The Interpreters' Newsletter Special Issue 2025 (30), X-XX
DOI: 10.13137/2421-714X/37800 ISSN 1591-4127 (print) ISSN 2421-714X (online)
https://www.openstarts.units.it/dspace/handle/10077/2119

Turn-transition in video-mediated interpreting: a case study of a multiparty and multi-sited public conversation

Han Wang
University of Bologna

Abstract

Video-mediated interpreting has become a prominent form of interpreting; however, the emergence of diverse configurations in the post-pandemic era presents new challenges and calls for further investigation. This study examines a video conference involving four primary interlocutors and an interpreter connected from distributed locations, with an audience observing the interaction, in order to elucidate how turn-transition (TT) is managed by multiple remote participants. The paper first reviews previous studies on TT and video-based interaction from a multimodal perspective and outlines the methodology applied. To contextualise the analysis, background information about the event and its participants is then provided, with particular attention to the interactional ecology and technical setup. Drawing on the multimodal approach in Conversation Analysis, TT instances are analysed by focusing on intra-turn gaps and overlapping turn-beginnings. Findings show that lengthy gaps are common, while overlaps occur seldom. This pattern reflects a controlled behaviour pattern of participants which is shaped by the multiparty and multi-sited configuration. The presence of the audience creates a formal dimension, which also impacts participants' behaviour. Overall, the study contributes to a more comprehensive understanding of diversified technology-based interpreting practices.

Keywords

Video-mediated interpreting, consecutive interpreting, turn-transitions, multiparty interaction, multimodal analysis.

Video-based technology undoubtedly opens up new scenarios of interpersonal communication, enabling interactants to see each other despite the physical distance. However, their perceptions are strongly conditioned by what the webcam 'sees' and how the software represents it. Interlocutors appear to each other as "talking heads" (Due/Licoppe, 2020: 8): their gaze and head are oriented towards the screen, and bodily movements below the chest level and hand-forearm gestures with a wider trajectory may not framed by the camera. As a result, the potential to draw on gaze and body orientation as floor-management cues diminishes significantly.

Interpreting activity also involves mutual signalling and backchanneling through various multimodal cues (Mason 2012; Pasquandrea 2012; Davitti/Pasquandrea 2017; Vranjes *et al*. 2018). Consequently, video-mediated interpreting (VMI) is likewise subject to visual constraints. Various studies have investigated the impact of such constraints on turn-transition (TT), focusing on overlapping speech (De Boe 2021), chunking (Davitti 2019; Licoppe/Veyrier 2020; Hansen/Svennevig 2021), and gaze (Vranjes 2023); how participants navigate these challenges has also been discussed (Davitti/Braun 2020; Klammer/Pöchhacker 2021). However, this growing body of research has predominantly focused on three-participant interactions in community settings, with one party taking part remotely. While this *pas de trois* (Wadensjö 1998) is perhaps the most emblematic VMI setup, video-based interpreting may also involve more than three participants and unfold across diverse social contexts and technical configurations.

To contribute to a more comprehensive understanding of this evolving practice, the present study examines an authentic event that took place among four guest speakers and an interpreter, named "Conversation on the occasion of the exhibition 'My Brilliant Friend': When Literature Appears on Screen" (hereinafter CMBF). The participants were connected to a videoconferencing system from distributed locations, and their conversation was followed by an audience both on-site and via webcast. This configuration raises two key research questions. First, in a situation in which the employment of multimodal cues is challenged by technical constraints, how do multiple participants manage the floor over a video link? Secondly, how does the presence of an audience, which brings the multiparty conversation into the public domain, affect the participants' behaviour? To address these questions, the paper begins with a literature review on turn-allocation in VMI and video-based interaction from a multimodal perspective. Following the introduction of the methodology, the organisational and technical setup of the interaction is discussed. An analysis of representative TT instances occurring in the data is then presented, and the findings discussed.

1. Turn-transitions and video-based interaction

Talk-in-interaction unfolds on a turn-by-turn basis (Sacks *et al*. 1974). When participants are physically co-present, they mobilize a wide range of multimodal re-

sources to facilitate TT (Goodwin 2000; Mondada 2006, 2007). With the advent of internet and video technology, remote gatherings have emerged as a cost-effective alternative to traditional in-person meetings. However, research on video-mediated communication has increasingly reported an undermined effectiveness of nonverbal cues and problematic visual access to co-participants (Croes *et al*. 2019; Due/Licoppe 2020; Oittinen 2022).

Similar issues have also arisen in VMI. For instance, gestures can serve as tokens inviting interlocutors to take or relinquish the floor (Streeck 1993). Yet, in video-based communication, forearm movements may extend outside the frustum of the webcam and fail to perform this signalling function (Braun 2013; Hansen/Svennevig 2021; Cavents *et al*. 2025). Even when arm movements are adequately framed by the webcam, they still have to be spotted by co-participants to fulfil the regulatory purpose. However, constant monitoring appears more problematic in VMI than in face-to-face interaction. In physical proximity, our peripheral vision can readily detect movements, making it possible to spot co-participants' behaviour without directing one's gaze towards them. By contrast, if one's sightline is diverted from a two-dimensional representation, such as a screen, monitoring activity becomes affected (Heath/Luff 1991). This is particularly relevant to interpreters when engaging in notetaking: their visual attention may be drawn away from the screen by the notes, potentially compromising monitoring of the ongoing interaction (De Boe 2021). Chunking, the practice of dividing extended contributions into shorter turns to enable rendition delivery, is a collaborative activity between the interpreter and primary parties (Licoppe 2023). However, chunking attempts frequently fall short in VMI (Licoppe *et al*. 2018; Davitti 2019; Licoppe/Veyrier 2020; Hansen/Svennevig 2021) since bodily hints, such as sideways gaze, postural change and in-breath, are more likely to go unnoticed in remote communication (Braun 2013; Davitti 2019; Vranjes 2023). To navigate such constraints, participants upgrade subtle nonverbal hints into explicit verbal interventions to make themselves noticed by remote co-participants (Braun/Taylor 2012; Licoppe/Veyrier 2020; Cavents *et al*. 2025).

