



# Open Research Online

---

The Open University's repository of research publications and other research outputs

## Video conferencing and multimodal expression of voice: Children's conversations using Skype for second language development in a telecollaborative setting

### Journal Item

#### How to cite:

Austin, Nick; Hampel, Regine and Kukulska-Hulme, Agnes (2017). Video conferencing and multimodal expression of voice: Children's conversations using Skype for second language development in a telecollaborative setting. *System*, 64 pp. 87–103.

For guidance on citations see [FAQs](#).

© 2017 Elsevier

Version: Accepted Manuscript

Link(s) to article on publisher's website:

<http://dx.doi.org/doi:10.1016/j.system.2016.12.003>

<http://www.sciencedirect.com/science/article/pii/S0346251X16304110>

---

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's data [policy](#) on reuse of materials please consult the policies page.

---

[oro.open.ac.uk](http://oro.open.ac.uk)

System Special Issue: Telecollaboration

Article 7

Title:

Video Conferencing and Multimodal Expression of Voice: Children's conversations using Skype for second language development in a telecollaborative setting

Authors:

Nick Austin (corresponding author) Colegio Luso Internacional de Porto;

Regine Hampel and Agnes Kukulska-Hulme, The Open University.

Email Address and Contact Details:

email: [austinnj@hotmail.com](mailto:austinnj@hotmail.com)

Telephone number: (Portugal +351)931103514

Address: Rua S. João de Brito,

No. 47, 2<sup>a</sup> andar

4100 454

Porto

Portugal

# **Video Conferencing and Multimodal Expression of Voice: Children's conversations using Skype for second language development in a telecollaborative setting**

## **Abstract**

This article explores how voice is expressed in a telecollaborative project using Skype to connect two groups of primary age English language learners across two countries. Voice is understood as the ways in which language and other semiotic means are used for communication (Blommaert, 2008). This theoretical view frames the qualitative study into how voice is expressed materially involving tools such as verbal language, body language, technology, and the spatial and temporal dimensions within which the children's conversation happens. A methodology for analysing the video recorded data was developed using Scollon and Scollon's concept of geosemiotics. This method of analysis investigates how language is materially assembled through interaction with others in the physical world. The study shows that telecollaborative conversations create particular conditions which affect the ways children express their voice. The implications discussed in the conclusion have the potential to initiate wider discussion in the context of early childhood education and language learning concerning the importance of a multimodal perspective on how children express voice to support their communication when using video conferencing.

Keywords: voice, computer-mediated communication, language learning, video conferencing, social semiotics, telecollaboration

## ***1. Introduction***

The technologies now available to many schools facilitate the creation of partnerships between language learning classrooms across different countries, allowing students to experience learning in a different way to previous generations. Teachers are, therefore, challenged to forge new skills in language lessons by embedding intercultural dialogue and the development of children's use of information and communication technology (ICT) (Council of Europe Committee of Ministers, 2006; Department for Education, 2003). This means moving their practice beyond delivering face-to-face lessons in the classroom by incorporating computer-mediated communication. This can be done through telecollaboration, which Belz (2003) describes as involving 'the use of Internet communication tools by internationally dispersed students of language in institutionalized settings in order to promote the development of (a) foreign language (FL) linguistic competence and (b) intercultural competence' (68).

However, the use of communication technologies in education is changing the way we learn, and so our manner of viewing the classroom and how it operates must change too (Mahiri and Sablo, 1996; Levy, 2009; Dicks *et al.*, 2011). As computer-mediated conversations are becoming a part of children's learning experience in the classroom it is important to explore what the implications are for how children communicate in this environment. Online sites allow for different ways of interacting with a much wider community of learners and experts who can be in dispersed locations. An integral part of this shift in approach is the idea that children have a need to express themselves in a range of contexts and thus must be supported to learn with a sense of agency.

To explore this further, the concept of voice was employed within a social semiotic framework to find out more about how children communicate in an online telecollaborative setting using video conferencing technology. Conversations took place in Skype between

students from two primary schools in different countries who are second language (L2) speakers of English.

## ***2. The research questions and an overview of the literature***

To address children's use of voice in synchronous online conversations, the present study examined the following first research question: *How is voice experienced and expressed in a video conferencing environment?* This study begins with the supposition that voice conceptualises the way in which people produce meaning during online exchanges and, in particular, in video conferencing environments. We define voice as the ways in which 'people use language and other semiotic means in attempts (...) to make themselves understood by others' (Blommaert, 2008, p. 427). The individual character of a person's voice is transmitted through the choices they make over which signs highlight and portray those aspects about themselves that they wish to express. For if the speaker's voice 'is to become significant to others, he [sic] must mobilize his activity so that it will express during the interaction what he wishes to convey' (Goffman, 1959, p. 40). However, the process of voicing our ideas is complex and unpredictable because what is expressed is not necessarily perceived or understood. Reaching a shared understanding with others requires negotiation through dialogue, making the expression of voice an inherently social process (Bakhtin, 1986).

Interlocutors build on each other's ideas in order to get things done in the social world. Consequently, voice is seen materialistically as the conversion of socially meaningful resources into socially meaningful action. For a speaker's voice to carry meaning it must communicate something to others and therefore be intrinsically dialogic, incorporating elements of addressivity and responsivity to others in conversation (Bakhtin, 1986, p. 105).

However, despite the important role played by voice in the everyday activities in educational and online environments, the literature makes little mention of how children use

their voices to express themselves in video conferencing conversations. This qualitative study of online conversations between primary age learners, therefore, sets out to capture the interaction between communication means (such as language, gaze, gesture or artefacts), producers and users of those communications and the immediate context. It considers the children's expression of voice to be a multimodal accomplishment and shows how a multimodal perspective can help structure the analysis of children's voices as they engage them through the video communication service Skype.

A second, related research question asked: *What effect do the affordances of Skype have on how voice is expressed?* This question explored the ways in which the online environment mediates children's voices. Both Vygotsky (1978) and Bakhtin (1986) believed that the development of mental functioning in the individual is the result of learning conversations with others. Like Vygotsky, Goffman (1981) describes how the way in which people use the tools at their disposal (their bodies and other material means) in the presence of others supports collaboration with them. The distances that are maintained between people, the way in which gaze is used, the clothes that are worn, the responses they anticipate, how they interact with the physical spaces where people live all contribute to what they wish to say. People's bodies and objects from the material world become tools which, alongside verbal speech, can be used to signal the type of social role they are assuming and the actions that they will take. Communication in a conventional classroom happens face-to-face and is mediated through a range of semiotic tools including tasks, physical settings, institutional and cultural assumptions, time frames and language. By incorporating the use of internet voice communication, the material that we use to make meaning through interaction online is further expanded to include technology (Lamy and Flewitt, 2011). Kern (2014) observes that how we communicate in this environment is dependent on the ways in which our voice is mediated. The hardware and software through which ideas are expressed filter and transform

communicative activity influencing the choice of how best to convey those ideas (Hampel, 2014). This study follows the view of Develotte, Guichon and Vincent (2010) that Skype video conferencing software provides a new cultural tool that potentially restructures the way in which voices interact through a whole range of meaning making resources in new situations (see also Guichon and Cohen, 2014).

