Chapter 5

Audio vs. Video Conferencing for Language Learning: Choosing the Right Tool for the Right Job

Natalia Barley

University of Illinois at Urbana-Champaign, USA

ABSTRACT

This chapter will trace recent trends in research on audio and video conferencing for language learning and teaching and outline the affordances and mediating effects of each platform that have communicative, socio-affective, and cognitive implications. On the basis of these characteristics, the author will argue that both audio and video conferencing have the potential to foster language learner interaction; however, neither medium is appropriate for all tasks, learners, and learning contexts. The focus of future inquiry should be driven by the need to better understand the complex interrelationship of interaction and the multimodal context in which it occurs. The chapter will conclude with considerations for selecting the platform most appropriate for task goals and learning setting.

INTRODUCTION

Ubiquitous access to networking technologies has transformed language learning by defying distance and providing rich opportunities for learner interaction in synchronous, multimodal environments. Oral synchronous computer-mediated communication (CMC) tools, such as audiographic (henceforth also referred to as audio conferencing or audio CMC) and video conferencing have been particularly promising as they support oral interaction. In the past two decades, research investigating oral synchronous CMC in the context of second language acquisition attempted to determine its suitability as a language teaching environment. In his work on promoting oral interaction in distance language learning, Rosell-Aguilar (2005) positioned the research inquiry into oral synchronous CMC in the context of the following question: “Whether audio or video conferencing is preferable for distance language learning is under discussion” (p. 418).

DOI: 10.4018/978-1-7998-6745-6.ch005
Communication in video conferencing may intuitively seem superior to audio CMC, as it most closely approximates face-to-face (F2F) environments. Its most salient characteristic is the interlocutors’ ability to see each other and observe non-verbal dimensions of communication, known to convey important, and, at times, critical information (Gullberg, 2010; McNeill, 2000). However, this assumption often neglects the fact that, along with its affordances, video conferencing introduces profound mediating effects on communication that may result in multiple unexpected consequences on language learner interaction. Similarly, in audio conferencing environments learners have to adapt to the unique characteristics and limitations of the medium. Audio CMC does offer some unexpected benefits, however, such as the security and relative anonymity often sought after by language learners for whom fear of making mistakes and threats to loss of face are a constant. Additionally, relying on single communication mode has been shown to put less strain on processing capacity, as it provides less information (Ferran and Watts, 2008; Hinds, 1999).

It appears that the question of medium preference for distance language learning may not accurately reflect the richness and the potential of each medium to contribute to interaction. This chapter adopts a sociocultural perspective with the focus on the mediating effects of technology on communication. A review of current literature will trace the recent trends in research in audio and video CMC and outline the affordances and potential limitations of each medium. On the basis of these characteristics the chapter will call for an alternative perspective highlighting the idea that each oral synchronous CMC medium has a place in language instruction and can potentially serve as an equally effective platform for language learner interaction, when its affordances are adapted to specific tasks, learners and contexts. Therefore, the focus of inquiry should be driven by the need to better understand the mediating effects of each mode within specific learning contexts with the goal to inform language teaching practice. The chapter will conclude with considerations for selecting the medium most compatible with instructional objectives, tasks, setting, and individual learner profiles.

**BACKGROUND**

CMC can be broadly defined as “any communication pattern mediated through the computer” (Metz, 1994). In language learning, CMC emerged in the 1990s when asynchronous text-based networking opportunities were first offered to language students (Lamy & Hampel, 2007). CMC then quickly evolved with the increasing availability of and easy access to a variety of digital media, such as hypertext graphics, sound, animation, and video.

There are two broad categories of CMC: asynchronous, such as discussion boards, blogs, and wikis, and synchronous, e.g., text-based instant messaging, chat, multiuser virtual realities, and audio and video conferencing. The key differences between the two categories are: 1) the need for participants to be online at the same time; and 2) the amount of time participants generally have to edit and revise their messages (Levy & Stockwell, 2006).

Early CMC environments were predominantly realized through one mode: written communication. Much of the early research reflected that trend by focusing on reading and writing skills, with listening and speaking significantly underrepresented (Jung, 2005; Liu et al., 2002).

It was characteristic for most of the early studies in CMC to focus their inquiry on comparing and contrasting the new technology with traditional, F2F settings. Many of the studies of that period found that learners reported an overall positive experience with CMC (Beauvois, 1994; Kern, 1995) and an
increase in learner participation (Blake, 2000; Chun, 1994; Kern, 1995; Warschauer, 1996). Students were also observed to produce more language in online discussions due to the decentralization of the instructor role (Chun, 1994; Kern, 1995). Other studies documented greater grammatical accuracy (Salaberry, 2000) and the same or greater syntactic complexity in learner output compared to F2F settings (Kern, 1995; Warschauer, 1996).

As CMC technologies became more widely integrated into language instruction, the research focus shifted toward psycholinguistic approaches that no longer sought to determine the value of CMC for language learning and teaching. Instead, the conversation centered on the effective ways of integrating technology in the language learning and teaching process. Much of the psycholinguistic-oriented research was guided by the Interactionist theory of second language acquisition (Blyth, 2008) which posits that language acquisition is promoted by interaction. More specifically, adjustments and modifications made during conversational interactions facilitate second language development by providing learners with opportunities to produce language, to receive comprehensible input and feedback on the correctness of their utterances, and to notice the difference between their interlanguage and target language forms (Mackey & Goo, 2012). When learners negotiate for meaning, they communicate their level of comprehension, interpret the interlocutor’s perceived comprehension, and then adjust their utterances accordingly until they are understood (Long, 1996). These negotiations offer both productive and receptive linguistic benefits by connecting “input, internal learner capacities, particularly selective attention, and output in productive ways” (Long, 1996, pp. 451–452). A large body of research has demonstrated the benefits of interaction for language development (e.g., Mackey, 2012; Mackey & Goo, 2012) in different contexts with various participant populations, providing empirical evidence for the benefits of interaction for learners from diverse language backgrounds. Similar observations were made in CMC environments. Learners actively negotiated for meaning in text-based synchronous CMC (Blake, 2000; Smith, 2003), audiographic conferencing (Yanguas, 2010) and video conferencing (Wang, 2006).

The latest CMC research adopted a sociocultural stance in response to the “social turn” (Block, 2003) in second language acquisition and the need to examine the socially constructed language learning process. Computer-assisted language learning and CMC research conducted from the sociocultural perspective has focused on the social, cultural, and discursive implications of using computers in language teaching.

One thread in sociocultural research is the exploration of mediating effects of technology on communication. The early view of technology relegated it to the status of an empty vessel: “This time, we are using the computer and the Internet as an empty transport medium like the telephone…The medium does not interfere with or impose itself on communication, it just lets through what is produced at the two ends of the line” (Jung, 2005, p. 13). However, as the understanding of CMC evolved, many scholars (Chun, Smith, & Kern, 2016; Lamy & Hampel, 2007; Levy & Stockwell, 2006) expressed strong disagreement with this view. As Levy and Stockwell (2006) insist, “Any discussion of CMC needs to take into consideration the effects of the computer on the communication that occurs through it as well as on the communication partners” (p. 84). In a similar vein, Hampel (2014) adds, “Adding a computer to communicate also means adding another layer of mediation through computer hardware, computer software as well as the particular architecture of the environment. …text and context become indistinguishable, with both being used to make meaning” (p. 5).

Many CMC environments available today incorporate multiple modes for making meaning: written and oral language, visual representation, audio, gestures, and space (Cope & Kalantzis, 2009). Both audio and video conferencing constitute multimodal environments. Most audio conferencing platforms offer
multiple features such as voice and text chat, interactive whiteboards, file sharing, etc. In addition to these features, video conferencing provides access to the webcam that allows participants to see each other.

A perspective that can shed light on how we make and interpret meaning through the diversity of communicative forms, including language, image, sound, gesture, touch, and smell (Lamy, 2012) is multimodality. Gunther Kress, a leading theorist in the field of multimodality, argues that meaning-making is not achieved through investigating each mode separately, but through interpreting all the integrated modes holistically (Kress, 2010). Each mode contributes a different dimension to one whole, thereby creating a richer meaning (Kress, 2010).

Language learning in any context is multimodal by nature. In a physical classroom, language learners have access to multiple semiotic resources, such as teacher’s and peers’ voices, facial expressions, gestures, written text, and images. The increasing availability of sophisticated CMC environments opens new possibilities for meaning-making and subsequently introduces changes in interactional conventions in multimodal online environments. This shift calls for renewed attention to multimodality and reenvisioning it in a new light (Guichon & Cohen, 2016). The intricate mosaic of modes offers novel possibilities of expression, but at the same time places additional demands on learners, and on language learners in particular. As language learners are in the process of mastering the spoken and written target language, its grammar, at the same time they face the challenge of learning the ‘grammar’ of other modes (Hampel & Hauck, 2006). The multimodal nature of oral synchronous CMC environments demands the knowledge and the skills to use multiple semiotic modes to derive and express meaning. Hence, it is of critical importance for learners in oral synchronous CMC environments to be well-versed in the simultaneous use of multiple modes to make meaning.

Oral synchronous CMC contexts are, therefore, complex multimodal environments that introduce multiple layers of mediation into the communication process. There is emerging research that probes these environments in an attempt to understand their mediating effects on the dynamics of interaction. The following sections will provide a review of the studies that have been conducted to date to better understand the limitations and affordances of audio and videoconferencing, make recommendations, and outline directions for future research.

MEDIATING EFFECTS OF AUDIOGRAPHIC AND VIDEO CONFERENCING

Oral synchronous CMC is now widely used in a range of language teaching contexts (e.g., Berglund, 2009; Cohen & Wigham, 2019; Satar, 2015; Yamada & Akahori, 2007). Audio and video conferencing are the tools of choice for the purpose of telecollaboration in European higher education institutions (Helm, 2015; O’Dowd, 2013). Many US universities also embrace the latest CMC technologies to offer quality online language education to students (e.g., Kern, 2014; Yanguas, 2010).

Studies on audiographic conferencing report on successful language learning and teaching experiences using the medium (Blake, 2005; Bueno Alastuey, 2011; de los Arcos & Sanchez, 2006). A number of studies have been carried out in the early 2000s in Lyceum, an audiographic conferencing platform developed by the Open University in the United Kingdom highlighting the potential of audio conferencing to support collaborative learning through social interaction (e.g., Coleman et al., 2010; Hampel et al., 2005; Hampel & Hauck, 2004).

Similarly, video conferencing was found to be a compelling platform for developing language skills through interaction (Wang, 2004a, 2004b, 2006). Recent studies have explored added affordances of the
webcam, such as access to non-verbal communicative resources that include facial expressions, gaze, gestures, and body movement, and the ways in which they contribute to successful communication and improved comprehension (Develotte, Guichon, & Vincent, 2010; Guichon & Cohen, 2014; Telles, 2009). Significant gains in fluency and comprehensibility were reported for learners participating in video conferencing sessions (Hung & Higgins, 2016; Saito & Akiyama, 2017).

Multiple studies consistently demonstrated the efficacy of interaction in both oral synchronous CMC environments, as language learners actively negotiated for meaning (e.g., Blake, 2005; Sarré, 2011) and demonstrated positive gains in second language development (Ziegler, 2016).

The global trend identified in current research indicates the overall efficacy of both audiographic and video conferencing for language learner interaction. It is difficult to determine whether one of the modes is more suitable than the other for specific tasks or contexts as direct comparisons of the two mediums are presently scarce. Intuitively, video communication may seem superior to audio CMC, as it allows interlocutors to see each other, which, in turn, gives them access to nonverbal cues, allows them to see how words are pronounced, and helps maintain social and emotional connection with each other. Now that videoconferencing platforms (e.g., Zoom, Microsoft Teams, GoToMeeting, Google Meet) are easily accessible, the question arises whether they should be the preferred choice over audiographic conferencing.

Interestingly, in their literature review on audio CMC, de Freitas and Neumann (2009) reference a number of studies (e.g., Gale, 1990; Heath & Luff, 1993) that do not provide any evidence that video-conferencing improves performance. Moreover, in his transparently titled book chapter “Visual Cues in Computer-Mediated Communication: Less is More,” Walther (2011) claims: “Research exploring the utility of visual images of participants in CMC has tended to reflect two trends. First, great consistency in users’ ratings of the desirability of video communication, and, second, no consistent support for the material benefits of the channel, despite its appeal. This discrepancy suggests that, in the truest sense of the word, visual cues in CMC are ‘overrated’ (p. 18).

To determine whether the question of medium preference can (and should) be answered, a better understanding of both the affordances and limitations of each mode is paramount. This section provides a detailed review of the studies that evaluate the mediating effects of oral synchronous CMC. These effects are broadly classified into communicative, socio-affective, and cognitive categories.

**Communicative Considerations**

As mentioned in the previous section, the key distinction between audiographic and video conferencing is the access to non-verbal means of communication. The importance of non-verbal communication in language acquisition has been documented by Gullberg (2010) and McNeill (2000), among others. When communicating, we use visual cues, such as eye contact and body language to establish and maintain joint attention. More specifically, gestures and gaze allow speakers to establish a common point of reference, to elicit help from an interlocutor (Bavelas & Chovil, 2000), and to signal attention to the speaker’s message (Argyle & Dean, 1965).

*Within the view of conversation as a joint activity… interlocutors establish successful communication by converging in their use of both linguistic forms and visual cues (gesture, posture, laughs, yawns) through what is known as interactive alignment. Just as alignment at one linguistic level can facilitate alignment at another linguistic level, convergence in visual cues, like eye gaze, can lead to shared interpretations… (McDonough, et al., 2015, p. 564)*
Gestures may potentially represent a particularly important resource for language learners. Dahl & Ludvigsen (2014), for instance, discovered that the availability of gestures had a measurable effect on language learner comprehension. Similar findings on the relationship between second language comprehension and gestures were reported by Kida (2008) and Sueyoshi and Hardison (2005). Gullberg, Roberts, & Dimroth (2012) found that gestures also aid attention and ‘noticing’ of language items, which have been demonstrated to promote language acquisition. “These findings suggest that gestures in input can serve to highlight linguistic forms, not just meaning, helping learners with the first essential learning task, namely to segment the incoming sound string” (Gullberg, 2010, p. 85). Additionally, gestures were found to facilitate interaction management functions, such as turn-taking (Streeck and Harteg, 1992).

Eye contact, or mutual gaze (Gale & Monk, 2000), is another powerful non-verbal resource that serves as a channel for both communicating and receiving information between the interlocutors. Eye gaze, as a component of interaction, is an emerging area of research in language learning. Some studies identify its value in facilitating meaning negotiation. For example, McDonough et al. (2015) discovered that eye gaze plays a role in the effectiveness of interactional feedback provided by language instructors when the second language speaker produces a non-target-like language form, i.e., makes a mistake. The researchers found that mutual eye gaze increased the likelihood of the language learner producing target-like (i.e., correct) responses. Gaze was also found to play a role in managing conversations by helping to either maintain or to end speaking turns, as well as to signal the switch between the verbal modalities (i.e., audio and text) in technology-mediated environments (Wigham, 2017).

Naturally, gestures, facial expressions, and gaze are not accessible in audiographic conferencing. This experience has been compared to interaction “in the dark” (de los Arcos & Sanchez, 2006, p. 78). To compensate for the lack of non-verbal language, speakers have to rely predominantly on verbal input occasionally augmented with emoticons available in text chat. Based on the non-verbal communication literature reviewed above, it can be anticipated that this limitation would present challenges in two areas: interaction monitoring and comprehension. The former refers to communicating and ensuring interlocutor’s understanding, managing turn taking, mode switching, and demonstrating empathy which is commonly conveyed through non-verbal means in F2F interactions. Indeed, some studies on language learner interaction in audiographic conferencing do argue that this may, in fact, be the case. When lacking video input, learners find it more difficult to tell whether they are understood unless explicit verbal confirmation of understanding is offered by the interlocutor (Hampel et al., 2005). Managing turn taking also becomes complicated. Participants are often unable to judge when a speaker has finished talking which results in either long silences or overlaps (de los Arcos & Sanchez, 2006; Guichon & Cohen, 2014; Hampel & Baber, 2003).

The second challenge is that of comprehension. A handful of studies compared comprehension rates in audio and video CMC arriving at conflicting results. Yanguas (2010) investigated Spanish language learners’ negotiation for meaning in F2F, audio, and video conferencing task-based communication. He found the amount of negotiation F2F and in video conferencing to be fairly similar, with a slightly higher number of negotiation patterns in audio conferencing, which initially suggested that audio conferencing could be more beneficial for language learning. However, upon investigating the outcomes of these negotiations, Yanguas found that interlocutors were able to reach complete understanding only 45% of the time in audio CMC, compared to 64% and 70% in videoconferencing and F2F conditions respectively.

In another study, Yanguas and Bergin (2018) also analyzed the effect of video and audio conferencing as well as task types on the number of language related episodes (LREs) in interaction in non-native speaking dyads. The authors proposed the following definition of LRE: “instances in the conversation...
Audio vs. Video Conferencing for Language Learning

in which learners turn their attention to the meaning, form, or use of a lexical item or grammatical form in the context of the L2 [second language] task being carried out” (p. 65). The researchers found no significant differences for the number of LREs for task or mode; however, similar to the findings of Yanguas’s 2010 study, a significantly larger number of unresolved LREs was found in the audio CMC. As the authors note, “… the AudSCMC group had more instances of non-communication than the VidSCMC group, which provided visual resources that could be used to aid the conversation” (Yanguas & Bergin, 2018, p. 76).

Guichon & Cohen (2014) arrived at somewhat different conclusions. They investigated the value provided by a webcam in an online second language teacher-learner interaction by comparing the audio and video conferencing conditions. The students attending audio CMC sessions reported that they understood the teacher slightly better than the students in video conferencing, although the difference was not statistically significant. Based on the study’s results, the authors concluded that “in a language learning and teaching online interaction, being able to see the image of the interlocutor and oneself during an video-conferencing interaction may in fact be distracting for some learners who, as a consequence, will be less focused on the verbal components of the teacher’s message, thus hindering understanding to some extent” (Guichon & Cohen, 2014, p. 349).

