

Chenxi Li\*

# The role of gaze in meaning negotiation episodes in video synchronous computer-mediated interactions

https://doi.org/10.1515/jccall-2022-0005 Received March 19, 2021; accepted January 9, 2022; published online February 23, 2022

**Abstract:** The outbreak of COVID-19 led to the replacement of many traditional classroom teaching and learning activities with online interactions through synchronous audio or video conferencing tools. Technically, the key difference between audio and video conferencing is the latter's use of a webcam. Therefore, this study aims to investigate the webcam's role in video synchronous computermediated communication (SCMC). Negotiation for meaning episodes, in which interlocutors try to resolve non-understanding through verbal interaction, were selected for analysis because they are widely believed to be helpful for second language acquisition. Specifically, this paper focuses on where interlocutors look during negotiated interactions in video SCMC and whether a statistical relationship exists between looking at a peer's video image and successful meaning negotiation. Four dyads of English learners completed information gap tasks with potentially unknown words through learner-learner video SCMS interactions. Students' multimodal performance was screen recorded for a gaze analysis and one-on-one interviews were conducted to confirm the students' gaze directions. The results reveal a statistically significant positive relationship between the duration of one's gaze at the peer's video image and the number of successful meaning negotiation episodes in a dyad.

**Keywords:** gaze; multimodality; negotiation for meaning; SCMC

# 1 Introduction

The last 20 years have witnessed the rapid development of online learning and teaching across the world. Specifically, synchronous computer-mediated communication (SCMC) has attracted increasing attention in recent years. Particularly with the global outbreak and spread of COVID-19 in 2020 and 2021,

<sup>\*</sup>Corresponding author: Chenxi Li, The Open University, Milton Keynes, UK, E-mail: cecilia.lcx@outlook.com

Open Access. © 2022 Chenxi Li, published by De Gruyter. © This work is licensed under the Creative Commons Attribution 4.0 International License.

traditional face-to-face teaching has been largely replaced by online teaching using SCMC technology for millions of online learners in China and throughout the world (Crawford, Butler-Henderson, Rudolph, & Glowatz, 2020; Huang, Liu, Tlili, Yang, & Wang, 2020; iiMedia Research, 2020). Therefore, research on SCMC for online language learning is urgently needed and has significant practical and pedagogical value for schools and universities worldwide. In the online teaching practice in China, one of the most frequently asked questions by online teachers and students is whether they should open the webcam for video conferencing or only use audio chat. Yet very few existing studies address this question directly.

In video conferencing classrooms, the teaching and learning process is mediated by technology; therefore, the affordances of different types of technology play an important role in how learners communicate and learn languages in the mediated environment (Hampel & Stickler, 2005, 2012; Stockwell, 2010). Of all possible modes of communication (textual, aural, visual, etc.), the visual mode afforded by the webcam is the most complicated one owing to the wide range of multimodal information it provides, including the interlocutor's gaze, facial expressions, posture, gestures and surrounding environment. Furthermore, due to the lack of a shared physical communication environment and loss of partial body visibility, gaze has become one of the most effective resources for interpreting an interlocutor's attitude, stance and behaviour (Sindoni, 2014). The direction of online learners' gaze in video conferencing classrooms can affect what information they receive from the screen, which may, in turn, affect how they react to peers both linguistically and with paralinguistic cues (e.g. facial expressions and gestures). Therefore, where students look in video conferencing classrooms and how their gaze affects their online language learning are the key issues this paper aims to explore.

# 2 Literature review

# 2.1 Rationale for a statistical gaze analysis

The present study uses a statistical method to analyse the relationship between the direction of students' gaze and their language learning online. To accurately measure how gaze affects online language learning, this study adopts "negotiation for meaning" episodes for analysis because they involve resolving nonunderstanding in a conversation in the target language, which is widely believed to show a certain degree of second language acquisition (SLA), according to the interaction hypothesis (Ellis, 2000; Long, 1996, 1988). This theoretical framework in SLA has also been used in many prior SCMC studies (Hubbard &

Levy, 2016). The rationale for choosing the specific research objective and method is presented below.

First, there seems to be disagreement in the literature on the role of the visual mode in video SCMC. Some studies argue that video can be distracting for students when they are trying to focus on the language during task interactions (e.g. Lee, 2006; Van der Zwaard & Bannink, 2014, 2016). However, other studies report positive effects of video for second language learning in SCMC environments (e.g., Wang, 2006; Wang & Tian, 2013; Yamada & Akahori, 2009). Moreover, existing SCMC studies suggest that some students do not look at peers' video images during meaning negotiations in video SCMC (Guo & Möllering, 2016; Lee, 2006; Wang & Tian, 2013). Consequently, they often miss important multimodal information from their peers during such interactions. Conversely, students who tend to look at their peers' video images during negotiated interactions seem to complete more successful meaning negotiations than those who seldom do so (Wang & Tian, 2013). This generates an initial hypothesis that there might be a positive statistical relationship between the time participants spend looking at their peer's video image and their success in meaning negotiation episodes (MNEs). Yet there does not appear to be any existing research exploring this specific question. Therefore, the following review focuses on the role of gaze in video SCMC interactions, which is closely related to the abovementioned hypothesis.

# 2.2 The role and types of gaze in video SCMC interactions

The role of gaze has been widely studied in many non-online communication environments concerning language learning, psychology, communication studies, etc. Argyle, Ingham, Alkema, and McCallin (1973) summaries the following key functions of gaze in face-to-face interactions: 1) seeking information and feedback; 2) signalling attitude; 3) controlling the synchronization of speech and 4) managing/avoiding intimacy. In video SCMC interactions, however, the causes and effects of gaze can be very different from those in face-to-face communication. First, gaze can be determined by many factors, including the interlocutor's cultural background, the technological tools, task design and the interlocutor's surrounding physical environment (Develotte, Guichon, & Vincent, 2010; Lamy & Flewitt, 2011; Satar, 2013). Moreover, in video SCMC, either through a built-in or external webcam, mutual eye contact is technically impossible. Therefore, Sindoni (2014) comments that gaze and its role in facilitating SCMC interactions "are not easily gauged by analysts" (p. 340).