As already highlighted above, the predominant configuration investigated within the literature on VMI is represented by three-party and two-sited interaction, where one party is connected via video link and the other two are co-located. In this configuration, if the seating and technical devices are properly positioned and form a triangular spatial arrangement, participants can still leverage bodily cues to facilitate TTs (Licoppe/Veyrier 2020; Klammer/Pöchhacker 2021). However, changes to the setup are likely to alter the interactional dynamics: as highlighted in Verhaegen's work with simulated data (2023, 2025), participants' behavioural patterns during TT in a three-point configuration, where all participants are situated separately, differ from those observed in the traditional two-sited situation. The involvement of more than three participants via video link may further increase the complexity of interaction. As multiple participants can potentially bid for the floor and negotiate the transition-relevant place (TRP), more efficient coordination becomes necessary. Yet, when separately located, participants' visual access to others relies exclusively on video technology. To show all the connected participants, the screen is divided into a larger number of smaller segments, which

challenges the visualization of bodily conduct. In co-present conversations, participants form an interactional space where they can publicly address a co-participant and display their turn-management intentions by shifting their visual and body orientation (Kendon 2010; Mondada 2013; Chen/Brandt 2021). In contrast, in videoconferencing, participants only appear as images arranged on a two-dimensional "face wall" (Hochuli/Jud 2023: 194). Even when orienting toward a co-participant, one typically maintains a frontal posture, with minimal adjustments in gaze and body direction. This is where the "Mona Lisa" effect (Luff *et al.* 2016) comes into play: to whoever the gaze is directed at, all co-participants may have the illusion that the person is orienting to them. Hence, gaze and body orientation cannot be relied on as TT hints in this context.

The number of participants and their locations may influence turn management, as does the presence of an audience. Research on broadcast news interviews (Clayman 2013) indicates that the presence of an audience determines the purpose of the interaction, which becomes a public conversation that "should be managed as 'talk for overhearers', so that audience members do not feel that they are listening in on a purely private conversation" (Heritage/Clayman 2011: 216). The overhearing participants thus take the speech event beyond the private domain and add a formal layer to it. In this type of institutional talk, TT usually follows a well-established question–answer sequence (Drew/Heritage 1992: 25-27), departures from which are deemed "problematic" and "sanctionable" (Heritage/Clayman 2011: 222).

To our knowledge, the implications of multiparty and multi-sited participation, along with the presence of an audience, are still relatively unexplored in VMI research. The present case study contributes to the existing literature by providing a preliminary account of how TT unfolds in a context characterised by these features.

2. Methodology

The data analysed are drawn from the video recording of the CMBF webcast. The author did not attend the event but accessed the data on the social media account of the organiser. Therefore, the participants' behaviour was not affected by the researcher's presence, and the phenomena observed can be considered as naturally occurring.

Embracing the perspective of Discourse Analysis in considering VMI as a communicative event situated within a specific context (Braun 2017; Pöchhacker/Klammer 2023), the analysis examines first of all the organisational and technical configuration of CMBF. The second part of the analysis zeros in on four actual instances of TT by adopting the multimodal approach to Conversation Analysis (Stivers/Sidnell 2005; Streeck *et al*. 2011; Deppermann 2013; Mondada 2014, 2016). In their seminal work regarding turn-management, Sacks *et al.* (1974: 700-701) affirmed that "transitions (from one turn to the next) with no gap and no overlap are common. Together with transitions characterised by slight gap or slight overlap, they make up the vast majority of transitions". This implies that in

ordinary conversation TTs are typically achieved at a fast pace. By looking into gaps and overlaps in CMBF, we explored the general traits of TTs in this VMI event and discussed how they are shaped.

As turns at talk in CMBF were often extensive and sometimes lasted up to three minutes, the excerpts in the following section do not include the full turns but only the transition. The annotation is aligned chronologically from left to right and top to bottom. A timeline shows the duration of the actions and pauses. In order to include both verbal and nonverbal features, the original utterances, English translation, gaze orientation and other kinetic behaviours, such as nodding and body leaning, were arranged on four tiers. Only variations from frontal gaze and body orientation were indicated in the transcription; otherwise, it can be assumed that participants were facing the screen without any notable bodily conduct. When a feature is not present – for example, when no talk is produced – the relative tier is removed, and the remaining tiers are kept in the same vertical order specified above.

3. Analysis

To elucidate the setting in which the participants interact, section 3.1 unpacks the interactional framework and technical arrangement of CMBF. Then, the analysis proceeds to the examination of the gaps between turns and overlaps so as to sketch out the salient features of TT practice.

3.1 The communicative event

Adapted from Italian writer Elena Ferrante's series of novels, *L'amica geniale* (My Brillant Friend) is a TV show which has achieved worldwide success, including in China. To enable the Chinese audience to meet the crew of the show, in 2022 the Italian Institute of Culture in Shanghai and a local art gallery co-organised CMBF. Due to international travel restrictions still in force in China, the event was held in a hybrid configuration: the speakers connected from distributed locations and interacted on a videoconferencing platform. A small group of show enthusiasts gathered at the art gallery, while a larger audience joined via webcast.

The institutional greetings and final Q&A with the on-site audience have been excluded from the analysis, which focuses exclusively on the eighty-minute interaction among the following guests: the Chinese-speaking director of the art gallery (AGD) and three Italian-speaking participants, namely a film critic (FC), the TV show director (SD) and the leading actress (LA). An interpreter (INT) supported them with Italian-Chinese consecutive interpreting. Throughout the conversation, FC and AGD took on the responsibility of introducing topics and posing questions, to which SD and LA provided answers and comments. Considering the defined roles and question-answer sequence, as well as the presence of an audience, the event resembled a news interview with a defined procedure in the shift of floor (Heritage/Clayman 2011: 215-226; Clayman 2013: 630-656). Thus, a normative

TT framework can be identified, as outlined in Table 1. As section 3.2 will show in detail, actual dynamics sometimes did not progress as smoothly as expected or deviated from this template. Nevertheless, we still refer to this normative framework in the following analysis to indicate where TT can be expected to take place.

Turns	Question	*TT1*	Rendition	*TT2*	Answer	*TT3*	Rendition	*TT4*
Participants	AGD/FC		INT		SD/LA		INT	

Table 1: the normative TT framework in CMBF

Given that CMBF took place in a remote space, and participants' perceptions of one another relied heavily on the videoconferencing technology, it is worth mentioning how their images were presented. Two different layouts were found in the data: the first was the gallery view that displayed participants' video feed in a grid; the other was the speaker view, in which the active speaker was shown in full screen, while others were arranged in a smaller strip at one side of the screen. Due to the secondary nature of the data, we lack precise information on whether the layout on the participants' side matched the one shown in the recording. Nevertheless, both visualizations may pose challenges to mutual monitoring on the part of participants. The grid view is generally the default layout in multiparty videoconferencing. As Figure 1 shows, in the event analysed, SD and LA were co-located and shared the same video window, AGD was present at the art gallery with the on-site audience, FC and INT were connected via video link from separate sites.[1] This layout offers a comprehensive view of all the participants. However, their images are distributed across the screen. Due to the reduced peripheral vision in two dimensions mentioned above, proper monitoring of the co-participants would imply gazing at their frames one by one, which requires time and close attention. The smaller images of the participants may also hinder visualisation of nonverbal signals. On rare occasions throughout the event, speaker-centred view was selected. This option provides a closer view of the current speaker; however, their image remains pinned and enlarged at the conclusion of their turn, while other potential next speakers still appear in reduced size or even remain hidden. Pre-beginnings hints displayed by the latter thus risk being overlooked.

1 The fifth image pertains to the officer from the Italian Institute of Culture in Shanghai and his interpreter. Neither of them intervened during the session analysed in the present study.

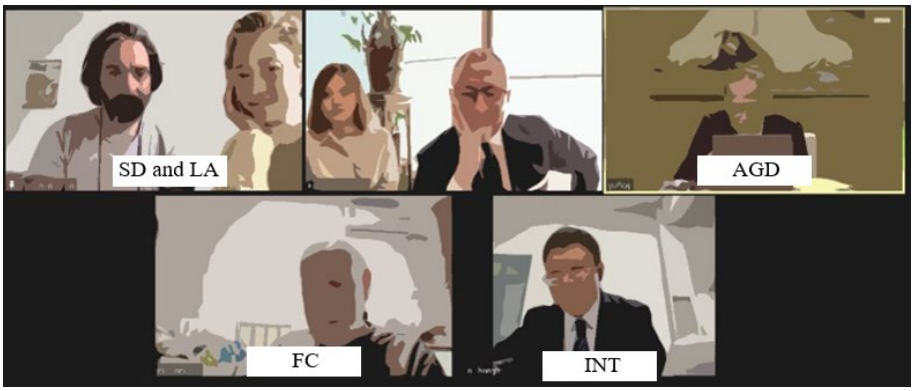


Figure 1: Grid view of the participants in CMBF

3.2 Turn-transitions

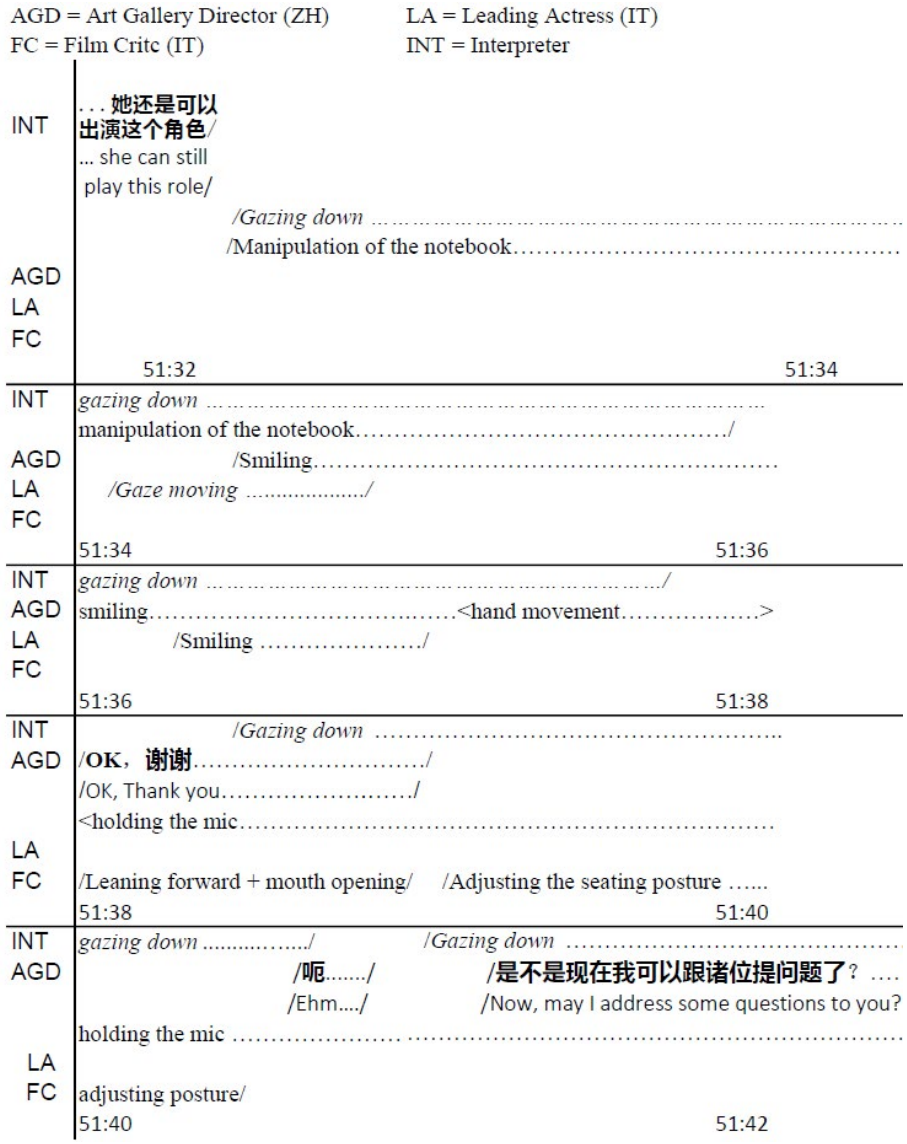
In the following section, we outline the characteristics of intra-turn gaps and overlaps in CMBF and present key examples to provide an index of the progression of TTs in the analysed event.