More recently, Barley (2020) also compared language learners’ interaction with a native speaking instructor in the two oral synchronous CMC conditions examining negotiation of meaning and comprehension rates. She found that learners actively and successfully negotiated for meaning both in audio and video conferencing with no statistically significant differences in the number of negotiation routines or comprehension rates.

The incongruity in the findings of these four studies can possibly be attributed to the differences in the type of interlocutor. Yanguas (2010) and Yanguas & Bergin (2018) documented interaction between intermediate Spanish learners who may lack the linguistic resources necessary to effectively maintain and repair conversation when communication breakdown occurs. When deprived of access to non-verbal language, they may struggle to compensate for it through the verbal channel. Conversely, language instructors not only have the language proficiency, but the experience necessary to sustain interaction with language learners and assist with conversation repair, which may have contributed to successful interaction in audio CMC in the other two studies.

Although video CMC allows for access to non-verbal dimensions of communication, it is important to remember that, just like any other aspect of communication in online spaces, it is mediated by the technology used. According to Hampel (2014), “there are fundamental differences, with communication in online learning spaces being characterized by an additional layer of mediation through the computer generally, the material characteristics and specific functionalities of the tool, and the specific combinations of modes that the tool provides” (p. 9).

First of all, unlike F2F settings where the speaker’s body is often fully visible, the camera frame captures either the user’s face only or the face and the upper body, limiting the view of the interlocutor (Guichon & Wigham, 2015; Hampel, 2014). The size of participants’ video window is another potential constraint. Depending on the platform interface, layout options, and the number of participants, the size of the video window can vary dramatically, from full screen to thumbnail view, potentially further limiting participants’ ability to see the speaker’s body language and facial expressions.

The effects of these constraints were documented by Guichon & Wigham (2015) who closely investigated communicative functions of gestures in video conferencing in interactions between French language learners and trainee teachers. The researchers noted that, during the examined interactions,
most of the teachers’ gestures were invisible to the webcam. Subsequent interviews with students showed that the visibility of mimics and gestures did have an impact on the perceived quality of the pedagogical interaction. When communicative gestures were visible and sustained long enough to be noticed by learners, they were more likely to contribute to mutual comprehension.

Another challenge faced by participants in videoconferencing is the inability to establish eye contact with each other. When speaking, users’ gaze is usually directed toward others’ image on the screen rather the webcam that is usually located at the top of the computer screen (Hampel, 2014; Satar, 2013). Lamy and Flewitt (2011) explored gaze in technology-mediated tandem learning documenting four types of gaze: looking at one’s own image, the camera, the chat window, and at the interlocutor. Follow up interviews with participants revealed that they felt uneasy looking at the webcam and that, when doing so, they could not establish eye-contact with the interlocutor.

Satar (2013) also examined gaze as one of teacher trainees’ strategies for using the webcam in a language learning setting. Similar to Lamy and Flewitt’s (2011) findings, the participants felt that eye-contact was practically impossible to establish in videoconferencing, at least with the technology used in the study. They preferred looking at their interlocutor, as it felt more natural and to avoid missing their partner’s visual feedback.

Compounding the challenges of oral CMC environments are technology and connectivity issues that continue to be reported by users (Rosell-Aguilar, 2005; Wang, 2004b) causing lag and transmission delays that disrupt synchronization between the audio and video feed. In some cases, disruptions result in loss of video and/or audio altogether. Kern (2014) described the effects of such desynchronization:

At moments when the delays become more pronounced, students sometimes wondered whether the desynchronized smiles, gestures, or facial expressions they saw onscreen were in response to what they were saying at that moment or whether they corresponded to what they had said a moment earlier. Tutors sometimes attributed a transmission lag to students’ hesitation. These ambiguities, combined with dropped frames, which resulted from bandwidth limitations and produced a jerky appearance to body motions, sometimes presented real challenges to understanding. (p. 348)

Unfortunately, even today, technology-related interruptions and connectivity concerns continue to plague online learning environments (Barley, 2020).

**Socio-Affective Considerations**

**Social Presence**

When mediated by communication technologies, language learner interactions are marked by social and psychological distance between interlocutors (Satar, 2013). The construct of social presence (SP) describes this phenomenon. Presented within Garrison, Anderson, and Archer’s Community of Inquiry framework (2000) alongside cognitive and teaching presence, SP was defined as “the ability of participants in the Community of Inquiry to project their personal characteristics into the community, thereby presenting themselves to the other participants as “real people” (Garrison et al., 2000, p. 89). More recently, Garrison, Anderson, and Archer (2010) added social and emotional connection among participants to their description of social presence. The markers of SP can be classified into affective (expression of emotions, use of humor, and self-disclosure), interactive (asking questions, continuing a thread, quoting
Audio vs. Video Conferencing for Language Learning

or referring to others’ messages, complimenting, expressing appreciation or agreement), and cohesive (using vocatives, using inclusive pronouns, phatics and salutations) (Rourke et al., 2001). Establishing SP is considered crucial for fostering interaction and learning in distance education settings (Satar, 2015).

Several scholars explored the degree of language learners’ SP in oral synchronous CMC. Yamada and Akahori (2007) explored the effect of interlocutors’ SP on language productive performance, more specifically, number of turns taken, number of utterances with target expressions, number of errors, etc. across four different conditions: text-based chat with and without interlocutors’ image, video conferencing, and audio conferencing. During the study, English language learners in each group engaged in task-based communication with their non-native speaking peers. The results revealed that the presence of the partner’s image increased the perceived consciousness of the partner’s presence. The combination of image and voice raised the learners’ consciousness of presence in a manner similar to F2F communication: “The ability of learners to see their partners’ images helped them to understand each other by allowing the use of non-verbal devices. … subjects using non-image systems tended to feel anxiety in communication, because of the lack of social cues such as nodding” (Yamada & Akahori, 2007, p. 57).

As far as language productive performance is concerned, video conferencing had “the greatest average number of turns taken, grammatical errors, self-corrections, utterances in the native language, and interruptions, as well as the shortest average interval between utterances by a significant margin” (Yamada & Akahori, 2007, p. 58). The authors hypothesized that, despite numerous grammatical and lexical errors observed in video conferencing, non-verbal behaviors helped subjects to speak the target language without frustration. The authors concluded that the medium’s affordances (video and voice) are motivating and conducive to communication.