Indeed, it appears that only three articles have focused extensively on the role of gaze in video SCMC interactions. Develotte et al. (2010) explore the types and

role of gaze that online language teachers use during teacher-learner interactions for pedagogical purposes in video SCMC. Their findings suggest that webcam images play a complementary role in contributing to the information contained in a verbal message and could potentially be distracting. However, when a webcam is used, facial expressions (e.g. smile, frown) and gestures (e.g. nod) take on various empathic and interactional functions. Satar (2013) identifies five learner gaze patterns in learner-learner video SCMC interactions: manipulating gaze constantly, manipulating gaze strategically, avoiding gaze completely, directing gaze and free gaze. Satar also emphasizes that video SCMC requires manipulating interlocutor's own image. The conclusion, then, is that the video SCMC environment, at least before 2010, could not provide immediacy, as proposed by Argyle et al. (1973), in face-to-face communication due to "the disembodied and limited representation, delays and distortions in audio and video, and the lack of eye contact" (Satar, 2013, p. 139). Using another gaze classification scheme, Lamy and Flewitt (2011) identify four types of gaze: looking at one's peer, one's own image, camera and chat window (as cited in Satar, 2013). This study offers an easy way to classify gaze types according to the direction of interlocutors' gaze in video SCMC.

The three cited studies use different methods to classify different types of gaze during video SCMC interactions. Develotte et al. (2010) identify five degrees of webcam use and gaze, indicating a hierarchy of competence in multimodal video SCMC. Lamy and Flewitt (2011) categorize gaze according to the part of the video screen on which SCMC interface interlocutors focus, and Satar's (2013) framework of gaze is identified according to the learner's intentions. These three studies demonstrate that there is no established framework for analysing and classifying gaze types in video SCMC interactions. This leaves researchers to develop their own approach to gaze analysis according to their particular research objectives, participants, devices, contexts and interface of the video SCMC software. The scant research on the role of gaze in video SCMC also indicates that this topic is in its infancy and requires further research attention. Moreover, as the quality of webcams has improved substantially since the 2010s, the role of gaze may also be greater today. Therefore, further research and more detailed evidence are needed on the role of gaze in current online video conferencing environments.

Methodology-wise, all relevant studies have been purely qualitative. Thus, there exists a methodological gap, as no one has quantitatively measured the durations of interlocutors' gaze on different parts of the video conferencing screen and the potential effects of such gaze on online language learning. In terms of linguistic episodes for analysis, none of the above studies focuses on MNEs, which are a key process for SLA (Long, 1996; Varonis & Gass, 1985). Therefore, this study aims to fill in this research gap by exploring if there exists a statistical relationship

between the duration of language learners' gaze on an interlocutor's video image and the success of their meaning negotiation during task interactions.

## 2.3 Research questions

Therefore, this study aims to answer the following two research questions:

- 1) How do interlocutors use their gaze during MNEs in video SCMC?
- 2) What is the statistical relationship (if any) between the amount of time interlocutors spend looking at their peer's video image and their success in meaning negotiation in video SCMC?

# 3 Methodology

## 3.1 Research context

This study was conducted in a prestigious higher education institution (HEI) in Beijing which provides both independent online language courses and qualification courses at the undergraduate and postgraduate levels. Its students are usually full-time employed adult learners who study online in their spare time to gain further degree qualifications, expand their knowledge and improve their language proficiency.

This project was conducted by the author as part of her doctoral research. The author designed an online course which was provided for free and students' performance was not related to their assessment at the HEI. Two online teachers were invited to deliver this online course, including giving task instructions, facilitating task interactions when needed and offering post-task feedback to students. All eight participants were recruited from this HEI and all had at least half a year of online language learning experience. They were all female adult learners with a proficiency level around B2 according to CEFR criteria.

The video conferencing system used in the online course (Figure 1) consisted of presentation slides, the online teacher's video image, attendance list, students' video images, students' text chat area and some control buttons. The interface allowed for placing a peer's video image either in the centre (central view) or the top right-hand corner (corner view), which would have important implications for the direction of students' gaze. The teachers had overall control of the system. They were in charge of managing students' access to audio/video channels for verbal or multimodal interactions with online teachers and peers.

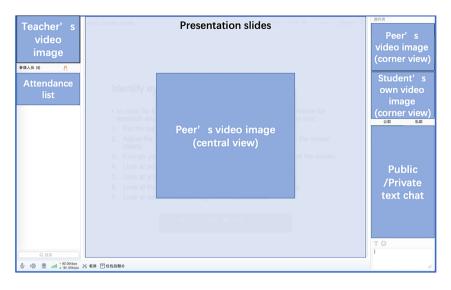


Figure 1: The video conferencing system interface.

# 3.2 Research procedures

Data for the overall doctoral research project were collected in three stages (Table 1). The first stage was designed to familiarize participants with both their peers and audio/video peer interaction and to test their general English

Table 1: Research procedures.

Stages	Session	Content	Data
Stage 1: Preparation	SCMC session 1	Introduction, pairing, ice-breaking; pre-task vocabulary test (video only)	Not used for analysis
	SCMC session 2	Mock IELTS speaking test; opinion gap tasks (audio and video)	
Stage 2: Main tasks	SCMC session 3	Spot-the-difference tasks (1 task in audio and 1 task in video)	8 h of screen video recordings
	SCMC session 4	Problem-solving tasks (1 task in audio and 1 task in video)	
Stage 3: Interviews	Face-to-face interview	<ul><li>a) Video stimulated recall interview about negotiated interactions</li><li>b) Normal interview about students' opin- ions and backgrounds</li></ul>	12 h of audio recordings

proficiency and knowledge of target lexical items. Then each dyad performed two types of tasks – namely, spot-the-difference and problem-solving tasks – in both audio and video modes. In each dyad, the two participants had different task sheets (see the task examples in Appendix). They were asked to describe the pictures or items in their task sheets to each other and work out the differences or make decisions together. The target lexical items were embedded in the tasks, requiring students to negotiate the meanings of these words to complete the tasks. Their performance was screen recorded to produce the main data for analysing their gaze and multimodal performance in the audio/video SCMC classrooms. After each task session, the author watched the recordings, identified meaning negotiation stances and prepared related questions for the video stimulated recall interview (VSRI). The interview was designed to confirm the directions of students' gaze and their understanding of the target lexical item at certain points in the negotiation interactions.

## 3.3 Gaze data collection and the coding scheme

## 3.3.1 The coding scheme for gaze direction

As the literature review revealed, there are no well-established methods for classifying and analysing gaze. Essentially, the coding scheme should satisfy two key criteria: 1) deductively, codes should be created in a way that supports answering the research question, and 2) inductively, the codes should reflect the key patterns in the data. Drawing on the two research questions, the codes should reflect where participants direct their gaze, particularly at the peer's video image, which is similar to Lamy and Flewitt's (2011) classification. After repeatedly watching and analysing the participants' gaze data during negotiated interactions, the author observed two key gaze directions: looking up at the peer's video image on the screen and looking down at one's own task sheet. Other gaze directions, including looking at one's own video image, the teacher's video image or places outside the screen, are classified as "other gaze directions". Finally, cases where the direction of a participant's gaze could not be clearly identified due to low clarity or fast eye movement are coded as "unidentified gaze directions". Therefore, the final coding scheme including these four codes is as follows: 1) gaze directed at peer's video image; 2) gaze directed at the task sheet; 3) other gaze directions and 4) unidentified gaze directions.



Figure 2: D4B\_T3: Gaze directed at task sheet.

## 3.3.2 Pre-task gaze direction test

Sections 3.3.2–3.3.4 describe the three methods of collecting and triangulating gaze-related data. To capture and present participants' gaze performance and other relevant multimodal features, this paper must display many screenshots of participants during the video SCMC task interactions. All participants were informed about the purposes and procedures of this research and provided consent for the author to use their data with pseudonyms in research publications.

Before each task, the online teacher asked the students to complete a gaze direction identification test. For the test, students were asked to put their task sheets on the desk, adjust the webcam, enlarge their peer's video image and then look at the peer's video image, followed by the teacher's video image and then their own video image for 3 s each. Where students directed their gaze during the test could provide a reference point for a student's gaze when she was looking at her peer's video versus other directions. For example, the two pictures in Figure 2 show D4B's gaze directed at her task sheet during the test and task interactions. Comparing students' gazes during the test and task interactions enabled an initial judgement about the students' gaze directions.

#### 3.3.3 Persistent gaze in one direction

During MNEs, participants exhibited two main gaze directions: at their peer's video image and at their task sheet. A gaze directed at the task sheet could be readily identified because students were asked to put the task sheet on the desk, and they had to look down to gaze at the task sheet. The main difficulty lay in distinguishing whether a student who is looking up at the screen is looking at her peer's video image or at another part of the screen.



**Figure 3:** D1A's gaze directed at peer's video while talking.

Although all students were asked to take the gaze direction identification test, they were still able to make changes during their task interactions. For example, many students closed the peer's video window in the middle of the screen, which moved their peer's video window automatically to the default location in the top right-hand corner of the screen. This prevented the author from identifying their gaze direction in task interactions by comparing it with the student's gaze in the test. However, if a student used a persistent gaze in a particular direction, especially when talking to their peer during negotiated interactions, and their gaze was either directed at the middle of the screen or at the top right-hand corner, it was highly likely that their gaze was directed at the peer's video image. This method is based on the eye-mind hypothesis that a gaze at a certain time correlates to the focus of one's attention (Duchowski, 2003). For example, Figure 3 shows a screenshot of D1A's gaze while talking to her peers. The video recordings showed her looking in this particular direction repeatedly, especially while talking to her peers. D1A's gaze was also directed at the top right-hand corner of the screen in the pre-task gaze identification test. This consistent pattern suggests strongly that D1A's gaze was directed at D1B's onscreen video image.

#### 3.3.4 Video stimulated recall interview

Although the first two methods of identifying gaze directions can be highly accurate in most cases, it is still possible that a student looking at the top right-hand corner of the screen was looking at her own video image rather than that of her peer. This is due to the two participants' video frames covering a relatively small area of the screen and being displayed vertically adjacent to each other (see Figure 1). To resolve ambiguity, a third data source, VSRI, was used. VSRIs were conducted within two days of the last video SCMC session. These were one-on-one, face-to-face interviews where the author played segments of screen recordings to trigger a student's memory about specific MNEs and to ask specific questions regarding the direction of the student's gaze at certain points and her understanding of certain

words during meaning negotiations. The interview was conducted mostly in English, although some participants occasionally used Chinese to express their ideas clearly. Questions asked about the participant's gaze during negotiated interactions included, for example, "Do you care how you look in the camera?"; "Did you look at your own video image and/or your peer's video image very often during task interactions?"; and "What information (if any) can you obtain by looking at your peer's video?" VSRI data could be useful in the following three ways: 1) to confirm participants' gaze directions; 2) to clarify potential problems and determine whether students understood the correct meaning of the lexical item through negotiated interactions and 3) to investigate the reasons for their gaze direction choices.

To summarise, three approaches were combined to identify participants' gaze directions during negotiated interactions: 1) a pre-task gaze identification test; 2) persistent gaze in one direction, particularly while talking to the peer, and 3) the VSRI. When used in combination, these three methods offer strong evidence of the direction of a student's gaze, especially for identifying whether the student is looking at their peer's video image or not.

## 3.3.5 What episodes are included?

The above-described procedure generated sufficiently compelling data to lay a solid foundation for the data coding process as presented below.

Before coding and analysing the students' gaze data, it is important to define the episodes included in the analysis, as these could directly influence the results of the analysis. In this study, 37 MNEs were identified in video SCMC interactions. These include both successful (the respondent reached the correct understanding) and unsuccessful episodes (the respondent did not understand the meaning of the negotiated word), which may be incomplete. Analysing all episodes offers a fuller picture of the students' gaze and its relationship to the meaning negotiation results. One anomaly worth mentioning is that Dyad 2 did Task 6 instead of Task 5 because, on the day of class, one student could not find the Task 5 sheet. This has no significant influence on the analysis because both Tasks 5 and 6 were of the same type and had a similar level of difficulty; they only differed regarding the specific lexical items used in the negotiated interactions.

## 3.3.6 Coding with ELAN

ELAN is computer software offering a professional tool to manually and semiautomatically annotate and transcribe audio or video recordings. It has a tier-based data model that supports multi-level, multi-participant annotation of time-based

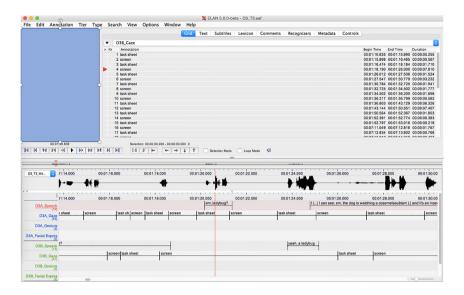


Figure 4: The interface of ELAN multimodal annotation process.

media ("ELAN\_Software," 2020). In this study, four tiers were used for each component of multimodal communication: speech, gaze, gesture and facial expression. For ease of coding, the analysis focuses on the two participants' video frames, instead of the whole video conferencing interface. Furthermore, the gaze analysis focuses on the coding of gaze directions in MNEs. Figure 4 shows a screenshot of a sample ELAN annotation interface for coding participants' gaze directions.

The coding process involved watching the video frame by frame in ELAN, then selecting one of the four available codes for the gaze direction and attributing it to the relevant section of the video. Each frame had a duration of 0.013 s to ensure extremely fine-grained coding. The start and end times of each coded gaze and its duration were recorded and exported to Excel for statistical and regression analyses.

# 4 Findings

This section summarises the findings from the VSRIs and the statistical and regression analyses in relation to the two research questions, which were presented at the end of the literature review.

# 4.1 Confirmation of gaze directions from the video stimulated recall interview

Before doing any statistical calculation or analysis, it is important to verify the validity of the data by reporting the findings from the VSRIs.

As described in the methodology section, the main reason for using VSRIs was to confirm that students' gaze was directed at their peer's video image, particularly as opposed to at their own image. The VSRIs revealed that students rarely looked at their own video image and mostly focused on the peer's video image during task interactions. For example, when asked about whether she was looking at her peer's video image or her own, D2B said, "when we were very engaged in our talking/ conversation, I already forgot and ignored how I looked in the camera; I was completely focused on the talking". The VSRIs also helped the author clarify a student's screen interface and where their peer's video image was located, which is important for the accuracy of coding. For example, in Extract 1, D2A admitted that her peer's video image during the task interaction was not in the middle of the screen but in the upper right-hand corner. All other students also confirmed their gaze directions during the VSRI, allowing the author to code the gaze directions more accurately.

Gaze VSRI Extract 1: D2A's gaze direction confirmation and the SCMC interface

Researcher: and when you were talking to her, in her video images in the middle of the screen? D2A: it's not even middle, it's just like this (on the upper right corner of the screen), when I was talking to her I was trying to see her facial expressions, so I didn't focus too much on my video

In addition to confirming gaze directions, VSRIs assisted in clarifying potential anomalies during MNEs. For example, when watching the task recording and looking for MNEs, the author was confused when D1B was looking at the screen with a wide range of up and down eye movements, together with some hand movements that were accompanied by the sound of typing. When asked about this in the interview, D1B admitted that at that point, she had minimized the SCMC system window and was looking up an unknown word, "razor", online (Extract 2). Therefore, this MNE did not count as successful. Since the second research question concerns the relationship between the time interlocutors spend looking at their peer's video image and their success in meaning negotiation, the author needed to verify whether students arrived at the correct understanding of an unknown word through their visual and verbal interactions with their peers or in alternative ways, such as using an online dictionary.

## Gaze VSRI Extract 2: D1B's clarification of the anomaly

Researcher: so here, you got it from?

D1B: yeah, I looked it up

Researcher: from the dictionary? [online dictionary]

D1B: yeah

Moreover, participants were asked to share comments or preferences regarding the use of video for task interactions, which could offer an additional explanation as to why interlocutors used their gaze in the ways they did. In particular, their answers to this question could relate to how looking at a peer's video image may enhance their negotiation for meaning. For example, Extracts 3 and 4 respectively demonstrate D4B's and D1B's preference for video-based communication over audio-only interaction, because they could see their peers' facial expressions and gestures, guess their feelings/attitudes, pick up the right turn-taking points and judge whether or not they understood what had been said, all of which could ensure that their communication ran smoothly. The interview results show that all participants, except for D1A, preferred video to audio SCMC because they were able to obtain more multimodal information when looking at their peer's video image, which could promote their meaning negotiation process and make the communication smooth. This qualitative evidence can strengthen and triangulate the subsequent quantitative findings.

Gaze VSRI Extract 3: D4B"s comments on video SCMC

D4B: I prefer video. Researcher: why?

D4B: because I can, we can have eye contacts for better communication, gestures, smile face, and facial expressions ...... video, hmm, I am able to find out, according to her facial expression, if she understands me or not, if she has a problem or not, and if she wants to go on talking or not, and her attitude towards me, when she talks, yes, in this way, communication is better because it's more smooth, clearer.

Gaze VSRI Extract 4: D1B's comments on video SCMC

D1B: the benefits of video is, when you are communication, you can guess what is your partner's feeling or attitude according to her facial expression or her hand gestures

In conclusion, the findings from the VSRIs helped the author confirm where students' gaze was directed and clarify some potential anomalies, thus contributing to accurate coding and statistical analysis. These findings also indicated that

most participants preferred video interactions to audio-only interactions because they could gain more visual information while looking at their peer's video image.

## 4.2 Overview of the gaze data

Section 4.1 reported findings from the interviews. Section 4.2 addresses the first research question: How do interlocutors use their gaze during MNEs in video SCMC?

It must be stressed that students were under no time pressure at all to ensure that they could perform naturally during meaning negotiation in video SCMC. In total, the gaze analysis includes codes for 1,020 gaze directions by four dyads in all MNEs in video SCMC. The overall coded time is 3,663.585 s (1 h 1 min 3.585 s). Table 2 summarises each dyad's gaze data during the MNEs. On average, during interactions devoted to negotiating meaning, students spent more than half the time (53.09%) looking at their peer's video image (PVI), while 38.58% of their time was occupied by looking at the task sheet (TS) and only 7.26% was spent looking in other directions (OD). The "unidentifiable" (UI) category took up only 1.06% of their time and had no discernible influence on the results of the analysis.

## 4.2.1 Gaze time on the peer's video image

Table 2 demonstrates huge differences in gaze direction choices by different dyads in different tasks. For example, in Task 5, D3A spent 91.47% of her time looking at her peer's video image, and in Task 3, D4B did this for only 6.19% of the time. There also exist significant differences in the time spent looking at the peer's video image both within dyads and between different dyads. In comparing different dyads, it can be seen that Dyads 2 and 3 spent more time overall looking at each other's video images than Dyads 1 and 4 in both tasks. Within each dyad, there is no consistent pattern in gaze direction choices. The following cases emerged: 1) both participants preferred looking at their task sheet (e.g., Dyad 4 Task 5); 2) one looked more at the task sheet while the other chose to focus more on the peer's video image (e.g., Dyad 2 Task 3) and 3) both interlocutors looked at their peer's video image most of the time (e.g., Dyad 4 Task 3).

## 4.2.2 Gaze directed at the task sheet

Clearly, those dyads that spent less time looking at the peer's video image, such as Dyads 1 and 4, tended to spend more time focusing on the task sheet. For example, Dyad 4 in Task 3 spent more than 92.78% of the time looking down at the task sheet.

Table 2: Gaze data in meaning negotiation episodes in video SCMC.

Dyad_Task	PV1 (s)	%I/A	TS (s)	%S1	(s) QO	% <b>00</b>	(s) IN	%IN	Overall time (s)
D1A_T3	26.973	40.60%	37.846	26.97%	1.613	2.43%	0.000	0.00%	66.432
D1B_T3	14.848	21.97%	40.972	60.62%	11.772	17.42%	0.000	0.00%	67.592
Sum_D1_T3	41.821	31.20%	78.818	58.81%	13.385	%66.6	0.000	0.00%	134.024
D1A_T5	22.988	27.99%	50.128	61.03%	7.772	%97.6	1.245	1.52%	82.133
D1B_T5	22.686	27.53%	49.840	60.48%	9.883	11.99%	0.000	0.00%	82.409
Sum_D1_T5	42.674	27.76%	896.66	%9′.09	17.855	10.73%	1.245	0.76%	164.542
D2A_T3	94.843	68.52%	38.064	27.50%	5.519	3.99%	0.000	0.00%	138.426
D2B_T3	36.479	26.29%	84.115	%69.09	18.149	13.08%	0.000	0.00%	138.743
Sum_D2_T3	131.322	47.38%	122.179	44.08%	23.668	8.54%	0.000	0.00%	277.169
D2A_T6	646.524	72.49%	122.067	13.69%	85.598	%09.6	37.728	4.23%	891.917
D2B_T6	554.407	62.08%	262.147	29.35%	76.488	8.56%	0.000	0.00%	893.042
Sum_D2_T6	1200.931	67.28%	384.214	21.53%	162.086	%80.6	37.728	2.11%	1784.959
D3A_T3	51.513	20.68%	50.125	49.32%	0.000	0.00%	0.000	0.00%	101.638
D3B_T3	36.902	36.30%	63.512	62.48%	1.235	1.21%	0.000	0.00%	101.649
Sum_D3_T3	88.415	43.49%	113.637	25.90%	1.235	0.61%	0.000	0.00%	203.287
D3A_T5	225.856	91.47%	12.370	5.01%	8.695	3.52%	0.000	0.00%	246.921
D3B_T5	143.773	58.33%	86.933	35.27%	15.774	%07.9	0.000	0.00%	246.480
Sum_D3_T5	369.629	74.91%	99.303	20.13%	24.469	<b>%96</b> .4	0.000	0.00%	493.401
D4A_T3	$15.913^{1}$	7.08%	206.086	91.75%	2.627	1.17%	0.000	%00.0	224.626
D4B_T3	13.937	6.19%	211.066	93.81%	0.000	%00.0	0.000	%00.0	225.003
Sum_D4_T3	29.850	6.64%	417.152	92.78%	2.627	0.58%	0.000	0.00%	449.629
D4A_T5	7.866	10.06%	50.556	64.63%	19.798	25.31%	0.000	0.00%	78.220
D4B_T5	29.610	37.79%	47.734	60.92%	1.010	1.29%	0.000	%00.0	78.354
Sum_D4_T5	37.476	23.94%	98.290	62.78%	20.808	13.29%	0.000	0.00%	156.574
Overall sum	1945.118	53.09%	1413.561	38.58%	265.933	7.26%	38.973	1.06%	3663.585

In the same task, Dyad 2 spent less than half that time (44.08%) looking down at the task sheet.

In terms of task type, all four dyads focused more on their peer's video image in the problem-solving tasks than in the spot-the-difference tasks. For example, Dyad 2 spent 1,200.931 s (67.28%) looking at their peer's video image in Task 6 but only 131.322 s (47.38%) looking at their peer's video image in Task 3. Meanwhile, all dyads except for Dyad 1 spent more time looking at the task sheet in the spot-thedifference task (Task 3 or 4) than in the problem-solving task (Task 5 or 6). For example, Dyad 4 spent 92,78% of the time (417.152 s) looking at the task sheet in Task 3 but only 62.78% of the time (98.290 s) looking at the task sheet in Task 5. For Dyad 1, they spent only a slightly longer percentage of time looking at the task sheet in the problem-solving task (60.76%) than in the spot-the-difference task (58.81%). A potential cause of this different time allocation for different types of tasks might be the amount of information on the task sheet. Specifically, the spotthe-difference task sheet, which is a picture full of details to be described to the peer, is more information-dense than the problem-solving task sheet, which only has four items to be explained (see sample tasks in Appendix).

## 4.2.3 Gaze in other directions and unidentifiable gaze

Despite the different amounts of time spent looking at their peer's video image and the task sheet, most participants spent less than 10% of their time looking elsewhere, except for D1B and D4A, who occasionally consulted online dictionaries, looked for a pen or were interrupted by their surroundings. As for unidentifiable gaze, only 38.973 s out of more than 1 h of recordings were coded in this category, most of which was caused by blurry pictures due to limited internet speed.

# 4.3 Regression analysis and findings

Section 4.3 summarises findings relating to the second research question: What is the statistical relationship (if any) between the amount of time interlocutors spend looking at their peer's video image and their success in meaning negotiation in video SCMC?