3.2.1 Intra-turn gaps

In face-to-face dialogue interpreting, gaps longer than 2.2 seconds are considered as "extreme" cases (Vranjes/Oben 2022: 634). While in CMBF, long silences between turns were recurrent: 40 out of 93 floor transfers exceeded two seconds, during which the participants closely scrutinised the screen before taking the floor. This phenomenon recalls what Davitti/Braun (2020: 287) describe as "awkwardness", which emerges when the interaction does not proceed "as smoothly or gracefully as one may expect, without necessarily hindering the outcome of the event". While it is known that internet latency can cause delayed data transmission and bring about an out-of-sync effect (Seuren *et al.* 2021), the analysis reveals other factors that potentially intensified the participants' hesitation during TTs.

One such factor concerns next speaker selection, which is particularly relevant at the potential conclusion of each question-answer sequence. As mentioned in section 3.1, both the roles of questioner and addressee were covered by two participants. Generally, an addressee was explicitly designated during the question turn, who took the floor in TT2 (see the normative TT framework in Table 1). However, in TT4, when an answer was provided and translated, local negotiation of the next speaker might become necessary: a questioner could initiate a new sequence, or the same addressee could continue the ongoing sequence by expanding on their answer with additional turns. Across 21 total occurrences of TT4, 11 lasted more than two seconds. Excerpt 1 offers insights into this scenario. Here, LA has responded to a question posed by FC,

and when INT concludes the rendition of LA's answer, a lengthy silence occurs. All participants remain still for two seconds. During this time, AGD first gives a hint of a smile and then extends his hand to the microphone to get ready to talk.



Excerpt 1: Long gap in TT4

At the TRP, LA might have self-selected and continued developing her answer, FC could have raised other follow-up questions, or AGD might have decided to introduce new topics since he had not intervened until that moment. However, the visual cues did not appear straightforward enough for them to infer other's intent: FC did not change his posture and

gaze direction, while LA briefly glanced at the screen. AGD continued staring and smiling at the screen, yet nearly no movements from the co-participants can be observed. It is only after LA smiled slightly that AGD finally grasped the microphone. At the same moment, FC leaned towards the computer. He first opened his mouth but held back immediately and then began to adjust his posture. AGD paused again during FC's movements and when he finally resumed, he did not immediately elaborate his talk but checked tentatively with the co-participants "May I address some questions to you?". Summarily, the constant monitoring without taking actions, the faint smiles on the part of AGD and LA, as well as the verbal checking question of AGD, all seem to suggest their uncertainty regarding co-participants' plans for the ongoing transition. A larger number of participants arguably complicates the next speaker selection and contributes to increasing "awkwardness" (Davitti/Braun 2020: 287).

Another aspect likely to bring about longer gaps and slower TT progression is the identification of TRPs. Given the multiparty nature of CMBF, sequences in which both the questioner and the addressee spoke Italian could occur. In these cases, the rendition in Chinese of their talk was delivered for the overhearing AGD and the audience but could not be grasped by the participants who were interacting. They could therefore not rely on semantic and syntactic clues to determine the end of the rendition and initiate their turn as promptly as in monolingual conversation. Out of 17 transfers under these circumstances, 13 were characterised by gaps exceeding two seconds. Excerpt 2 exemplifies one of these instances, in which FC addresses a question to SD in Italian and INT translates it into Chinese. When relinquishing the floor, INT shifts his gaze from right to left, where in the video something similar to a booklet is visible under his left hand (Figure 2). SD initially makes no moves but starts to glance across the screen a while later. After approximately two seconds, he lowers the hand covering his mouth and inhales. Then he takes an additional moment to gaze at the screen before finally taking the floor.