In a similar study, Yamada (2009) confirmed the original findings. The presence of image and voice resulted in an increased number of utterances and self-corrections. She found that image and voice promoted consciousness of natural communication, whereas a text-based communication enhanced confidence in grammatical accuracy. Yamada noted that

> Videoconferencing which promotes social presence, promotes active communicative language learning from the viewpoint of the affective side. Moreover, the signs concerned with social presence such as social cues seem to be effective on reflective learning. In fact, learners tried to modify the grammatical errors in their utterances based on the facial expressions shown by their partners... This point suggests social presence leads to learning consciousness, enhancing learning performance as a result. (2009, p. 831)

In his investigation of the value of webcam in teletandem sessions, Telles (2009) also discovered that it offered a sense of self-confidence and security during conversations, and fostered proximity and intimacy. Similar observations were made by Develotte et al. (2010) who reported higher learner engagement and motivation and the sense that interactions were “more natural.”

Projecting SP in the absence of non-verbal cues presents a challenge for language learners. While interactive and cohesive SP indicators can be, and typically are expressed verbally, affective markers, such as expression of emotions, are more difficult to communicate without relying on facial expressions or gestures. In such cases, interlocutors’ communicative repertoire is limited to paralinguistic means, such as tone and prosody or the use of emoticons in chat.

It appears that availability of webcam may be more advantageous for establishing SP, however, limitations introduced by video CMC still present some challenges. According to Satar, “the DVC [desktop videoconferencing] environment could enhance trust, yet, for some participants it was still insufficient
to permit the generation of immediacy due to disembodied and limited representation, delays and distortions in audio and video and lack of eye-contact” (2013, p. 139).

**Foreign Language Anxiety**

Foreign language anxiety (FLA) has been one of the most studied obstacles to language learning since the term was originally coined by Horwitz, Horwitz, and Cope (1986). It affects F2F and online learners alike, and often manifests itself as “… negative self-comparisons, excessive self-evaluations, worry over potential failures, concern over thoughts of others” (Gregersen & MacIntyre, 2014, p. 1). These negative thoughts predominantly stem from learners’ inability to “communicate their authentic self,” which causes frustration and embarrassment (Côté & Gaffney, 2018, p. 2). As a result, the classroom activity that is most frequently reported to cause learner anxiety is language speaking practice (Horwitz et al., 1986). Highly anxious learners may find speaking activities very distressing. They may experience fear of negative evaluation by peers and teachers due to perceived deficiencies in communication skills and inability to create a proper social impression. This often results in learners’ reluctance to engage in communication using the target language, which, in turn, interferes with learning and performance (Horwitz, 2000).

In addition to specific classroom situations, there are a number of other factors contributing to foreign language anxiety. Certain personality characteristics, such as low self-esteem or trait anxiety, cultural and regional differences, and language learning motivation are some of those factors (Yan & Horwitz, 2008; Young, 1991).

Language learners may experience anxiety independent of instructional context. Yet, a number of researchers have attempted to find out whether online environments also have an impact on FLA. Multiple studies investigating written CMC allude to its potential to reduce anxiety, increase motivation, and encourage collaboration (Beauvois, 1994; Côté & Gaffney, 2018; Kern, 1995; Shield, 2000). Some research investigating synchronous audiographic conferencing reported that absence of visual cues and body language during oral interaction in the target language caused learner anxiety and lowered motivation (Hampel, 2003; Hampel & Stickler, 2005). For instance, Hampel et al. (2005) described a collaborative task completed by students of German in an audiographic environment and found that for some learners, “loss of embodiment” (p. 16) induced anxiety independent of their level of proficiency. Speaking to an unknown and unseen audience, inability to gauge others’ understanding, and feeling judged solely based on language skills seemed to increase student anxiety. The researchers also found it difficult to elicit input from “less linguistically able members,” who showed reluctance to speak in audio conferencing (p. 17).

However, the majority of studies on FLA have discovered affective gains in audio conferencing. Rosell-Aguilar (2005) and de los Arcos and Sanchez (2006) suggest that audiographic conferencing has the potential to foster a comfortable and safe environment for language learners, as it allows them to remain anonymous, feel disinhibited, and speak more freely. It has also been noted in CMC literature not focused on language learning that the perception of anonymity in CMC inspires a feeling of freedom from constraints and responsibility, reducing anxiety and increasing self-disclosure (Thurlow, Lengel, & Tomic, 2004).

Yanguas and Flores (2014) took a step further to compare learners’ speaking performance in a task-based environment and trait-like willingness to communicate (WTC), another affective variable interrelated with anxiety (Dörnyei, 2005) in F2F and audio conferencing. The authors found that participants
produced more language, i.e., took significantly more turns and produced slightly more words in audio CMC, adding, “[T]he significant increase in turn taking that resulted in the AudCMC task might indicate that participants were less anxious, were more motivated, or perhaps had a more desirable interlocutor” (p.94). Participants’ comments corroborated these findings. The learners mentioned a number of benefits offered by audio CMC mode: increased external motivation and ability to communicate, decreased anxiety, and language support offered by the mode (e.g., looking up words online).

The study by de los Arcos, Coleman, and Hampel (2009) launched a more detailed exploration of whether learner anxiety in an audiographic conferencing application occurs as a result of speaking the target language or peer/teacher judgements, and to what degree the characteristics of the online setting influence learners’ anxiety. The authors concluded that audiographic conferencing does not induce anxiety in learners “who understand the setting as a separate learning context with its own conventions” (p.15). After initial apprehension, learners felt more at ease knowing that they could not be seen.

Research into affective variables in video conferencing generally pointed at learners’ heightened anxiety levels. For instance, Lee (2007) discovered differences in learners’ perception of interaction with native speakers in video conferencing based on their personality traits:

Outgoing students found desktop videoconferencing stimulating and a worthwhile experience for becoming self-confident when interacting with native speakers. The experience made them realize how valuable it was to use desktop videoconferencing to practice their oral skills with expert speakers... Less confident students, however, felt that speaking Spanish with experts provoked a certain level of anxiety, especially when the experts were instructors. (p. 643)

In her investigation of webcam use in one-on one tutoring via Skype, Kozar (2016) reported that tutors and students found webcams useful during the initial tutoring sessions as they helped reduce social distance and establish rapport. Afterwards, however, the use of webcams diminished dramatically, as students reported apprehension about being judged on their appearance, overall discomfort with being seen, as well as privacy concerns. Cunningham et al. (2010) also observed that the students in online English for Academic Purposes courses were reluctant to use webcams (p. 169).

Loss of Face

Another intriguing socio-affective concept relevant to this discussion is loss of face. A study by van der Zwaard & Bannink (2014), seemingly unrelated to affective factors of oral synchronous CMC, investigated whether the mode of communication (instant chat or video calling) affects negotiation of meaning during interaction between native speakers (NSs) and non-native speakers (NNSs) of English in a task-based advanced second language classroom. The NSs were asked to tell jokes that were difficult for NNSs to understand and likely required some negotiation of meaning. The authors found that negotiation of meaning episodes occurred in both types of real-time interaction: instant chat and video calling. However, the most interesting discovery was that in video calling, there were quite a few instances of negotiation of meaning not occurring, even in cases when NNSs could not have understood the joke. In contrast, none of the chat participants pretended to understand the trigger during the chat sessions. They asked detailed questions about the source of misunderstanding until understanding was reached.

The authors concluded that video conferencing can potentially trigger loss of face for the second language learner compared to instant chat. “…in our study it was found that the NS-NNS task perfor-
mance through video call tended to be more face-appropriate than task-appropriate; trouble sources and potential triggers tended to remain fuzzy and unresolved. The intrusive webcam, registering, and transmitting image as well as sound, seemed to pose a threatening and daunting communication environment where issues such as politeness and potential loss of face thwarted successful task completion” (van der Zwaard & Bannink, 2014, p. 146).

A follow up study by van der Zwaard and Bannink (2016) again investigated the frequency of meaning negotiations in video conferencing and text chat. This time, however, the study participants completed two tasks: a culturally specific task (telling jokes) and a decision-making task. The findings were consistent with the first study. Instead of initiating repair sequences, non-native speaking participants frequently did not engage in negotiation of meaning despite nonunderstanding. This was more often observed in video conferencing than in text chat. According to the data gathered in the post-task questionnaire, this decision was made mostly in the interest of saving face rather than in the interest of the task.

The authors further report,

*In the most authentic of L2 [second language] learning situations, such as genuine telecollaboration projects between NNSs and NSs of the target language, sociocultural factors like fear of losing face may hinder and jeopardize task performance. For TBLT [task-based language teaching] inspired research and pedagogies this may lead to the paradoxical situation that key desiderata and principles of TBLT may hamper rather than encourage the very negotiation of meaning it hopes to foster.* (van Der Zwaard & Bannink, 2016, p. 637)

Emerging literature on socio-affective considerations in oral synchronous CMC reveals a complex relationship between multiple variables. Some learners find that the absence of non-verbal cues in audio CMC, such as gestures and eye contact, triggers discomfort and anxiety, whereas video conferencing approximates “natural” F2F communication, nurtures the sense of personal connection, and fosters engagement and motivation. On the other hand, remaining anonymous in audio CMC can be seen as liberating by other learners, who feel less anxious and apprehensive about making mistakes and are more willing to take risks. These learners seem to thrive in audioconferencing environments but become anxious when webcams are used. Interaction with native speakers may serve as additional source of discomfort and anxiety for language learners. In NS-NNS dyads, the need for face saving may potentially present itself as the dominant consideration that supersedes the need to ensure successful interaction.

**Cognitive Cost**

While learning can certainly be enhanced by access to diverse modes, that access may carry potential cognitive cost. The challenge of multimodal environments in general, and video conferencing specifically, pertains to the attention that is required from teachers and learners to process information provided in different modes. Wang (2006) warns that

*Interaction in a videoconferencing environment is similar to that of face-to-face but differs from it in important ways. For example, the combination of video, audio, text and other links enriches the learning environment, but at the same time can place considerable pressure and strain on the teacher as well as the learner. In such a multimodal environment, coping with the intensive demands of the technology and the learner is a new challenge to teachers.* (p. 139)
Similarly, Guichon and Cohen (2016) highlight the challenge of managing interaction across several modalities, particularly when tasks are put in competition in terms of the learners’ cognitive resources. “In a web-mediated interaction,” they explain, “not only do learners have to pay attention to their interlocutors’ multimodal messages (text chat, voice chat, webcam image), but they also have to divide their attention between several tasks (e.g., using the keyboard, checking the webcam image, accessing various documents)” (Guichon & Cohen, 2016, p. 515).

An added challenge in multimodal environments is the need to at times process conflicting information delivered via different modes. In their observations of video conferencing interactions between language teacher trainees’ and language learners, Develotte et al. (2010) noted that the trainees did not always use webcam images in ways that are complementary, i.e., contributed to the information delivered via the verbal mode. At times, the webcam image contradicted the oral message altogether. As a result, the webcam image was more distracting than helpful.

Language learners completing speaking tasks in video conferencing on mobile devices (Lee, Hampel, & Kukulska-Hulme, 2019) encountered a similar challenge. Some of the gestures caused several instances of non-understanding: “[I]f the meaning of the gesture did not accurately reflect the meaning conveyed in language then learners may have been confused, especially where specific language items were recycled but their co-occurring gestures altered” (Lee, Hampel, & Kukulska-Hulme, 2019, p. 36).

Even when additional communication modes are not available (e.g., text chat, whiteboard, etc.), research points to the increased costs of video-mediated communication compared to F2F communication (Ferrán-Urdaneta & Storck, 1997; Ferran & Watts, 2008; O’Conaill, Whittaker, & Wilbur, 1993). According to Ferran and Watts (2008), “Communicating via videoconference presents the challenges of difficult audio localization, turn-taking and conversation pacing, changes in cue salience, asymmetrical personal distance, and heightened self-awareness” (p. 1567).