#### 4.3.1 Variables

The aim of Section 4.3 is to establish through regression analysis the extent to which looking at an interlocutor's video image in MNEs can contribute to the success of meaning negotiation. Therefore, the X variable, or the predictor or

explanatory variable, is the total amount of time each dyad spent looking at their peer's video image during MNEs. It is important to stress that the time is the sum of time both students in the dyad spent doing this. This is because the interaction is mutual and the interactants cannot be separated. For example, if Student A is looking at Student B's image, but Student B is not looking at Student A's image, Student A can still see Student B's modes of communication other than gaze, such as nodding, smiling, frowning or sitting forward. Such multimodal information offers evidence for Student A to make judgements on whether Student B understands what she is talking about. In cases where both students are looking at each other's video image, they can both see each other's multimodal information and, additionally, can engage in indirect eye contact through the webcam, which offers them a further indication of each other's (non-)understanding of the negotiated lexical item.

The *Y* variable in this analysis is the number of successful MNEs. Table 3 lists the numbers of MNEs and successful MNEs in eight video SCMC tasks completed by four dyads. In this study, a successful meaning negotiation refers to an MNE in which the respondent (the student who did not initially know the meaning of the lexical item) managed to arrive at the correct meaning of the lexical item through meaning negotiation with their peer. According to this criterion, 15 of 37 MNEs were successful, including eight from Dyad 2, three from Dyad 3, four from Dyad 4 and none from Dyad 1.

## 4.3.2 Regression analysis results

Section 4.3.1 explained what the *X* and *Y* variables represent. The final data for these two variables are summarised in Table 3. With this set of data as input, Excel was used to calculate the correlation coefficient between the two variables, generate a trend line and equation (Figure 5) and conduct a regression analysis (Table 4). The following paragraphs report the statistical results of the regression analysis and explain their meanings in context.

Table 3.	The \	Y variable: the	number of s	urcessful	meaning ne	ontiation	enisodes

Dyad_Task	PVI (X)	No. of MNEs	Successful MNEs (Y)
D1_T3	41.821 s	5	0
D1_T5	45.674 s	2	0
D2_T3	131.322 s	2	2
D2_T6	1,200.931 s	9	6
D3_T3	88.415 s	3	0
D3_T5	369.629 s	6	3
D4_T3	29.85 s	8	2
D4_T5	37.476 s	2	2

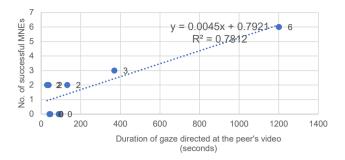


Figure 5: Linear regression between duration of gaze directed at peer's video image.

#### 4.3.2.1 Correlation coefficient

The correlation coefficient provides a basic method for measuring the degree to which two variables are linearly related. The correlation coefficient can range from -1 to 1, with 1 indicating a completely positive linear relationship and -1 indicating a completely negative linear relationship. It is also called the multiple *R* in a regression analysis. In the present analysis, the correlation coefficient is 0.88, which shows a relatively strong correlation. This indicates that the time students spent looking at their peer's video image is closely related to the number of successful MNEs. The next step of the regression analysis can further specify the relationships between these two variables from a statistical perspective.

## 4.3.2.2 R-squared

In a regression analysis, *R*-squared measures the extent to which a change in the *X* variable contributes to a change in the Y variable. In this case, a change in the duration of one's gaze directed at the peer's video image contributes to 78% of the change in the number of successful MNEs. R-squared serves to predict the likelihood of future events falling within the predicted outcomes. Therefore, another way of explaining the meaning of R-squared is that if we conduct the same experiment (video SCMC tasks) in a larger sample (with more dyads) in the same population, there is a 78% chance that this model (the equation/trend line in Figure 1) will be able to predict the number of successful MNEs. This is also a relatively strong prediction in statistical terms.

## 4.3.2.3 *p*-value

Another important measure is the p-value, which is used to determine the statistical significance of a hypothesis test. In other words, it indicates the extent to which the regression result occurs randomly. The smaller the *p*-value is, the more significant the result is. The accepted threshold for determining statistical

Table 4: The regression analysis results for the gaze analysis.

				Summary output				
Regression statistics	atistics							
Multiple <i>R</i> R Square Adjusted <i>R</i> Square	uare							<b>0.88386819 0.78122299</b> 0.74476015
Standard error Observations	ı							1.0260918
				ANOVA				
	df		25		MS		F	Significance <i>F</i>
Regression Residual Total	1 6 7	22	22.5578137 6.3171863 28.875	22.55 1.052	22.5578137 1.05286438	21.4251845	45	0.003582438
	Coefficients	Standard error	t Stet	<i>p</i> -value	Lower 95%	Upper 95%	Lower 95.0% Upper 95.0%	Upper 95.0%
Intercept X Variable 1	0.79206701 0.00445395	0.43167667 0.00096224	1.83486177 4.62873466		0.11620264       -0.26420774         0.00358244       0.002099437	1.84834177 0.00680847	-0.26420774 0.002099437	1.84834177 0.00680847

significance is 0.05. If the p-value is smaller than 0.05, it means we can state with 95% confidence that the regression analysis result is statistically significant. In this study, the p-value is 0.0035 – that is, much smaller than 0.05 and even smaller than 0.01. This means that the regression analysis result is statistically significant, because we can state with more than 99% confidence that the result did not occur randomly. The p-value can also be used to predict future events. This means that if we repeat the experiment (video SCMC tasks), it is highly likely (with more than 99% confidence) that we can obtain a similar result in the regression analysis. In other words, the chances of the sample data occurring randomly are extremely low (less than 0.01).

## 4.3.3 Interpretation of the findings

In response to the second research question, it can be concluded from the above regression analysis results that the more time students spent looking at their peer's video image during MNEs, the more likely their meaning negotiations were to be successful. Although the sample data are limited to eight video tasks, the regression analysis is statistically significant and can be used to predict the results of future/repeated experiments. It should be noted that the result is based on gaze analysis in MNEs only, rather than full video SCMC interactions, in line with the focus of the research question.

The *X* variable in this analysis is the actual amount of time (how many seconds) the members of each dyad spent looking at each other's video image. It is not the ratio or percentage of time spent looking at the peer's image as opposed to the overall time for meaning negotiation (which also includes time spent looking at task sheets and in other directions). Therefore, it does not matter how much time students spent looking at the task sheet or in other directions. As long as they spent more time looking at their peer's video during MNEs, they will likely be more successful in negotiating for meaning. The same regression analysis was carried out to examine the relationship between time spent looking at the task sheet and the success of meaning negotiation. Yet no statistically significant result was found between these two variables.