SD = Show Director (IT) INT = Interpreter

INT	**...在连贯性方面非常的突出**/
	… is outstanding
	in terms of coherence/
	/Gazing down to the left ……
	holding the booklet with his left hand……
SD	*/Gaze moving* ……/
	covering his mouth with one hand……
	30:28 30:30
INT	*gazing down to the left* ……//*Gazing down to the right* …….
	holding the booklet……..<Right hand moving……
SD	/**HBO (.) la prima volta con (.)**
	/HBO (.) the first time with (.) -
	hand down…./Inhalation + leaning forward/
	30:30 30:32

Excerpt 2: Long gap after the rendition

Figure 2: INT at the conclusion of the rendition

A participant taking the floor after a turn delivered in a foreign language is not the typical practice in three-party dialogue interpreting, but it can happen. For instance, when chunking is performed, the speaker's talk is divided into several stretches. Every time that they regain the floor after the interpreter, they have to follow the translation in the hearer's language. Yet, the chunked turns are generally shorter, which facilitates the speaker in recognising the boundary based on the duration and prosodic contour. It is also in the speaker's interest to regain speakership after the rendition, and therefore, they may monitor the interpreter more closely. Turn-final cues such as gaze and body orientation of the interpreter have been found to aid speakers in this task (Vranjes *et al*. 2018; Davitti 2019). In CMBF, the narrative nature of the event seemed to lead the speakers to deliver extended talks. In the case illustrated, INT's rendition in Chinese lasted nearly one and a half minutes. For SD, an Italian speaker, identifying prosodic turn-final cues in such a multi-unit talk produced in Chinese might arguably be more challenging than in shorter, chunked turns. Moreover, INT's bodily conduct provided few clues about the conclusion of his turn: although he looked away from his notes, he did not return to a stand-by posture. Instead, he directed his sightline to something at the edge of the screen, to which SD, as a remote co-participant, did not have enough visual access.

3.2.2 Overlaps

In focusing on TTs, the analysis presented in this section regards exclusively simultaneous talk of more participants arising at a TRP. This occurs when the next speaker takes the floor before the current one has completed their turn, or when two or more participants claim the floor simultaneously. In the data, these occurrences were limited to 14, yet they remain relevant to the purpose of the analysis.

A notable type of overlap in CMBF is related to floor-bidding between INT and one of the Italian speakers, which occurred three times. As mentioned earlier, in question-answer

sequences taking place between Italian speakers, INT was expected to provide the Chinese translation for AGD and the audience. In other words, INT can be considered the designated next speaker in TT1 and TT3. However, Excerpt 3 illustrates an instance in which FC did not follow the normative sequence. When SD leaves the floor at 21:56, FC does not wait for the rendition but takes the floor, overlapping with INT. As soon as he realises the overlap, he drops out with a smile and extends one hand forward.

SD = Show Director (IT) INT = Interpreter
FC = Film Critc (IT)

SD	**… come è fatta fisicamente…..……/**	
	… how she physically looks like……/	
INT	…*gazing down* ……	
	… writing ……	
FC	…*gazing down* ……	
	21:56	21:58
SD	/*Looking away* ……	
INT		**/呃/**
		/Ehm/
	…*gazing down* ……	
	… writing…… ……/	
FC	…*gazing down* ……../	/*Gazing down*
		/Nodding/
	21:58	22:00
INT	**/呃就是跟##…….** ./	**/呃 (.) 嗯**？ /
	/Eh, so, in relation to ##…../	/Eh (.) ehm?/
	…*gazing down* …… /	
FC	**/Tu pensi che/**	
	/You think that/	
	…*gazing down* /	
	22:00	22:02
INT		**/Scusami** ## ……/
		/Excuse me ## ……/
	/Gazing down ……	
	/Soft laughter ……/	
FC		/## ……/
	/Smiling + leaning back……/	/Extending one hand forward/
	22:02	22:04

Excerpt 3: Overlap between FC and INT

Although FC's voice was not audible at 22:04, it was possible to read from his lips that he uttered "*vai tu*" (go ahead), therefore inviting INT to proceed. FC's behaviour could be considered a reflexive reaction to a previous turn in his language; nevertheless, the instance reflects a distinctive aspect of CMBF with respect to three-participant interpreting. In the latter, interpreter-as-next-speaker represents not only a normative procedure in the consecutive mode but also a primary necessity due to the linguistic barrier separating the two primary interlocutors. When one party is able to use the language of the other, direct exchanges may be temporarily established, yet they may still rely on the interpreter to ensure understanding and prevent miscommunication (Monteoliva-García 2020). In CMBF, in contrast, the two interacting participants were both native in Italian, for whom leaving space for the rendition in Chinese represented an interactional norm rather than a necessity. Thus, they may occasionally overlook this convention and compete for the floor with INT in TT1 and TT3.

Another type of overlap occurred when an inter-turn pause was mistaken for a TRP (De Boe 2021); this happened five times and INT was always involved. Excerpt 4 provides a case in point: in the ongoing turn at talk, SD explains to AGD that the author of the novels "My Brilliant Friend" watched and liked the TV show. The last utterance he produces is syntactically complete and ends with a falling intonation. However, instead of returning to a more relaxed posture (Mortensen/Hazel 2024) as he sometimes did when giving up the floor, SD continues to lean towards and gaze at the screen (Figure 3). In the meantime, INT is absorbed in notetaking and has not glanced at the computer. A moment later, SD resumes with a new utterance; almost simultaneously INT stops writing and takes the floor in overlap with SD. While INT raises his gaze to the monitor and withdraws immediately, SD continues with what he is saying.

SD = Show Director (IT) INT = Interpreter

SD	**… le è piaciuto molto ……..……../**	
	… she liked it very much …… ../	
	[No video feed.....]	
	[No video feed.....]/Leaning foward …	
INT	…*gazing down* …… ……	
	… writing …… ……	
	17:10	17:12
SD		/**Però ha capito che** .....
		/However she understood
	leaning foward ……	
INT		/**呃**..... /
		/Ehm.../
	…*gazing down* …… ……	
	… writing ……/	/Writing .
	17:12	17:14

Excerpt 4: Overlap between SD and INT

Figure 3: SD in forward-leaning posture

In co-present interactions, taking notes does not necessarily hinder the interpreter's possibility to monitor primary participants, since one's peripheral view can capture slight bodily movements or inhalation of people nearby. Video-based communication, on the other hand, requires higher levels of visual focus to recognize the bodily cues of others (Heath/Luff 1991; Cavents *et al*. 2025), and activities such as writing may impede the interpreter from constant monitoring, resulting in non-coordinated actions such as overlapping speech.