These observations are supported by empirical research investigating the cognitive costs of video conferencing. Lang (1995) conducted a literature review on information processing in single- and multiple-mode media, which led her to conclude that single-mode puts less strain on processing capacity, as it is less complex and contains less information.

Hinds (1999) conducted three studies analyzing the cognitive and interpersonal costs of video. One of the studies examined the relationship between the cognitive load and two types of communication technology: audio-video (where an image of the communication partner and an image of self was visible) and audio only. Hinds found that interacting over video conferencing requires a greater cognitive load than interacting over audio only. She made the following observation:

*Communication technology may increase cognitive load by providing more social information. Such technologies can increase the amount of social information transmitted by increasing the number of channels (e.g., auditory and visual) through which information is transmitted and by increasing the actual volume of information transmitted on each channel (e.g., number of video images on the screen). For instance, audio-video communication relies on multiple channels (i.e., visual and auditory) and has the potential for conveying multiple messages over each channel. Audio-video systems allow for multiple images, including screens of one’s partner(s), oneself, and several screens of text or graphical information. Audio-only systems, in contrast, rely on a single channel and tend to use that channel for one audio message at a time.* (p. 286)
These observations highlight another benefit of teaching and learning in audiographic conferences: “By not having any visual information of fellow participants, there are also fewer distractions and one tends to concentrate more and listen carefully” (de los Arcos & Sanchez, 2006, p. 91).

It is important to be careful not to draw any conclusions at this stage, as the research in this domain is very scarce. The studies conducted to date appear to indicate that each additional mode may increase learners’ cognitive load. It would, therefore, stand to reason that video conferencing creates the potential for additional cognitive load on learners. A cautious approach would be to carefully consider the necessity of each mode prior to introducing it into the learning environment.

SOLUTIONS AND RECOMMENDATIONS

In an attempt to answer the question of whether audiographic or video conferencing is better suited for distance language learning, this chapter attempted to capture the key mediational features of each CMC environment.

The mediating effects of oral synchronous CMC on the communication process are most notable. Absence of webcam in audiographic conferencing negatively impacts conversation monitoring and, to some degree, comprehension. Interlocutors may struggle with turn-taking or making sure they are understood resulting in awkward silences or overlaps. Comprehension seems to be more affected in NNS-NNS dyads, as language learners, in the absence of non-verbal communication channel, may lack the linguistic resources necessary to effectively maintain and repair conversation when communication breakdown occurs. Even in video conferencing environments, technology heavily mediates non-verbal dimensions of communication: mutual gaze is impossible to achieve and many gestures often appear outside the webcam frame and, therefore, are not always visible.

Socio-affective implications of mediation are also significant. SP is easier to project in video conferencing but the limitations of the mode, such as disembodied representation, desynchronization of audio and video, and lack of eye-contact still require careful consideration and planning to establish trust and decrease the psychological distance between interlocutors. In audio conferencing this task is even more challenging, as affective markers of SP are difficult to communicate through emoticons alone or paralinguistic means, such as tone or prosody.

The effect of the medium on learners’ FLA levels is largely defined by their personality traits, general anxiety levels, and language proficiency with more anxious and/or less proficient learners preferring the anonymity and security of the audio conferencing environment. Interaction with native speakers may serve as additional source of discomfort for some.

Finally, the cognitive cost of each added mode is another important consideration. Processing information provided in different modes appears to place considerable strain on teachers and, especially on language learners. From a cognitive perspective, it may be said that less is more.

Each environment is clearly a part of an intricate mediational process interwoven into communication. The unique characteristics of each CMC mode introduce multiple communicative, socio-affective, and cognitive variables that have a profound impact on language learner interaction. However, to best understand and appreciate the complexity of interactional dynamics in oral synchronous CMC, other considerations also need to be taken into account. A number of factors, such as individual learner profiles, language proficiency, instructional tasks and setting, and the type of interlocutor appear to interact with the specific CMC mode properties and, as a result, further mediate communication within these environ-
Audio vs. Video Conferencing for Language Learning

ments. In other words, each instance of technologically mediated interaction is uniquely influenced by and adapted to its participants, instructional tasks, as well as the medium in which that interaction takes place. Therefore, it is not possible to identify any one medium as preferable for distance language learning, as neither audiographic nor video conferencing is universally suited for all tasks, learners, interlocutor types, and learning contexts. Each CMC medium has the potential to foster rich multimodal interaction when the affordances of each communication mode are carefully considered and selected to best support specific learners and instructional objectives. Thus, a more relevant discussion should be focused on establishing a better understanding of the complex interrelationship of computer-mediated interaction, the technological, instructional, and social context in which it occurs, and other pertinent variables.

To maximize the potential benefits of oral synchronous CMC for language learning, more research is needed to further investigate language learner interaction in oral synchronous CMC in various settings with different participants.

It is also important to support language practitioners in their efforts to deliver instruction in oral synchronous CMC environments. Teacher training should focus on building a repertoire of strategies for online teaching and learning. Multiple scholars (Develotte et al., 2010; Guichon & Wigham, 2015; Hampel & Stickler, 2012) call for online teachers to develop critical semiotic awareness and semio-pedagogical skills to mediate online interaction “by combining or dissociating modalities...that are adapted to objectives and to the cognitive requisites of the task” (Develotte et al., 2010, p. 293). It is, therefore, crucially important to design and develop tasks that take into consideration which combination of modes is most appropriate for the objectives of those tasks. Additionally, both teachers and learners alike need to develop strategies that overcome the absence of body language to 1) build immediacy and foster socialization; 2) support interaction monitoring; 3) improve comprehension. Acquiring these skills and strategies will help realize the full potential of oral synchronous CMC in distance learning settings.

FUTURE RESEARCH DIRECTIONS

The potential of oral synchronous CMC platforms for language learning is particularly relevant in today’s context, as many educational institutions have embraced online teaching. Moreover, at the onset of COVID-19 pandemic, distance education offered learning opportunities that would not have been possible otherwise. Therefore, the need to identify best practices in audio and video conferencing is now even more pressing.

Some studies have been conducted in the field, many of which were reviewed in this chapter. However, the field is still characterized by paucity of research (Satar, 2013; Wang, 2006). According to Yanguas and Bergin (2018),

Research on oral SCMC [synchronous computer-mediated communication] … is exceptionally scarcer than research on text-based SCMC… Oral SCMC research has only produced a handful of studies that have addressed a variety of topics in diverse contexts and from different theoretical perspectives. Therefore, drawing any conclusions on the use and efficacy of oral SCMC in CALL [computer-assisted language learning] becomes an arduous endeavor. (p. 68)

The review of currently available literature on audio and video conferencing allowed to draw some preliminary conclusions. Still, it leaves a number of questions unanswered. First of all, it is unclear how
and to what degree the lack of non-verbal cues impacts comprehension in audio conferencing. Given the fact that comprehensible input is considered a key component of language acquisition (Krashen, 1985), this question is of crucial importance. A related consideration is whether the interlocutor type (language learner vs. native speaker vs. language teacher) has any effect on the success of negotiated interactions, i.e., comprehension. Second, there is little research on strategies that can effectively compensate for limited or absent non-verbal communication channels. Finally, an equally important question is the effect of multimodal environments on language learners’ and teachers’ cognitive load, a domain that is currently unexplored.

Research directions that move us closer to gaining a better understanding of these phenomena will inform online language teaching practices and will guide educational practitioners in providing effective online language instruction.

CONCLUSION

The review of existing research indicates that audiographic and video conferencing are equally efficacious for supporting language learner interaction because both mediums appear to promote negotiation of meaning (de los Arcos & Sanchez, 2006; Rosell-Aguilar, 2005; Wang, 2006). The studies, however, revealed unique characteristics of each oral CMC environment that introduce multiple layers of mediation that, in turn, have communicative, socio-affective, and cognitive implications.

Most closely resembling face-to-face interaction, video conferencing is the only CMC medium that provides a channel for non-verbal communication (e.g., gestures, gaze, facial expressions, body language) that has been shown to support interaction management (Wigham, 2017), improve comprehension (Dahl & Ludvigsen, 2014; Kida, 2008; Sueyoshi & Hardison, 2005), and vocabulary acquisition (Kelly, McDevitt, & Esch, 2009; Tellier, 2010). However, videoconferencing was found to mediate interaction in a number of ways, for example, by introducing distortions, lag, limited view of the interlocutor and her gestures, and lack of eye contact. This limits access to non-verbal channels of communication and diminishes SP. As Satar (2013) explains, “Certain features of face-to-face communication, which enhance feelings of being physically together with others, such as eye-contact (oculesics), touch (haptics), and physical distance (proximity), are still not available via DVC [desktop videoconferencing]” (p. 123). Interaction in audio conferencing, which lacks non-verbal cues altogether, is characterized by reliance on verbal messages almost exclusively. This impacts conversation monitoring and, to some degree, comprehension leading to longer pauses and silences (Guichon & Cohen, 2014), and, in some cases, unresolved negotiation episodes (Yanguas, 2010). Issues with comprehension are more frequently observed in NNS-NNS dyads, as language learners, in the absence of non-verbal communication, may lack the linguistic resources needed for conversation repair. Interaction with language instructors, on the other hand, appears to have similar comprehension rates in audio and video conferencing.

The impact of the CMC medium on learners’ FLA is largely influenced by their personality traits, general anxiety levels, and language proficiency. More anxious and/or less proficient learners seem to prefer the anonymity of audio CMC because of its potential to create a comfortable and safe environment (de los Arcos & Sanchez, 2006; Rosell-Aguilar, 2005). Interaction with native speakers may serve as additional source of discomfort for some and threaten loss of face, particularly in videoconferencing.
Another important benefit of audio conferencing is its potential to offer reduced cognitive load, as processing information provided in different modes appears to place considerable strain on both teachers and language learners.

In the context of these findings, it can be concluded that neither CMC medium is universally preferred for all tasks, learners, and learning contexts. Therefore, neither is inherently “better” than the other for distance language learning. An alternative, and more productive perspective should highlight the idea that each CMC environment undoubtedly has a place in language instruction and can serve as an equally effective platform for language learner interaction, as long as the affordances and limitations of each medium are taken into consideration and each mode is selected to best match the specific learners, learning context, and instructional objectives. The discussion should be focused on establishing a better understanding of the complex interrelationship of interaction, the technological, instructional, and social context in which it occurs, as well as individual learner variables. In light of this, language educators should carefully consider the affordances and limitations of each medium before selecting the one that best suits their learners, instructional setting, and objectives. Harnessing that knowledge will allow to envision and design future learning environments that capitalize on the potential of each medium to create engaging and effective learning opportunities conducive to language development.

REFERENCES


Audio vs. Video Conferencing for Language Learning


Hampel, R., & Stickler, U. (2012). The use of videoconferencing to support multimodal interaction in an online language classroom. ReCALL, 24(2), 116–137. doi:10.1017/S095834401200002X


Audio vs. Video Conferencing for Language Learning


Audio vs. Video Conferencing for Language Learning


van der Zwaard, R., & Bannink, A. (2014). Video call or chat? Negotiation of meaning and issues of face in telecollaboration. *System, 44*(1), 137–148. doi:10.1016/j.system.2014.03.007


Walther, J. B. (2011). Visual cues in computer-mediated communication: Sometimes less is more. In A. Kappas & N. C. Kramer (Eds.), *Face-to-face communication over the Internet* (pp. 17–38). Cambridge University Press. doi:10.1017/CBO9780511977589.003


**ADDITIONAL READING**


KEY TERMS AND DEFINITIONS

Audiographic/Audio Conferencing: A conference between two or more participants by using computer networks to transmit live audio, text, and image data.

Computer-Mediated Communication: Any communication mediated through technology.

Foreign Language Anxiety (FLA): Apprehension experienced in learning or using a target language.

Negotiation of Meaning: A process during which the speakers communicate their level of understanding, clarify, or paraphrase their utterances until they reach clear understanding of each other.

Non-Native Speaker: A person who did not learn the target language as a child in a ‘natural’ setting.

Non-Verbal Communication: Gestures, gaze, facial expressions, body language.


Video Conferencing: A conference between two or more participants by using computer networks to transmit live audio, video, text, and image data.