# 5 Discussion and conclusion

Exploring gaze in video SCMC interactions can offer insights into what role the visual mode plays in the technology-mediated communication environment. The key difference between the affordances of audio and video SCMC is the availability of the visual mode. In the absence of a shared physical communication

environment and limited visibility of bodily gestures and posture, gaze has become one of the most important sources of information in video SCMS interactions (Sindoni, 2014). This study mainly contributes to the research field by quantifying the time participants spent on different gaze directions, correlating this factor with the outcomes of meaning negotiation in video SCMC and identifying a statistically significant and positive relationship between the two factors.

On one hand, the findings of this study support the positive effects of video SCMC on meaning negotiation that Wang (2006) and Wang and Tian (2013) have observed. Some students' answers in the interviews also confirm Yamada and Akahori's (2009) findings that the webcam can facilitate online communication by reducing the interlocutor's anxiety and unease and by enhancing metacognition and comprehensibility. However, this positive finding about the webcam does not necessarily indicate a controversy among those researchers who have reported some negative or even disturbing effects of the webcam on SCMC, such as Lee (2006), Satar (2013) or Guo and Möllering (2016). As Wang and Tian (2013) point out, participants might have different levels of competence in using the webcam, possibly due to familiarity with their peers, their interpersonal and communication skills, their computer skills and other social and environmental factors. In the current study, participants were trained to use the web conferencing software, they became reasonably familiar with their peers and the online teacher during the preparation stage, and all had more than half a year of online learning experience. These factors might have contributed to their competence in using the webcam to obtain visual information by looking at their peer's video image and their preference for video SCMC as opposed to audio-only SCMC. Despite the disagreements on the webcam's effects on video SCMC, researchers widely agree that more training is needed for both online teachers and learners to further develop their competence in SCMC environments (Guo & Möllering, 2016; Lee, 2006; Wang & Tian, 2013).

On the other hand, this paper is intended to contribute to the methodological development and innovation of multimodal research in video SCMC contexts in several ways. First, the tasks were designed with many lexical seeds to elicit participants' negotiated interactions. These lexical items were carefully chosen to encourage and enable participants to exploit a wide range of multimodal resources to facilitate their meaning negotiation, for instance, "robot", "perfume" and "razor". Second, the methodological design of using hard copies of task sheets played an important role in the collection and interpretation of gaze direction data. As Chanier and Lamy (2017) argue, in video SCMC, meanings are constructed "through learners' physical relationship to tools ..., through learners' engagement with still and moving images" (p. 431). In the current study, hard copies of the task sheets were placed on learners' desks during all task interactions. This setting made it possible to clearly and easily distinguish whether a participant was

looking at the task sheet or the screen. Guichon and Wigham (2016) and Develotte et al. (2010) also highlight the physical elements of the communication context beyond "the screen's edge" (Jones, 2004, p. 24). In this study, the gaze being directed at the task sheet demonstrates that the physical set-up of the wider communication environment beyond the screen can have a substantial influence on how interlocutors negotiate meaning in video SCMC. Most importantly, the coding scheme and the triangulated gaze identification methods are key to determining the role of gaze in video SCMC. The coding scheme was highly suitable because it offered a good classification covering all gaze directions and it helped in answering the research questions. Moreover, three methods were used to identify and confirm different gaze directions, including a pre-task gaze direction test, constant gaze in one direction and a VSRI. The triangulated gaze identification methods enhanced the validity of the findings. Above all, the fine-grained frameby-frame coding procedure in ELAN ensured the accuracy of data analysis.

However, the research also had several limitations. For example, only eight adult female students participated in this study, and each dyad only performed two tasks in video SCMC. The statistical findings would have been more convincing and generalizable if the data had been collected on a larger scale in terms of the numbers of students and tasks. Moreover, the gaze identification and coding procedures were highly demanding and time-consuming, making it hard to replicate the research in exactly the same way. Finally, the paper exclusively focuses on participants' gaze directions and is, therefore, limited in providing a comprehensive picture of participants' multimodal orchestration in video SCMC interactions and revealing the relationships among different modes and semiotic resources, including linguistic output, facial expressions, hand gestures, etc. A related analysis of the overall multimodal analysis and findings from the same doctoral research project will be published in the future.

In conclusion, this study employed mixed methods to code and analyse gaze data in video SCMC and uncovered that the more time interlocutors spent looking at their peer's video image, the likelier they were to succeed in negotiation for meaning. Such a finding has implications for both future research and online teaching and learning practice.

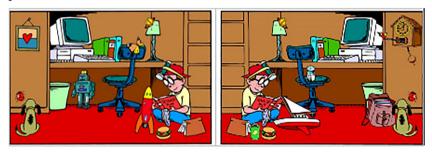
Research-wise, more studies on the role of gaze in video SCMC are certainly needed. In fact, the statistical relationship between gaze direction and the success of meaning negotiation possibly represents one of the first attempts to manually code students' gaze directions in video SCMC and to conduct a statistical analysis to explore the role of gaze in negotiated SCMC interactions. Therefore, this needs further examination in various research contexts and with different task designs. Future research could also investigate the effects of other factors on students' gaze performance, such as cultural background, language proficiency, level of familiarity with peers and motivation in the online video SCMC environment.

Teaching-wise, in response to the question raised earlier about whether online teachers and students should use the webcam or only audio conferencing, the statistical analysis result in this paper offers some evidence to encourage webcam use in synchronous web conferencing environments, particularly in online language learning through learner—learner verbal interactions. As demonstrated in the VSRIs, most participants preferred the use of a webcam because it can offer important visual information to facilitate meaning negotiation and smooth communication. Nonetheless, both online teachers and learners require more training to improve their ability to make proper use of different modes and semiotic resources in the multimodal synchronous computer-mediated environment.

# **Appendix**

# Task 3: Spot-the-difference task

You and your partner each have a picture. There are 7 differences in the two pictures. Please describe your picture to each other and find as many differences as possible.