4. Discussion and conclusion

This study has explored TT features in CMBF, a virtual public conversation involving five participants connected from four locations. The data have shown that perceptible lengthy gaps occurred in over 40% of TTs, during which the participants kept gazing at the screen but delayed taking actions. Instances of overlaps were not frequent and rarely led to significant breakdowns in the communication.

These findings differ not only from the minimal-gap-and-overlap pattern found in ordinary conversations (Sacks *et al*. 1974), but also from the interpreter's strategy of taking turns "as soon as the opportunity arises" in dialogue interpreting (Englund Dimitrova 1997:149). Overlapping speech has also been found to occur frequently in remote settings due to participants' tendency to "leav[e] no or little time in between turns" (De Boe 2021: 145), and in such settings usually turns out to be more disruptive. Our result aligns instead with contrastive studies conducted on monolingual conversations, which have shown longer time between turns and fewer overlaps in video-based settings compared to face-to-face interactions (Boland *et al*. 2022; Tian *et al*. 2023). Based on simulated data, Verhaegen (2025) confirmed the same result of a slower progression of TT in VMI; however, the analysis revealed only subtle differences in floor-transfer timing between two-point and three-point configurations.

We argue that the TT patterns emerging from our data are shaped by some distinct traits of CMBF, which introduced new challenges to participants in perceiving and interacting with each other.

Despite an overarching ritualised configuration, local next speaker selection and establishment of TRP among multiple participants turned out to be more complex than

in situations involving three participants. The Italian speakers did not need INT's aid to communicate with each other. Nevertheless, due to the presence of an overhearing audience, they still had to align with the interpreter-as-next-speaker rule and cope with a previous turn delivered in a foreign language. Navigating such situations may lead to hesitation among participants and prompt them to monitor others more closely.

The analysis also highlighted hindered visual access to remote participants and its potential implications for negotiating TTs through nonverbal features (Braun 2013; Davitti 2019, Davitti/Braun 2020; Hansen 2020; Hansen/Svennevig 2021; De Boe 2021). A distinguishing feature of CMBF in this regard is that almost all the participants took part remotely, making their perception of one another even more "fractured" (Luff *et al*. 2003). Across the multiple small images of co-participants, subtle cues, such as body leaning or inhalation, are therefore likely to become more difficult to discern. Therefore, the participants' sustained and focused monitoring, as observed in the data, may have also delayed floor-taking, resulting in lengthy gaps.

As concerns overlaps, their limited occurrence may arguably be attributed to the participants' attentive monitoring and prudent turn-taking behaviour, which is visible in the recordings and evident from the transcripts. The prolonged turns reduced the number of shifts between speakers and, consequently, the possibilities of overlaps. In addition, the ritualised interview-like framework of the event limited the participants from taking as much initiative as in ordinary conversation, thereby lowering the risk of simultaneous bidding for the floor. This format also provided a normative sequential procedure that facilitated the repair of overlaps: in most cases, the designated speaker continued, and the other withdrew without further complications.

Another factor that may have shaped the participants' controlled behaviour lies in the audience's presence, which imparted a formal dimension to the conversation (Drew/Heritage 1992: 25-27, Heritage/Clayman 2010: 215-226). For instance, despite the reduced potential of embodied cues, the floor-management remained primarily nonverbal and delicate. Upgraded cues, such as hand raising or verbal interventions, which have been observed in previous research on VMI (Braun/Taylor 2012; Licoppe/Veyrier 2020; Cavents *et al*. 2025), were rarely observed in our data. In particular, INT displayed a constrained embodied pattern and seldom used overt tokens when taking or leaving the floor. This conduct may have allowed him to avoid appearing like the author or the addressee of the talk, making it more audience-oriented (Clayman 2013: 636). Visibility concerns may have impacted his actions as well. In public events such as CMBF, invited speakers enjoy a prestigious position as focal figures. However, in the videoconferencing system, the interpreter's presence is rendered as visible as that of the speakers: in grid view, all participants' images are equal in size; in speaker-centred view, the interpreter is displayed full-screen while speaking, just as the speakers are. Due to the "Mona Lisa effect" (Luff *et al*. 2016), when an interpreter looks at a co-participant on the screen, it appears to everyone (including the audience) that the interpreter is gazing at them or seeking eye contact. In light of the formal layer of CMBF, the controlled behaviour of INT may therefore be understood as a virtual 'step back' to limit his visibility.

This apparently reticent behaviour seems to have implications for the fulfilment of the interpreting task. In CMBF, many of the turns at talk were of extended length. In similar conditions, interpreters often collaborate with the primary speaker to chunk

the speech into shorter turns (Licoppe 2023). As discussed earlier, INT rarely undertook such activity, arguably to avoid interrupting the speakers in public view and drawing attention to himself. Excerpt 4 may be counted as one attempt, yet it concluded quickly as he dropped out. INT dedicated considerable time to notetaking, as evident from the video data. However, writing distracted his direct sightline from the screen to the block notes and, eventually, hampered his ability to monitor the ongoing interaction. Navigating between notetaking and monitoring, or between chunking and waiting, implies delicate choices. Regardless of which aspect the interpreter focuses on, others may be affected. This dynamic not only highlights the multitasking nature of interpreting but also underscores the need for future evidence from similar contexts to inform recommendations and guidelines for interpreter training.

The findings of the study obviously require further validation, as it is based on a single case and utilises secondary video data collected online two years after the event. With this delay, gathering additional evidence directly from the participants, such as split recordings and interviews, turned out to be unfeasible. This made it impossible to isolate the impact of latency on the phenomena emerged, and some of the findings may be subject to the author's personal reconstruction of the event. Nevertheless, the analysis provides insight into some of the TT patterns in a previously unexplored setting and arrangement, in doing so contributing to a more comprehensive understanding of diversified technology-based interpreting practices. As much of the TT process relies on nonverbal hints, the validity of the multimodal perspective in shedding light on nuanced behaviour in virtual co-presence is also demonstrated.