The final question was: *What role does voice have in helping children think together?* It explored how primary age L2 learners' voices engage to make meaning in this environment. As mentioned above, a view of development through child-led activity takes as its premise an understanding that learning happens through interaction with others. That learning is achieved in communication between contemporaries and across generations is of particular significance in language development as language is both the medium for learning and the focus of study (Hauck and Youngs, 2008). Interacting in meaningful contexts that build on young language learners' lived experiences, home languages and cultural frameworks helps them to flourish as the potential for drawing on different meaning making resources is expanded (Spencer *et al.* 2011).

Although Skype provides different opportunities for dialogue, we cannot assume that Skype conversations between children in their L2 necessarily lead to their cognitive development. Wegerif *et al.* (2004) highlight the importance of the particular surroundings in which conversations are embedded, but in the context of face-to-face communication. They assert that in a learning situation the style of interaction that is socially appropriate will fall into one of three broad categories. These are termed *disputational*, *cummulative* and *exploratory* talk (Wegerif *et al.*, 2004). Disputational and cumulative types are considered unproductive as they lack constructive engagement. Wegerif *et al.* (2004) suggest that the principal means for thinking together is through exploratory talk in which children critically

engage with each other's ideas in order to reach a mutually accepted understanding. The change in thinking that this process of finding agreement entails may be considered learning.

The three research questions and the theoretical perspectives that underlie them are central to the enquiry process. The emphasis of this study is not on interpreting the meaning of what is expressed by the children, but rather on understanding *how* meaning making resources are employed to express children's voices. Examining voice in this environment means paying attention to the use of non-verbal as well as verbal speech to make meaning. The words that children speak are considered just one of a range of representations of voice that provide the material means for communicating with others (Goffman 1981, Scollon and Scollon 2003). The approach taken by the researchers towards the research data must account for the multimodal co-orchestration of these different semiotic systems as they are used simultaneously by children to make meaning. These semiotic systems must be viewed in their context; transmitted through a computer programme, located on a computer within a room, all of which is 're-presented' to the children in conversation through the computer screen.

### ***3. Research participants and setting***

Data were collected from the online collaboration of primary students using the voice-over-internet protocol (VOIP) service, Skype. The conversations analysed in this study were conducted between two groups of twelve 6 and 7 year old children from different schools. One school was an infant school located in England with a cohort of L2 English speakers who have Urdu or Punjabi as their L1. The other was an international school in Portugal whose students are also L2 speakers of English but with Portuguese as their L1. The children met weekly to talk on Skype during an extracurricular lunchtime club. They were all volunteers who gave up their time for free play to practise their L2 in an exchange with each other. The volunteers in this study were of varying language ability and so the main focus of the tasks



carried out by the children was on making meaning rather than practising particular language forms. The study consisted of seven sessions, each lasting 30 minutes that ran between late September and early December 2012. Data from the pilot study is also drawn from in this paper. The pilot study similarly consisted of seven sessions carried out by children of the same age from the same schools. The pilot study served as a trial for the structure of the sessions in the main study. For each session, the children were given a task to provide a starting-point for the interaction, such as conducting a quiz. The study complied with requirements for conducting ethical research with children.

While evidence from across the data set is used to inform the analysis, this article focuses in the main on a 9'21" extract (micro segment) of a 27'31" minute Skype session (macro segment) that features a conversation between eight children, two of whom live in Portugal and six live in England. This session occurred at the end of the series of Skype meetings and represents exchanges between the students with the least researcher and teacher involvement. The data excerpt was considered the most likely to yield relevant information in answer to the research questions. Another reason for selecting this excerpt is that it relates to all three research questions, demonstrating a social event (documenting the ways in which voice is manifested through dialogue and its use in helping students to think together) and a computer-mediated conversation (inviting scrutiny of the ways in which different semiotic systems intertwine to make meaning), These dimensions shape the analysis of the data extract.

For the Skype session used in the analysis here students from the Portuguese school had decided to create a quiz for the English students. This task required reaching a shared understanding in order to explain and conduct the activities, a semiotic phenomenon of particular interest to this study as it challenges the children to convey their respective voices in a clear way for a real purpose. The students had the opportunity to bring items from home, include any aspect of the classroom environment in their conversation, adjust their seating

position in front of the computer, move the webcam or microphone and use the functionalities of the Skype conferencing system (emoticons, messaging and video chat) and their related affordances. It is important to note that while instant messaging was available to the children they did not use this function as typing text was too slow for them during synchronous conversation. As this research is interested in identifying the different elements used to make meaning, it is crucial to know what choices the children made between different semiotic modes (words, images, sound or movement) to constitute their voice in this environment.

#### ***4. The approach to the data***

##### ***4.1 An analytical framework***

The focus of this study on how individuals employ multimodal resources to express their voice in an online conversation fits the theoretical emphasis of multimodal interactional analysis. Multimodal interactional analysis stands apart from other approaches to multimodal data through its emphasis on the notion of context and situated interaction which places the focus of analysis on what individuals express and react to in given situations; this interaction is seen as co-constructed between members of a conversation (Scollon and Scollon, 2003; Norris, 2011; Jewitt, 2009). A useful framework for this purpose is Scollon and Scollon's (2003) concept of *geosemiotics*, which provided the three main categories that were used to frame the analysis in this study. Geosemiotics brings together research from different areas (namely linguistic anthropology, social psychology, sociolinguistics, cultural studies, semiotics, visual anthropology, sociology and cultural geography) to systematically analyse how people express themselves materially in the world.

An interest in how voice is experienced and expressed has a logical starting point in looking at the ways in which people form social arrangements and produce social interactions among themselves. In a geosemiotic approach to communication this broad topic is termed

*the interaction order* (Scollon and Scollon, 2003; Goffman, 1981). Of particular interest to this present study are the notions of *singles* (a person who is alone in a social space among others), *withs* (two or more people who are seen to be together through their mutual focus of attention on each other) and *platform events* (a person performing for others who watch) (Scollon and Scollon, 2003, p. 61-2). A major organising system in this category is the words that are spoken to each other by the interlocutors. Goffman, however, cautions that ‘it might be argued that children learn to respond with actions before they respond with words’ (1981 p. 40). The primary focus in this category is, therefore, on all forms of embodied communication and not just on language.

A second category is termed *visual semiotics* (Scollon and Scollon, 2003). This focuses on how meanings are produced through visual artefacts such as pictures, photographs, film, objects, writing and any other forms of sign that refer to something other than itself and exists independently of the interlocutors’ bodies. An interest of this study is in how visual objects mean what they mean because of where they are used, and the way they are used to communicate things to others in the world.

The third category in geosemiotic systems is called *place semiotics*. The broad array of meaning systems which fall into this category are those not located in the communicators themselves or framed in artefacts. This order examines the significance of the place in which the conversational event occurs and how space is used within it by the conversationalists to give meaning to the semiotic resources they employ.

These three categories do not exist independently of each other in social action. How different resources, described in the three different categories, express meaning together in the material world is termed their indexicality (Scollon and Scollon, 2003). The indexicality of different semiotic resources will vary in meaning depending on the context in which they are used. For example, pointing a finger can mean giving directions, an accusation, a threat or a

dance move depending on the way in which this resource integrates with resources from the other categories. In this approach highlighting how semiotic resources are indexed in the material world is the key to identifying how different resource selections relate to and affect each other within the composite whole of the multimodal text.

#### ***4.2 Using geosemiotic sub-categories for analysis***

Once this initial set of categories had been identified the complete data was viewed several times with each of the three categories in mind. The purpose was to find and mark those places in the data where the most salient evidence of each category could be found. At this point the level of interpretation was limited to the question of whether the information related to the category.

In the following phase of analysis the data relating to each category were processed. The objective was to summarise the large chunks of data so that they could be more easily organised. Because of the need to develop consistency in analysis and annotation it was necessary to divide the three categories into sub-categories. Scollon and Scollon's (2003) framework for geosemiotics provided these sub-groups. Thus, the 27'31" minute macro excerpt of the data chosen for more detailed analysis was annotated according to each of the nine sub-categories discussed below. Appendix A shows an excerpt from the multimodal coding chart used to analyse the data.

##### ***4.2.1 The interaction order***

The category addressing the relationship between the embodied actions of the speakers (the interaction order) was further divided into four sub-categories. Each of these sub-categories represents the main resources for making meaning in the interaction order. The first is *the sense of time*. A person's sense of how quickly or slowly time is passing is attributed to either the urgency with which they want something done or the extent to which they focus on a task

(monofocal or polyfocal activity) (Scollon and Scollon, 2003, p. 50). Signs of impatience such as tapping or repeatedly glancing at a clock are examples of how someone's sense of time manifests itself through their embodied actions.

The second sub-category accounts for the ways in which *space* can be perceived and invoked through embodied action (Scollon and Scollon, 2003, p. 52). Squinting or shading the eyes with a hand, for example, can invoke a sense of visual space. Different actions index different perceptual spaces. It is to these different semiotic zones that we look for the interpretation of their meaning.

*Interpersonal distance* is the sub-category which refers to the space that separates people in a social place (Scollon and Scollon, 2003, p. 54). Intimate distance indicates touching to very close proximity. Personal distance is the distance in which we feel obliged to begin some kind of social interaction to either acknowledge or ignore the person in this space. Social distance suggests a space in which the presence of others is acknowledged without needing interpersonal engagement, for example the distance between the teacher and a student at the back of the classroom. These spaces between people index their different relationships with one another.

*The personal front* is the fourth aspect of embodiment that constitutes the interaction order. As Scollon and Scollon (2003, p. 57) define it, the personal front is virtually any visible or perceptible aspect of a person that gives meaning to others in a social situation. For the purposes of categorisation in this present study the definition of the personal front focuses on what embodied aspects of communication children do, or do not, bring to focal attention through dialogue. This quality of selecting what we pay attention to is termed 'civil inattention' by Goffman (1981). Goffman's concept shows how we are able to make sense of the busy and complex array of discourses present in everyday environments such as a

classroom by prioritising certain resources (e.g. the teacher standing at the front of class to speak) over others (peer talk, classroom signs and so on).

#### **4.2.2 Visual semiotics**

The broad typology that examines how the interaction order is represented through ‘disembodied’ resources such as images and signs is visual semiotics. The broad category of visual semiotics was further divided into three sub-categories to describe the data.

*Modality* is based on the linguistic idea of modals which modify statements to give them a greater degree of truth or credibility (Scollon and Scollon, 2003, p. 89). With respect to visual semiotics modality is the degree of authenticity that can be attributed to an image or sign. The extent to which an image has been modified beyond what is conventionally considered a naturalistic state provides information about the discourse that might take place. An example might be the oversized lettering and primary colours of a child’s writing on a whiteboard that indexes a different context of use for the word than if it were printed in the page of a book.

Where action, objects, signs and images are located within a frame such as a computer screen affects the meaning that they express. *Location* as a category identifies two basic information structures, centred and polarised (Scollon and Scollon, 2003, p. 92). A person located in the centre of the computer screen, for example, is given more attention and so is more able to express meaning than someone on the periphery.

The final sub-category used in this current study as part of the visual order is termed *text, image and/or object participants*. The discussion above suggested that the expression of voice is an inherently social process incorporating elements of addressivity and responsivity in relation to speakers communicating with each other (Bakhtin, 1986). There is always a dynamic dialogicality in play among signs that this category attends to. Of particular interest in the analysis are the ways in which objects, signs and images are made more or less salient through conversation.

### ***4.2.3 Place semiotics***

*Place semiotics*, the third broad typology, turns our focus away from the actions and activities of the conversationalists to investigate the places in which voice is expressed. The concept of *physical space* considers whether a location has semiotic systems, and if so, the kinds of discourses that these meaning making resources might be put to. Signs are situated because they reflect the physical environment in which they are placed. In a classroom that includes a display of high frequency English words on the wall, one might expect children to voice ideas about learning.

Another important aspect of the location in which conversations happen is the way in which they are organised to reflect the interaction order. How the material world intersects with the different ways of being together socially is accounted for by the sub-category *social context* (Scollon and Scollon, 2003, p. 169). One might imagine that the space within the classroom walls, the type of furniture and how it is laid out will affect the social use of the room and shape what is expressed there.

### ***4.3 Issues of representation***

Video lends itself to repeated viewings of an event and would appear to represent the complexities of multimodal dialogue. Wolfe and Flewitt (2010), for example, argue that while questionnaire and interview data can offer broad insights into practices around technologies, case study video data reveals the multimodal detail involved in computer-mediated interaction. The suggestion is that multimodal expressions of voice comprise a complex orchestration of different semiotic resources. Each of the semiotic resources relates to and affects each other within the composite whole of the multimodal text (Baldry and Thibault, 2006, p. 18-19). These meaning systems function together to create a *multiplying effect*

(Lemke, cited in Baldry and Thibault, 2006, p. 18) on the meaning made from the text in ways that are not predictable through discrete analysis of the individual modes.

However, the choices the researcher makes in representing the data will ultimately influence its interpretation. In recognition of this Wolfe and Flewitt (2010) highlight the underlying need for the development of robust frameworks for the analysis and representation of events when using visual media for data collection and analysis.

Appendix B shows how the data were represented in this study. Rather than using a conventional way of transcribing verbal language only, a matrix was chosen that shows the simultaneity of language, gaze, movement and actions through their horizontal positioning. The transcription incorporates the temporal sequence of a Skype conversation in the leftmost column. Time, therefore, becomes the principle around which all other information is organised. Following a similar framework to Baldry and Thibault (2006), the table also has screen shots inserted into the left hand column, representing the continual visual sequence as a series of sampled still images. The transcription favours a visual representation of the data by placing it on the left-hand side, reflecting an emphasis of the multimodal interactional analysis on the communicators. As the focus of this study is on how children express and experience voice through Skype in the classroom, attention is placed not only to the primary involvement of the children (their interaction on screen), but also to their potential secondary involvements (with other class members, objects or audio phenomena for example) from the off screen and on screen surroundings in which their conversations are embedded. Multimodal interactional analysis thus adopts a polyfocal perspective.

### ***5. Findings from the data and discussion***

A geosemiotical approach to analysis of the data from the study was used to identify how interactional, visual and place orders manifest children's expression of voice in a video

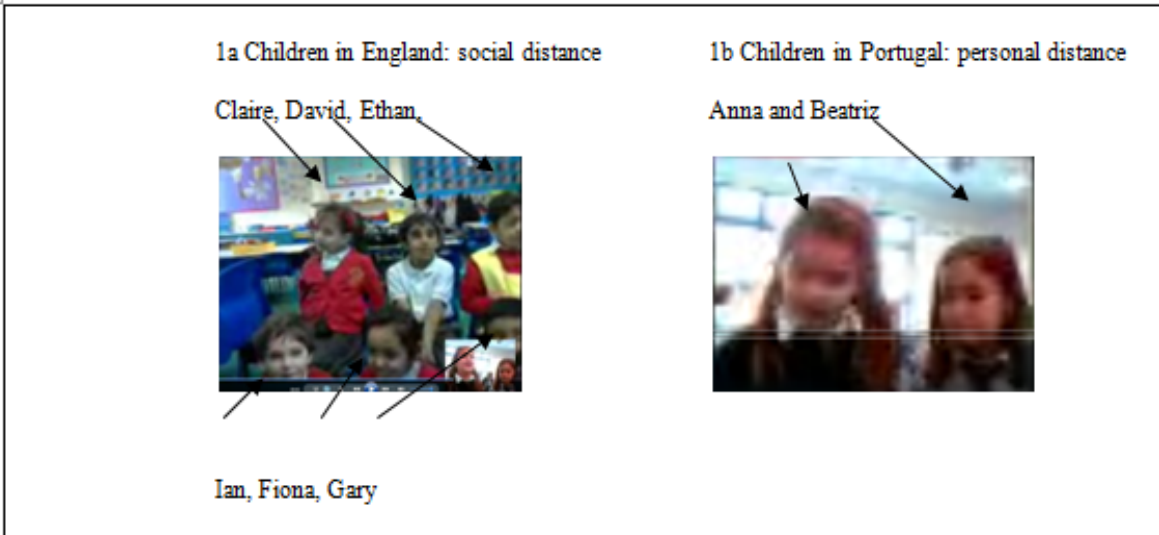


conferencing environment. The key findings from the analysis of the data will now be described as they relate to the research questions posed by this study. For a more detailed presentation of the findings see Austin (2015).

**5.1 How do children experience and express voice in a video conferencing environment?**

Webcam mediated online conversations create their own particular sets of conditions which affect the ways in which children are able to use resources in the interaction order to express their voice. A particular point of divergence from familiar patterns of communication is the way the children used the classroom space to negotiate different ways of being together.

In the data the distance represented by the webcam for the children in England would place the Portuguese children at a *personal* distance (18 inches to 4 feet from the respondent). In a face-to-face encounter, at this distance a person would be within what we sense to be our personal space so we feel obliged to engage them in some kind of social interaction (Scollon and Scollon, 2003, p. 54) (see Figure 1).

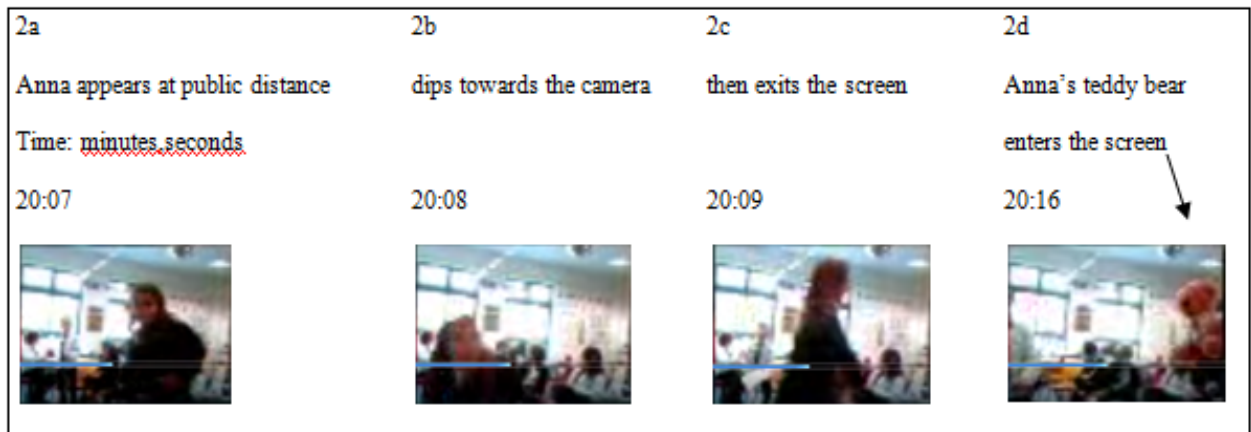


**Figure 1. Interpersonal distances represented through the webcam**

In a face-to-face conversation the participants' experience of interpersonal distance would be the same as each other's. In a webcam mediated communication this is not necessarily the

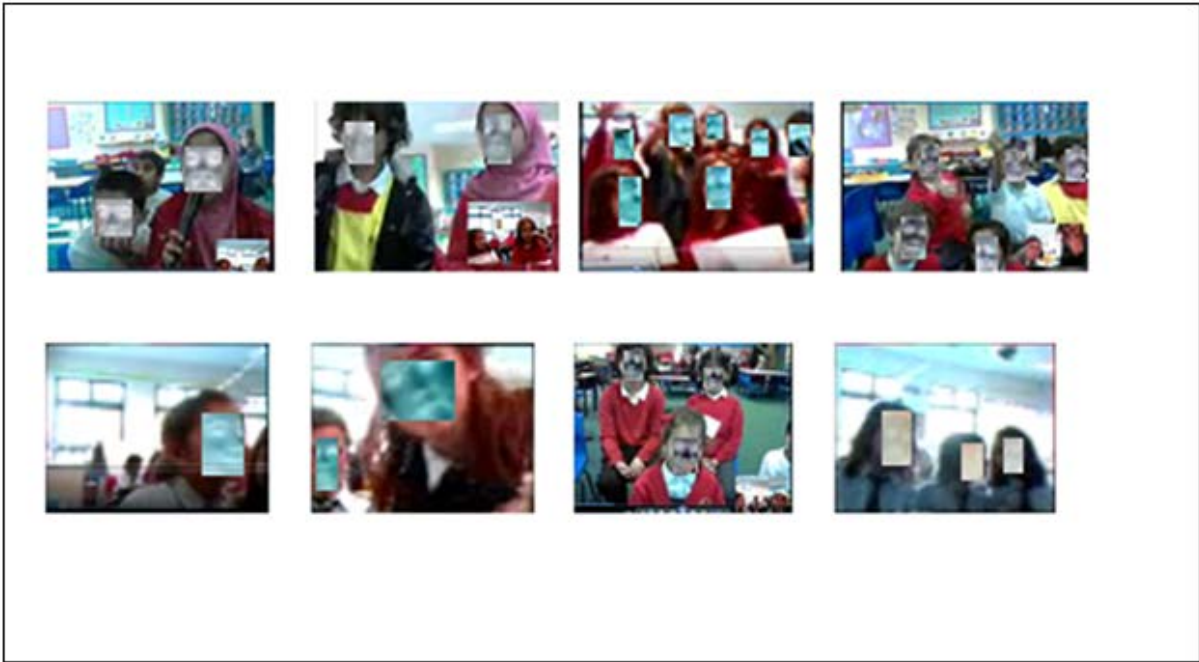
case as the space perceived by an interlocutor depends on the distance their conversational partner is from their webcam. This means that a speaker can affect the degree of social space that is represented to their partner, but they are unable to directly influence the distance at which their partner appears before them. Different interpersonal distances demand different types of behaviour from the people that experience them. This can be appreciated in the data from the roles assumed by the students in the exchange. Both Anna and Beatriz are in a position to engage their listeners directly while the six English children in Figure 1 are members of a group and subject to the dynamics of group interaction. Thus the ways in which interpersonal distances are mediated through the webcam can be seen to directly affect the way in which voice is expressed in this environment.

The represented interpersonal distance between the students is altered approx. 3 minutes later in the data. At this point Anna moves out of view from the capture of the webcam into the *backstage* area of the Portuguese classroom. She then reappears on the screen (*frontstage*) at a *public* distance of 12 to 25 feet only to disappear 4 seconds later (see Figure 2 parts a-c). The represented change in the interpersonal distance allows for different types of discourse to potentially enter into the exchange. Anna then introduces a Teddy bear into the screen shot at social distance from the children (see Figure 2 part d). If social distance provides the opportunity for interpersonal engagement without the obligation to necessarily do so, the bear might be considered an offer of further social interaction. The ways in which social distance is framed and represented by the computer thus becomes a semiotic tool, a part of the orchestration of meaning making resources used by Anna to express what she wishes to bring into the conversation.



**Figure 2. Screen shots of Anna**

Through the webcam the children were able to simultaneously manage visible (on screen) and hidden (off screen) areas of the video conferencing forum and – by adjusting their proximity to the camera – explore different representations of interpersonal distance. The freedom to move between these different spaces gave them access to different ways of unfolding their actions in relationship to each other. Through taking on diverse social role performances such as questioner, respondent or partner the children were able to play out their conversational aims. The ways in which dialogue was allowed to evolve was dependent on movement between these spaces and this movement was in turn dependent on the children's ability to index different ways of being together through their embodied actions (see Figure 3).



**Figure 3: Children use different representations of interpersonal distance to combine their voice with, or separate it from others, in different conversational roles**

Multimodal resources are thus seen to serve a variety of speaker and addressee purposes in the data. These functions include indicating different social spaces, managing roles in the conversation, indicating a lexical gap, enhancing the understanding of spoken language and representing something that cannot be voiced through spoken language by the child. Visual modes of expression are integrated with linguistic ones to manage and sustain the conversations.

A further example may be seen in Figure 4 where Violet and Wendy discuss a book. Representation of just the speech (see Figure 4, part a) from the data section fails to show movement between these spaces. In contrast, Figure 4, part b shows that frontstage and backstage activity is indexed in this section through embodied actions. In Wendy's view of Violet, Wendy is visible in the small screen located in the bottom right hand corner of each screen shot. The white square beneath her head is a book about Justin Bieber that she is showing to Violet.

a Excerpt from the language transcription

Time	Violet	Wendy
5:16	What's in your book?	
5:20		It's from Justin Bieber
5:22	[I like that book]	[I've got it to show because it's a non-fiction book. If it was a fiction book it would have pictures.]
5:35	I really like that	[It wou]
5:40		Do you like Justin Bieber?
5:45	Yes...very much	
5:52		I also like it too...I have loads of things

b Screen shots from the video data



Time (minutes : seconds)

5:16

5:22

5:35

5:45

**Figure 4. Two different representations of the conversation between Violet and Wendy**

The children's gaze direction, language and use of a non-fiction book present this object as the focus of their interaction and their attention is initially on each other. This is in keeping with the learning task which is for Wendy to describe what she had been doing in her English lessons (she had been learning the features and vocabulary associated with reference books). On hearing what is in the book Violet lowers her gaze and smiles. This embodied action indicates a possible side involvement with the subject matter of the book. Violet indicates her interest in the book backstage by looking there and saying 'I like that book.' Only when this interest has been tested and approved backstage does Violet then reiterate it frontstage to Wendy. Wendy responds with the question 'Do you like Justin Bieber?' The question and

Wendy's actions indicate her willingness to shift roles from a formal '*show and tell*' style activity with the book to a more equally balanced conversation about a popular musician.

Scollon and Scollon (2003, pp. 50-52) describe how time and space interact with each other. They refer to monochronism as a state of focusing on one thing at a time, displaying a sense of urgency and single activity. Polychronism, in contrast, refers to a person whose attention is divided and suggests a less laconic sense of time. In the example from Figure 4 the posture and activity of Violet would suggest a shift from focused activity in conversation with her partner to split attention between the off-screen area, activity on screen and the subject of the book. This move from monochronism to polychronism signals a shift from the original activity to digress on to the subject of Justin Bieber. This change in conversation would suggest a shift in the children's sense of time from the need to work through the task with urgency to an unhurried open conversation about popular culture.

The embodied actions present in the data would conform to the notion that gesture is not replaced by spoken language in children's language development. Instead, actions are combined with spoken words to express a voice (Hall *et al.*, 2013). In the context of this study this was achieved with varying degrees of success for different children suggesting a need for them to understand how different communicative resources index each other in this environment. Thus, an important outcome of this investigation is to endorse the need to attend to voice from a multimodal perspective.






### ***5.2 How does mediation through Skype affect the way in which children are able to express themselves?***

In the interaction order ambiguity over the presence and absence of the participants during the online conversation came from their ability to signal themselves as simultaneously present


and absent. The students' appearance before the webcam signalled their presence in a social encounter while their gaze vector might indicate their social absence.

The expressive "equipment" (as Goffman (1959) calls it) that constitutes the children's personal front is conspicuous across the data when important elements are absent, as in the case of eye contact between Anna and Ian in the situation described above. Figures 5 and 6 are examples from the data in which pupils were unable to express their voice clearly to each other. In each of these examples important elements of how the children would usually express themselves are missing. Figure 5 is based on data from the pilot study which shows the beginning of an exchange between Violet (from England) and Wendy (from Portugal). The verbal transcription shows how Wendy repeats her opening 'hello' and misses the question that Violet asks. It would appear that Wendy was not expecting Violet to speak. The reason for this false start in the conversation might be found in the personal front displayed by Violet. Her gaze direction is predominately to the bottom right hand corner of Wendy's screen signalling her social absence from the conversation. However, the children are using webcams that are separate from the monitor and the camera for the English children is located slightly above them and to the left of the children as they appear in the screen shot. This means that when the English children look directly at the represented image of the Portuguese children they look to the bottom right hand side of the screen. To look directly at their interlocutor and signal their social presence the children would need to look at the webcam and not the image on the monitor. By seemingly not making eye contact the children are unable to initiate dialogue and express their voices to begin with, leading to a false start in their conversation. In Figure 6 a poorly angled webcam leaves Wayne (from England) with only the top of his head and his verbal speech to express himself. In this instance the conversation is again at the beginning. Despite the fact that Wayne is addressing Zack with his voice the absence of any other expressive equipment causes Zack to ignore this verbal

contribution by talking over it. Without any visual cues to help the conversation her gaze then turns away from the screen signalling her social absence from her Skype partner. The data illustrate the importance of other people engaging with the speaker's voice if it is to make meaning in conversation. For the children to connect with a voice the speaker must signal that they are socially present.

Time	Screen shot	Violet indexed through language	indexed through the material world	Wendy indexed through language	indexed through the material world
5:00			Gaze to bottom right of screen	hello	gaze directly ahead
5:08			Gaze to bottom left of screen body angled to left (teacher)		body angled ahead (camera) holding book
5:10		[What is your book?]	Gaze to bottom right of screen holding microphone in an active position	[hello]	gaze directly ahead leaning forwards into the screen displaying face
5:13			Gaze to right body angled to left (teacher) microphone lowered	What? Can...May you repeat	gaze lowered (to mic) displaying face
5:16		What's in your book?	gaze to bottom right hand side of screen microphone raised		gaze ahead sat back holding book

**Figure 5. A false start in the conversation and analysis of eye gaze**

Time	Screen shot	Wayne indexed through language	indexed through the material world	Zack indexed through language	indexed through the material world
0:05		[Hello CLIP my name is ... I am going to be talking to you today]	---	[hello... hello, I can't see you]	gaze towards the screen body sat upright and angled towards the camera gaze turns to the left (teacher)









**Figure 6. A false start in the conversation and a poor camera position**

Tension between these conflicting states of being can be seen to influence whether the children were able to add their voice to the dialogue. If children mistakenly signal themselves as socially absent from their partner it can lead to their being excluded or overlooked during a conversation, or an activity or topic they wish to focus on may be subordinated in place of another, thus, leading to a breakdown in communication.



The visual way in which Skype mediates the children's conversations can support their ability to make meaning. The material environment around them provides them with the means to represent their ideas creatively, beyond spoken or written words. The children in the data used a variety of visual and gestural means in creative ways to voice their ideas to others. Gestures, signs and objects were employed to represent activities, interests and lexical items that are important to the children in their lives. If we assume that the meaning expressed by one's voice emerges somewhere between the speaker's intent and the response of the addressee (Bakhtin, 1986), then the representational relationships evoked through using gestures or objects in place of words play a key role in helping children understand what is expressed by a voice.

The data from the Skype session discussed earlier shows how Anna is able to manipulate the semiotic resources available to her and engineer a change in the type of conversation from task-based activity to talk that grows around the central topic of her teddy bear. The bear, therefore, acts as a tool to offer further – and less formal – interaction between the children. The imaginative way in which this is achieved points to Anna's strong sense of personal agency. This is further illustrated when she brings the bear to the forefront of the webcam, displaying the symbol of the English flag on its jumper and asks 'can you read his belly?' (Figure7). The bear is positioned to the side of Anna indexing new information; an attempt by Anna to establish an area of shared experience with her English Skype partners. Anna's resourceful manoeuvring of signs and objects shows how the culturally acknowledged tools of a teddy bear and a flag are individually shaded and toned when they are used to express Anna's voice.

20:44		(can you read) (his belly?) brings prop to forefront so that flag and 'England' become visible as new information on the left of the screen. A is establishing common ground with Skype partners					(I've got one at home)
20:51		Can you read his belly? 					
21:00		yes			(It's our flag)  England		
21:10		Um, my grandmother gave it from .eg London yes					
21:20					oh		shared connection prompts response— talking about home I've got a massive teddy bear at home 

**Figure 7. Example from the analysis of the visual order**

The screen through which the children express their voices when using Skype reflects both the tools for its expression as well as the background in which voices engage. Skype provides a visual medium for voice that allows children to represent objects, actions and feelings with something that stands for them. This ability may or may not be paralleled by children's corresponding ability to represent these in language (Cassell and Ryokai, 2001). In this situation the material means for expressing a voice is expanded to include *the silent language* (Hall, 1959) of paralinguistic or concrete objects that may be selected by a communicator to express their thoughts. The example above shows how a teddy bear is used to represent a Portuguese child's connection to England and to establish common ground with the other students.

The enhanced capacity for making meaning that representational resources (such as gesture or objects) bring thus motivates the children further to share their experiences with their social partners, and so binds their voice with others through dialogue. This relational model for how children express their voice together fits with the notion that dialogue grows informally around a central theme. The online conversations led by the children do not seem to follow formal 'drill' or 'initiation-response-feedback' genres associated with classroom talk (see Mercer, 2000). Instead, the affordances of Skype seem to suit a relational view of language in which talk is allowed to evolve as the children explore the affordances of the media and their developing inter-personal relationships.

### ***5.3 What role does voice play in helping children to think together?***

Wegerif *et al.* (2004) have shown how the active joint engagement of children with each other's ideas through *exploratory talk* will lead to learning. In exploratory talk, conversation develops from an initial concept according to the joint acceptance of well reasoned suggestions from each speaker. Thus, through a verbal exchange of challenges and counter challenges children arrive at shared meaning which, according to Wegerif *et al.* (2004), constitutes learning. In the case of the Skype conversations investigated in this current research the driving purpose behind the children's activity was to make meaning in their L2. With an emphasis on sustaining conversation it was socially appropriate for the children to focus on the free and open questioning that, in part, characterises exploratory talk.

Of course, these conversations do not take place in a void, isolated from the world around the learners. It has long been acknowledged that in order to explore how children's conversations allow them to think together it is important to look beyond 'the bonds of mere linguistics and be carried over into the analysis of the general conditions under which a language is spoken' (Ogden *et al.*, 1946, p. 277). The current study would attest to this

perspective. The free and open way in which the children were able to use different modes such as eye gaze, represented and physical space created possibilities for the children to develop their conversations and build on each other's ideas.

Figure 8 shows how the children use their gaze and posture to invite a response from their classmates to the question 'what is a jacket potato?' asked by a Portuguese student. The children's classmates are sat outside the capture of the webcam in the backstage area of the classroom. This shift in attention alters the social group from being with the Skype partners to being with the children in the classroom. In Figure 9 the children's body language and gaze show that they are focused on their respective class partners as they share ideas before contributing to shared dialogue about their best school trip so far that year. The children's activity creates two groups within a group (this would be two *sub-withs* within a *with* using Goffman's (2009, p. 19) terminology). The separation of these groups is reinforced by the fact that the Portuguese children confer using their L1 while the English children use their L2. The body language and eye gaze of Gary on the left hand side of the screen shot in Figure 10 signals a shift in his role from the main conversationalist as he passes a tricky question to the backstage area of the classroom. In Figure 11 David, on the right hand side of the screen, points and looks up in the direction of the lights to help him explain what a firework is. By moving between these different perceptual spaces the children are able to move between different roles in a conversation and respond and adapt to shared interests.






**Figure 8. Students turning their attention to the off screen areas of their own classrooms**

The Portuguese children are using their L1 while the English children use their L2



**Figure 9. The children's attention is on their class partner**

Time	Screen Shot	Paul	Gary
20:20			On Friday we celebrated Eid
20:28			... It's a Muslim festival'
20:55		What is a Muslim?	

**Figure 10. Gary signalling a move to the off screen area of the classroom**



**Figure 11. David (on the right hand side of the screen shot) using gesture and gaze**

Rather than just the engaged activity between two children in the video conferencing space represented on the computer screen, a wider lens on the data shows how the material reality of the classroom played an important role in the online conversation. The way in which children were able to move between their material and online surroundings allowed them to engage their voices in ways that do not conform neatly to the models for *exploratory* and *cumulative* talk, but contain features of both.

## **6. Multimodal construction of voice: learning in the third space and its challenges**

The exploration of the meaning making resources involved in articulating children's voices through Skype necessitated the in-depth study of a small sample of data. A microanalysis of the recorded data – which created a vast amount of information – was employed to identify the complex and creative ways in which children orchestrated the use of diverse meaning making resources (including words, eye gaze, gesture, objects, signs and the spaces around them). The findings have revealed how signs, objects and words are all used by children with agency to take the dialogue beyond what is often required in a school context. Resources such as the children's L1 or objects brought from home allow the children to connect aspects of their life outside of the school to their activity in school. In so doing they create opportunities to forge connections with the interests of their Skype partners and influence what their voices are able to express. In the data the learning task might serve as a helpful starting point for dialogue, however, the locus of control over the way in which the conversation evolves rests with the children rather than the direction of the set task.

It is difficult to make reliable predictions as to the kinds of language use and opportunities for learning that might arise out of such conversations. What each child's voice expresses is dependent on the contributions of the voices of others, which in turn are dependent on their particular locally determined experiences and goals. In other words, communication between the children no longer fits the predictable pattern of a psycholinguistic approach to language learning. The children blend the semiotic resources to which they have access to create a new social space which can be conceptualised as a third space (Gutiérrez, 2008). This is a space where children benefit from the structure of school set activity, while having the freedom to explore creative ways of expressing their voice in interaction with others. Thirdness in this sense is associated with the hybrid communicative practice that arises from blending the familiarity of school genre language with the idiosyncratic constructions of voice from child-led talk, with the computer as meditational

tool affording a space for learning in which the cultures of school and the various cultures from out-of-school come into contact. Gutiérrez (2008) shows how such spaces comprise learning and development supported through the movement of practices across various temporal, spatial, and historical dimensions of activity. The teaching and learning roles in this model are flexible as the cultural affiliations of the participants meet, clash and grapple with each other.

However, the study has also shown some of the issues that can arise in Skype mediated conversations. The integration of a webcam into children's conversations can lead to confusion that limits, rather than supports, communication as gestures, postures, gaze and body movements may not always index the inner psychological state or speech will of an interlocutor, their voice. Instead they might reflect the way the images are represented through the video conferencing medium to the conversationalists. It is often hard to determine which embodied states reveal insights about social performance and which reveal a lack of insight into the affordances of the technological environment. Much of what is read from a speaker's voice is conveyed through postures and movements to others in the same situation. An example would be the image capture from the video camera which is an upper body shot that represents a distance of 4 to 12 feet between the interlocutors through the digital image. Scollon and Scollon (2003, p. 96) suggest that this range in a face-to-face situation requires either interaction or civil inattention (purposeful avoidance) between the interlocutors. Also, the demand for eye contact is the first move in opening up a social space for further conversation. However, the webcam used to capture the data segment is separate from the screen so when the children are watching each other on the screen they appear to be looking away from each other in the represented image. In a face-to-face situation this would signal civil inattention and index a desire not to participate in conversation. Similarly, an embodied act such as standing up might indicate a desire to point out something in the room in the



interaction order, but the represented image of a body without a head indicates to the conversation partner that they are excluded from the exchange in the visual order. The limited field of view created by the webcam can also pose problems, with hidden or backstage spaces that fall out of range of the lens (Goffman, 1959).

## ***7. Conclusion***

An important purpose of this investigation was to discover how children experience and express voice through Skype. One major outcome of this study has been to endorse a multimodal perspective of voice to understand how children are able to express themselves with others through social software. By expanding their view of voice beyond linguistic performance to include other semiotic ways of communicating (such as gesture, intonation, eye gaze or material objects for example) educators can foster the development of activities which support children's communication and develop their spoken language skills.

Teachers wishing to use video conferencing to support language learning in the classroom need to recognise that the emphasis on linguistic skills and knowledge underlying psycholinguistic tasks focused on form are not enough for children to voice their ideas through sustained conversation. If children wish to express something in dialogue with others in this environment they must not only say something, but be visibly seen to say it. There is a need for them to be proactive and creative in finding ways to make meaning with others. An attempt by the teacher to control the performance of skills and knowledge could stifle children's ability to express their voice through Skype. The children in the examples from data in this study drew on a range of semiotic resources (including objects, signs, and gestures) to share their ideas about things that they might not have been able to talk about using just speech alone. They were able to use the resources in creative ways by leading the

conversations. Their teachers helped to set the general themes of each session but relinquished control over what would be said and how.

Nevertheless, while children might be familiar with learning conversations in the traditional classroom the particular circumstances of computer-mediated communication through Skype are different. It is possible that children who are not taught to express themselves through video conferencing will find that the opportunities to practise their L2 in online conversations are curtailed or limited. It is possible that for young children entering the unfamiliar surroundings of video conferencing conversations in the classroom, the different possibilities for expressing their voices including different ways of being socially present or absent, different routines and rituals could be confusing. Young learners need to be able to understand the ways in which they are able to voice their ideas in these circumstances. Teachers have an important role to play in helping children manage their telecommunication exchanges so that they are better able to engage their voices in sustained conversation. Through sustained conversation the children are able to build on each other's ideas and in this way think together.

The findings of this present research suggest that children's Skype-mediated communication provides them with rich opportunities to practise their developing communication skills when conversations are child-led. This context is related to a view of learning as social activity in which children's performance is contingent on their ability to connect with others. The role of the teacher in this situation is to monitor the conversations and limit their intervention to the minimum. This requires consciously stepping back from directing the activity and trusting the students to take the lead. At the same time, the teacher should be open and flexible to respond to the complexity of the context in which the children are interacting if the need arises.

Due to the absence of other comparable studies into children's voice a robust framework for analysis of the video data needed to be developed, and Scollon and Scollon's (2003) concept of geosemiotics was adopted for this purpose. To our knowledge, the multidisciplinary approach taken by geosemiotics has not so far been applied to children's conversations through Skype in any other studies. Through analysis of the interaction order, visual semiotics and place semiotics the geosemiotic framework has allowed this research to account for the complexities of expressing voice in a video conferencing environment. However, the categories used in this study required careful application as a tool kit for analysis, as the geosemiotic framework presumes relationships within the data that did not always fit with the particular context of this study.

The methodology developed and used in the current research illustrates one way in which the concept of geosemiotics might be used to model children's behaviour conversing through Skype. However, the perspective on voice offered might be used to inform and guide analysis in a variety of research designs aimed at answering related as well as broader questions. Such research might take the form of further empirical studies aimed at category development or broader mapping of the physical or material characteristics of voice to explore how their use may be different from place to place in the world. In addition, mindful of the opportunities for participation that online spaces offer children who are disadvantaged by traditional approaches to learning (Gomez, 2009; Levy, 2008; Marsh, 2003; Neuman and Celano, 2006; Warrington *et al.*, 2006) it is ultimately envisaged that a better understanding of how voice is expressed through telecollaboration could help children to develop their L2 more effectively in this non-traditional environment and support their class-based learning.

The value of a multimodal understanding of voice in video conferencing environments depends in part on its utility as a psychological tool that teachers might use to help young learners reflect on how they can engage their own voices to fully exploit the potential of

Skype as an online space to support their learning. To this end, conceptualising voice to include body language as well as linguistic performance might be used in a developmental research agenda that empowers children to communicate beyond the limitations of their verbal skills. From a Vygotskian (1978) perspective the cognitive load for making meaning is distributed between the speaker's brain and the tools at their disposal (including the material and computer-represented world as well as the minds of others). By making meaning through tools other than linguistic ones, speakers are able to express more, opening up opportunities for further language learning. Similarly the notion of different social groupings in the interaction order and how this relates to represented distances in video conferencing spaces might be used in a developmental research agenda aimed at alerting young learners as to the ways they can purposefully manipulate the roles they have in social encounters to enhance their ability to voice their ideas. When children find ways of connecting with others they encounter potentially rich opportunities for extending their personal networks while practising their L2. This seems particularly important as much of our communication today takes place at a distance and the use of digital communication tools is becoming part of the day-to-day lived experience of many people, including in educational contexts.

(approx. 9,500 words)

## **References**

- Austin, N. (2015). *Video conferencing and Multimodal Expression of Voice: Children's communication in a second language using Skype*. (Unpublished doctoral dissertation). Milton Keynes, United Kingdom: The Open University.
- Bakhtin, M.M. (1986). *Speech Genres And Other Late Essays*. Austin, United States of America: University of Texas Press.

- Baldry, A. and Thibault, P.J. (2006). *Multimodal Transcription and Text Analysis*. London: Equinox Publishing Ltd.
- Belz, J. (2003). Linguistic Perspectives on the Development of Intercultural Competence in Telecollaboration. *Language Learning & Technology*, 7(3), 68-99.
- Blommaert, J. (2008). Bernstein and poetics revisited: voice, globalization and education. *Discourse & Society*, 19(4), 425-451.
- Cassell, J. and Ryokai, K. (2001) 'Making space for voice: Technologies to support children's fantasy and storytelling', *Personal and ubiquitous computing*, vol. 5, no. 3, pp. 169-190.
- Council of Europe Committee of Ministers. (2006). *Recommendation of the Committee of Ministers to member states on empowering children in the new information and communications environment* [Homepage of Council of Europe]. Available from <https://wcd.coe.int/wcd/ViewDoc.jsp?id=1041181> [July, 2011].
- Department For Education. (2003). *Every Child Matters Green Paper*. Available from <https://www.education.gov.uk/publications/eOrderingDownload/CM5860.pdf/>.
- Develotte, C., Guichon, N., & Vincent, C. (2010). The use of the webcam for teaching a foreign language in a desktop videoconferencing environment. *ReCALL*, 22(03), 293-312.
- Dicks, B., Flewitt, R., Lancaster, L. and Pahl, K. (2011). Multimodality and ethnography: working at the intersection. *Qualitative research*, 11(3). 227-237.
- Dweck, C.S., Higgins, E.T. and Grant-Pillow, H. (2003) Self-systems give unique meaning to self variables, in Leary, M.R. and Tangney, J.P.(Eds.), *Handbook of Self and Identity*. New York, Guilford Press, pp. 239–252.
- Goffman, E. (2009) *Relations in public*. New York, Basic Books .
- Goffman, E. (1981). *Forms of talk*. Philadelphia: University of Pennsylvania Press.
- Goffman, E. (1959). *The presentation of self in everyday life*. Garden City: New York

- Gomez, K. (2009). "Living the Literate Life": How Teachers Make Connections Between the Personal and Professional Literate Selves. *Reading Psychology*, 30(1), 20-50.
- Guichon, N., & Cohen, C. (2014). The Impact of the Webcam on an Online L2 Interaction. *Canadian Modern Language Review/La Revue canadienne des langues vivantes*, 70(3), 331-354.
- Gutiérrez, K.D. (2008). Developing a Sociocritical Literacy in the Third Space. *Reading Research Quarterly*, 43(2), 148-164.
- Hall E. T. (1959) *The silent language*. New York, Anchor.
- Hall, S., Rumney, L., Holler, J. and Kidd, E. (2013). Associations among play, gesture and early spoken language acquisition. *First Language*, 33(3). 294-312.
- Hampel, R. (2014). Making meaning online: computer-mediated communication for language learning. In A. Peti-Stantić & M.-M. Stanojević (Eds.), *Language as Information*. Proceedings from the CALS Conference 2012. Frankfurt am Main: Peter Lang.
- Hauck, M. and Youngs, B.L. (2008) Telecollaboration in multimodal environments: the impact on task design and learner interaction. *Computer Assisted Language Learning*, vol. 21, no. 2, pp. 87-124.
- Jewitt, C., ed. (2009). *The Routledge Handbook of Multimodal Analysis*. Abingdon, Oxon: Routledge.
- Kern, R. (2014). Technology as Pharmakon: The Promise and Perils of the Internet for Foreign Language Education. *The Modern Language Journal*, 98(1), 340-357.
- Lamy, M-N and Flewitt, R (2011) Describing online conversations : insights from a multimodal approach, in Develotte C. Richard K. and Lamy, M-N., (eds), *Décrire la Conversation en Ligne : Le Face à Face Distanciel*. France, Lyon, ENS Éditions, pp. 71–94.

- Levy, M. (2009). Technologies in use for second language learning. *The Modern Language Journal*, 93(1), 769-782.
- Levy, R. (2008). "Third Spaces" Are Interesting Places: Applying "Third Space Theory" to Nursery-Aged Children's Constructions of Themselves as Readers. *Journal of Early Childhood Literacy*, 8(1), 43-66.
- Mahiri, J. and Sablo, S.(1996). Writing For Their Lives: The Non-School Literacy of California's Urban African American Youth. *Journal of Negro Education*, 65(2), 164-180.
- Marsh, J. (2003). One-way Traffic? Connections between Literacy Practices at Home and in the Nursery. *British Educational Research Journal*, 29(3), 369.
- Mercer, N. (2000). *Words and Minds: How we use language to think together*. Abingdon, Oxon: Routledge.
- Neuman, S.B. and Celano, D. (2006). The knowledge gap: Implications of leveling the playing field for low-income and middle-income children. *Reading Research Quarterly*, 41(2), 176-201.
- Norris, S. (2011). Three hierarchical positions of deictic gesture in relation to spoken language: a multimodal interaction analysis. *Visual Communication*, 10(2), 129-147.
- Ogden, C.K., Richards, I.A., Malinowski, B. and Crookshank, F.G. (1946). *The meaning of meaning*. New York: Harcourt, Brace & World New York.
- Scollon, R. and Scollon, S.W. (2003). *Discourses in Place: Language in the material world*. London: Routledge.
- Spencer, T., Falchi, L. and Ghiso, M. (2011). Linguistically Diverse Children and Educators (Re)Forming Early Literacy Policy. *Early Childhood Education Journal*, 39(2). 115-123.

Vygotsky, L.S. (1978). *Mind In Society*. Cambridge, Massachusetts: Harvard University Press.

Warrington, M., Younger, M., Neuman, S.B. and Celano, D. (2006). Working on the inside: discourses, dilemmas and decisions. *Gender and Education*, 18(3), 265-280.

Wegerif, R., Littleton, K., Dawes, L., Mercer, N. and Rowe, D. (2004). Widening access to educational opportunities through teaching children how to reason together. *Westminster Studies in Education*, 27(2). 143-156.




Wolfe, S. and Flewitt, R. (2010). New Technologies, New Multimodal Literacy Practices and Young Children's Metacognitive Development. *Cambridge Journal of Education*, 40(4). 387-399.



## