudosynchrony, and dissynchrony: measuring the entrainment process in mother-infant interactions." In: *Journal of personality and social psychology* 54.2 (1988), p. 243.
[34] A. Tsuchiya et al. "Body movement synchrony predicts degrees of information exchange in a natural conversation". In: *Frontiers in psychology* 11 (2020), p. 817.
[35] R. Dale et al. "Body synchrony in triadic interaction". In: *Royal Society open science* 7.9 (2020), p. 200095.
[36] K. Kodama et al. "Influence of Visual Information on Interpersonal Co-ordination of Head-and Body-Movement During Dyad Conversations". In: *Proceedings of the Annual Meeting of the Cognitive Science Society*. Vol. 44. 44. 2022.
[37] A. Paxton et al. "Interpersonal movement synchrony responds to high-and low-level conversational constraints". In: *Frontiers in psychology* 8 (2017), p. 235014.
[38] K. Miles et al. "Behavioral dynamics of conversation,(mis) communication and coordination in noisy environments". In: *Scientific Reports* 13.1 (2023), p. 20271.
[39] J. Zubek et al. "Dynamics of remote communication: Movement coordination in video-mediated and face-to-face conversations". In: *Entropy* 24.4 (2022), p. 559.
[40] C. Diao et al. "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. 2023, pp. 240–243.
[41] ITU-T Recommendation. P.1305. *Effect of delays on telemeeting quality*. 2016.
[42] S. Hemminger et al. "Network emulation with NetEm". In: *Linux conf au*. Vol. 5. 2005, p. 2005.
[43] F. T. Ramseyer. "Motion energy analysis (MEA): A primer on the assessment of motion from video." In: *Journal of counseling psychology* 67.4 (2020), p. 536.
[44] H. Sloetjes et al. *ELAN (version 6.7). Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive*. 2023.
[45] M. I. Coco et al. "Cross-recurrence quantification analysis of categorical and continuous time series: an R package". In: *Frontiers in psychology* 5 (2014), p. 72510.
[46] R. R Core Team et al. "R: A language and environment for statistical computing". In: (2013).
[47] S. Wallot. "Multidimensional Cross-Recurrence Quantification Analysis (MdCRQA)–a method for quantifying correlation between multivariate time-series". In: *Multivariate behavioral research* 54.2 (2019), pp. 173–191.
[48] S. Wallot et al. "Analyzing multivariate dynamics using cross-recurrence quantification analysis (crqa), diagonal-cross-recurrence profiles (dcrp), and multidimensional recurrence quantification analysis (mdrqa)–a tutorial in r". In: *Frontiers in psychology* 9 (2018), p. 365352.
[49] A. Main et al. "An exploratory analysis of emotion dynamics between mothers and adolescents during conflict discussions." In: *Emotion* 16.6 (2016), p. 913.
[50] D. Bates et al. "Package 'lme4'". In: *convergence* 12.1 (2015), p. 2.
[51] D. Mirman et al. "Statistical and computational models of the visual world paradigm: Growth curves and individual differences". In: *Journal of memory and language* 59.4 (2008), pp. 475–494.
[52] Z. Degutyte et al. "The role of eye gaze in regulating turn taking in conversations: a systematized review of methods and findings". In: *Frontiers in Psychology* 12 (2021), p. 616471.
[53] I. R. Roman et al. "Delayed feedback embedded in perception-action coordination cycles results in anticipation behavior during synchronized rhythmic action: A dynamical systems approach". In: *PLoS computational biology* 15.10 (2019), e1007371.
[54] A. Washburn et al. "Harmony from chaos? Perceptual-motor delays enhance behavioral anticipation in social interaction." In: *Journal of experimental psychology: human perception and performance* 41.4 (2015), p. 1166.