Picture for Student A

Picture for Student B

# Task 5: Problem-solving task

#### Part 1:

You and your friend are trying to decide on some gifts for your homestay family in the UK. Your host family has four members: Mr. Jones (father), Mrs. Jones (mother), Billy Jones (son, 15 years old), and Mary Jones (daughter, 14 years old). Below are some items you and your friend have noticed while shopping at the Mall, which

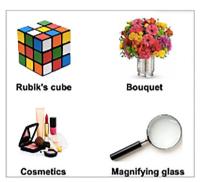
may make good presents. Your friend has been shopping at the Mall and has also seen some (different) things that he/she thinks might make good presents. Since the presents will be from both of you, you much decide together on one present for each family member (four total).

#### Part 2:

After you and your friend have decided on the four presents, discuss with your friend, estimate a reasonable price for each gift and calculate the total cost of these gifts.







Items for Student B

# References

- Argyle, M., Ingham, R., Alkema, F., & McCallin, M. (1973). The different functions of gaze. Semiotica, 7(1), 19-32.
- Chanier, T., & Lamy, M. N. (2017). Researching technology-mediated multimodal interaction. In C. A. Chapelle & S. Sauro (Eds.), The handbook of technology and second language teaching and learning (pp. 428-443). New Jersey: Wiley-Blackwell.
- Crawford, J., Butler-Henderson, K., Rudolph, J., & Glowatz, M. (2020). COVID-19: 20 countries higher education intra-period digital pedagogy responses. Journal of Applied Teaching and Learning (JALT), 3(1), 1-20.
- Develotte, C., Guichon, N., & Vincent, C. (2010). The use of the webcam for teaching a foreign language in a desktop videoconferencing environment. ReCALL, 22(3), 293-312.
- Duchowski, A. T. (2003). Visual attention. In Eye tracking methodology: Theory and practice (pp. 3-15). New York: Springer.
- ELAN Software. (2020). Wikipedia. Retrieved from https://en.wikipedia.org/wiki/ELAN\_software. Ellis, R. (2000). Task-based research and language pedagogy. Language Teaching Research, 4(3), 193-220.

Guichon, N., & Wigham, C. R. (2016). A semiotic perspective on webconferencing-supported language teaching. *ReCALL*, 28(1), 62–82.

- Guo, S., & Möllering, M. (2016). The implementation of task-based teaching in an online Chinese class through web conferencing. *System*, *62*, 26–38.
- Hampel, R., & Stickler, U. (2005). New skills for new classrooms: Training tutors to teach languages online. *Computer Assisted Language Learning*, 18(4), 311–326.
- Hampel, R., & Stickler, U. (2012). The use of videoconferencing to support multimodal interaction in an online language classroom. *ReCALL*, 24(2), 116–137.
- Huang, R. H., Liu, D. J., Tlili, A., Yang, J. F., & Wang, H. H. (2020). Handbook on facilitating flexible learning during educational disruption: The Chinese experience in maintaining undisrupted learning in COVID-19 outbreak. Beijing: Smart Learning Institute of Beijing Normal University.
- Hubbard, P., & Levy, M. (2016). Theory in computer-assisted language learning research and practice. In F. Farr, & L. Murray (Eds.), *The Routledge handbook of language learning and technology* (pp. 50–64). Oxford: Routledge.
- Jones, R. H. (2004). The problem of context in computer-mediated communication. In P. Levine, & R. Scollon (Eds.), *Discourse and technology Multimodal discourse analysis* (pp. 20–33). Georgetown University Press.
- Lamy, M., & Flewitt, R. (2011). Describing online conversations: Insights from a multimodal approach. In C. Develotte, R. Kern, & M. Lamy (Eds.), *Décrire la conversation en ligne: Le face à face distanciel* (pp. 71–94). Lyon: Presses Universitaires de Lyon.
- Lee, H. G. (2006). Meaning negotiation in task-based video conferencing. *The Journal of the Korea Association of Multimedia-Assisted Language Learning*, 9(2), 128–153.
- Long, M. (1988). Understanding second language acquisition. *Studies in Second Language Acquisition*, 10(1), 79–82.
- Long, M. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie, & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413–468). Academic Press.
- iiMedia Research. (2020). Business research survey on the online education for Chinese primary and secondary school students in the first quarter of 2020. Retrieved from https://www.sohu.com/a/384295245\_533924.
- Satar, H. M. (2013). Multimodal language learner interactions via desktop videoconferencing within a framework of social presence: Gaze. *ReCALL*, 25(1), 122–142.
- Sindoni, M. G. (2014). Through the looking glass: A social semiotic and linguistic perspective on the study of video chats. *Text and Talk*, 34(3), 325–347.
- Stockwell, G. (2010). Effects of multimodality in computer-mediated communication tasks. In M. Thomas, & H. Reinders (Eds.), *Task-based language learning and teaching with technology* (pp. 83–104). Continuum International Publishers Group.
- Van der Zwaard, R., & Bannink, A. (2014). Video call or chat? Negotiation of meaning and issues of face in telecollaboration. *System*, 44(2014), 137–148.
- Van der Zwaard, R., & Bannink, A. (2016). Nonoccurrence of negotiation of meaning in task-based synchronous computer-mediated communication. *The Modern Language Journal*, 100(3), 625–640.
- Varonis, E., & Gass, S. (1985). Non-native/Non-native conversations: A model for negotiation of meaning. *Applied Linguistics*, 6(1), 71–90.
- Wang, Y. (2006). Negotiation of meaning in desktop videoconferencing-supported distance language learning. *ReCALL*, 18(1), 122–146.

Wang, Y., & Tian, J. (2013). Negotiation of meaning in multimodal tandem learning via desktop videoconferencing. International Journal of Computer-Assisted Language Learning and Teaching, 3(2), 41-55.

Yamada, M., & Akahori, K. (2009). Awareness and performance through self- and partner's image in videoconferencing. CALICO Journal, 27(1), 1-25.

# **Bionote**

Chenxi Li The Open University, Milton Keynes, UK cecilia.lcx@outlook.com

Dr. Chenxi Li is a PhD graduate from The Open University UK. Her research interests include Computer Assisted Language Learning (CALL), Synchronous Computer-Mediated Communication (SCMC) through audio and video conferencing platform, negotiation for meaning patterns in SCMC, peer interaction in online English language teaching, task design for English language teaching, IELTS teaching and language assessment, English for academic purposes, and academic writing.