In the post-pandemic era, events or workplace meetings with people connected from separate places have become the "new normal" (Due/Licoppe 2020: 2). This shift is particularly the case when participants find themselves on different continents. Multiparty and multi-sited virtual meetings, such as CMBF, are not just occasional occurrences. Future explorations including participants' accounts of the event, along with video data recorded from their perspectives, can be expected to advance current knowledge in VMI studies.

Transcription conventions

/	The beginning and the end of each segment
…	The continuation in time of the annotated content until its conclusion
Bold	Original utterances
Italics	Gaze movement
< >	Indicates that the enclosed behaviour occurs simultaneously with the previously annotated one
(.)	Inter-phrase pause, hesitation
#	Inaudible content
[ ]	Notes of the annotator

Boland J. E. / Fonseca P. / Mermelstein I. / Williamson M. (2022) "Zoom disrupts the rhythm of conversation", *Journal of Experimental Psychology: General* 151/6, 1272-1282.

Braun S. (2013) "Keep your distance? Remote interpreting in legal proceedings: A critical assessment of a growing practice", *Interpreting* 15/2, 200-228.

Braun S. (2017) "What a micro-analytical investigation of additions and expansions in remote interpreting can tell us about interpreters' participation in a shared virtual space", *Journal of Pragmatics* 107, 165-177.

Braun S. / Taylor J. (2012) "Video-mediated interpreting: an overview of current practice and research", in S. Braun / J. L. Taylor (eds) *Videoconference and Remote Interpreting in Criminal Proceedings*, Guildford, University of Surrey, 27-57.

Cavents D. / De Wilde J. / Vranjes J. (2025) "Towards a multimodal approach for analysing interpreter's management of rapport challenge in onsite and video remote interpreting", *Journal of Pragmatics* 235, 220-237.

Chen Q. / Brandt A. (2021) "Speakership, recipiency and the interactional space: Cases of 'Next-speaker self-selects' in multiparty university student meetings", *Journal of Pragmatics* 180, 54-71.

Clayman S. (2013) "Conversation analysis in the news interview", in J. Sidnell / T. Stivers (eds) *The handbook of conversation analysis*, Chichester, Wiley-Blackwell, 630-656.

Croes E. A. J. / Antheunis M. L. / Schouten A. P. / Krahmer E. J. (2019) "Social attraction in video-mediated communication: The role of nonverbal affiliative behavior", *Journal of Social and Personal Relationships* 36/4, 1210-1232.

Deppermann A. (2013) "Multimodal interaction from a conversation analytic perspective", *Journal of Pragmatics* 46/1, 1-7.

Davitti E. (2019) "Methodological explorations of interpreter-mediated interaction: Novel insights from multimodal analysis", *Qualitative Research* 19/1, 7-29.

Davitti E. / Braun S. (2020) "Analysing interactional phenomena in video remote interpreting in collaborative settings: implications for interpreter education", *The Interpreter and Translator Trainer* 14/3, 279-302.

Davitti E. / Pasquandrea S. (2017) "Embodied participation: What multimodal analysis can tell us about interpreter-mediated encounters in pedagogical settings", *Journal of Pragmatics* 107, 105-128.

De Boe E. (2021) "Management of overlapping speech in remote healthcare interpreting", *The Interpreters' Newsletter* 26, 137-155.

Drew P. / Heritage J. (1992) *Talk at work*, Cambridge, Cambridge University Press.

Due B. L. / Licoppe C. (2020) "Video-Mediated Interaction (VMI): Introduction to a special issue on the multimodal accomplishment of VMI institutional activities", *Social Interaction. Video-Based Studies of Human Sociality* 3/3, <https://doi.org/10.7146/si.v3i3.123836>.

Englund Dimitrova B. (1997) "Degree of interpreter responsibility in the interaction process in community interpreting", in S. E. Carr (ed.) *The Critical Link:*

Interpreters in the Community, Amsterdam/Philadelphia, John Benjamins, 147-164.
Goodwin C. (2000) "Action and embodiment within situated human interaction", *Journal of Pragmatics* 32/10, 1489-1522.
Hansen J. P. B. (2020) "Invisible participants in a visual ecology: Visual space as a resource for organising video-mediated interpreting in hospital encounters", *Social Interaction. Video-Based Studies of Human Sociality* 3/3, <https://doi.org/10.7146/si.v3i3.122609>.
Hansen J. P. B. / Svennevig J. (2021) "Creating space for interpreting within extended turns at talk", *Journal of Pragmatics* 182, 144-162.
Heath C. / Luff P. (1991) "Disembodied conduct: Communication through video in a multi-media office environment", in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 99-103.
Hochuli K. / Jud J. (2023), "Non-talking heads: How architectures of digital copresence shape question-silence-answer-sequences in university teaching", in D. vom Lehn / W. Gibson / N. Ruiz-Junco (eds) *People, Technology, and Social Organization*, London/New York, Routledge,181-206.
Heritage J. / Clayman S. (2011) *Talk in action: Interactions, identities, and institutions*, Oxford, Wiley-Blackwell.
Kendon A. (2010) "Spacing and orientation in co-present interaction", in A. Esposito / N. Campbell / C. Vogel / A. Hussain / A. Nijholt (eds) *Development of Multimodal Interfaces: Active Listening and Synchrony*, Berlin/Heidelberg, Springer, 1-15.
Klammer M. / Pöchhacker, F. (2021) "Video remote interpreting in clinical communication: A multimodal analysis", *Patient Education and Counseling* 104/12, 2867–2876.
Licoppe C. (2023) "Consecutive interpreting and multimodal sequences", in L. Gavioli / C. Wadensjö (eds) *The Routledge Handbook of Public Service Interpreting*, London/New York, Routledge, 155-174.
Licoppe C. / Verdier M. / Veyrier C.-A. (2018) "Voice, power and turn-taking in multi-lingual, consecutively interpreted courtroom proceedings with video links", in J. Napier / R. Skinner / S. Braun (eds) *Here or There. Research on Interpreting via Video Link*, Washington, Gallaudet University Press, 299-322.
Licoppe C. / Veyrier C.-A. (2020) "The interpreter as a sequential coordinator in courtroom interaction: 'Chunking' and the management of turn shifts in extended answers in consecutively interpreted asylum hearings with remote participants", *Interpreting* 22/1, 56-86.
Luff P. / Heath C. / Kuzuoka H. / Hindmarsh J. / Yamazaki K. / Oyama S. (2003) "Fractured Ecologies: Creating Environments for Collaboration", *Human-Computer Interaction* 18, 51–84.
Luff P./ Heath C. / Yamashita N. / Kuzuoka H. / Jirotka M. (2016) "Embedded Reference: Translocating Gestures in Video-Mediated Interaction", *Research on Language and Social Interaction* 49/4, 342-361.

Mason I. (2012) “Gaze, positioning and identity in interpreter-mediated dialogues”, in C. Baraldi / L. Gavioli (eds) *Coordinating Participation in Dialogue Interpreting*, Amsterdam/Philadelphia, John Benjamins, 177-200.
Mondada L. (2006) “Participants’ online analysis and multimodal practices: Projecting the end of the turn and the closing of the sequence”, *Discourse Studies* 8/1, 117-129.
Mondada L. (2007) “Multimodal resources for turn-taking: Pointing and the emergence of possible next speakers”, *Discourse Studies* 9/2, 194-225.
Mondada, L. (2013) “Embodied and spatial resources for turn-taking in institutional multi-party interactions: Participatory democracy debates”, *Journal of Pragmatics* 46/1, 39-68.
Mondada L. (2014) “The local constitution of multimodal resources for social interaction”, *Journal of Pragmatics* 65, 137-156.
Mondada L. (2016) “Challenges of multimodality: Language and the body in social interaction”, *Journal of Sociolinguistics* 20/3, 336-366.
Monteoliva-García E. (2020) “The collaborative and selective nature of interpreting in police interviews with stand-by interpreting”, *Interpreting* 22/2, 262–287.
Mortensen K. / Hazel S. (2024) “The Temporal Organisation of Leaning in Social Interaction”, *Social Interaction. Video-Based Studies of Human Sociality* 7/4, <https://doi.org/10.7146/si.v7i4.152386>.
Oittinen T. (2022) “Negotiating collaborative and inclusive practices in university students’ group-to-group videoconferencing sessions”, *Linguistics and Education* 71, 101107.
Pasquandrea S. (2012) “Co-constructing Dyadic Sequences in Healthcare Interpreting: A multimodal account”, *New Voices in Translation Studies* 8, 132-157.
Pöchhacker F. / Klammer M. (2023) “Ensuring understanding in video remote-interpreted doctor–patient communication”, in E. De Boe / J. Vranjes / H. Salaets (eds) *Interactional Dynamics in Remote Interpreting: Micro-analytical Approaches*, London/New York, Routledge, 66-90.
Sacks H. / Schegloff E. A. / Jefferson G. (1974) “A simplest systematics for the organization of turn-taking for conversation”, *Language* 50, 696-735.
Seuren L. M. / Wherton J. / Greenhalgh T. / Shaw S. E. (2021) “Whose turn is it anyway? Latency and the organization of turn-taking in video-mediated interaction”, *Journal of Pragmatics* 172, 63-78.
Stivers T. / Sidnell J. (2005) “Introduction: Multimodal interaction”, *Semiotica* 156, 1-20.
Streeck J. (1993) “Gesture as communication I: Its coordination with gaze and speech”, *Communication Monographs*, 60/4, 275-299.
Streeck J. / Goodwin C. / LeBaron C. (2011) “Embodied interaction in the material world: An introduction”, in J. Streeck / C. Goodwin / C. LeBaron (eds) *Embodied Interaction: Language and Body in the Material World*, Cambridge, Cambridge University Press, 1-26.
Tian Y. / Liu S. / Wang J. (2023) “A Corpus Study on the Difference of Turn-Taking in Online Audio, Online Video, and Face-to-Face Conversation”, *Language and Speech* 67/3, 593–616.

Verhaegen M. (2023) “Exploring turn-taking in video-mediated interpreting: A research methodology using eye tracking”, *The Interpreters’ Newsletter* 28, 151-169.
Verhaegen M. (2025) *Turn-taking Management in Video-mediated Interpreting and Face-to-face Interpreting: A comparative Study using Eye Tracking*, unpublished PhD Thesis, University of Antwerp.
Vranjes J. (2023). “Where to look? On the role of gaze in regulating turn-taking in video remote interpreting”, in E. De Boe / J. Vranjes / H. Salaets (eds) *Interactional Dynamics in Remote Interpreting: Micro-analytical Approaches*, London/New York, Routledge, 113-134.
Vranjes J. / Brône G. / Feyaerts K. (2018) “On the role of gaze in the organization of turn-taking and sequence organization in interpreter-mediated dialogue”, *Language and Dialogue* 8/3, 439-467.
Vranjes J. / Oben B. (2022) “Anticipation and timing of turn-taking in dialogue interpreting: A quantitative study using mobile eye-tracking data”, *Target* 34/4, 627-651.
Wadensjö C. (1998) *Interpreting as Interaction*, London/New York, Longman.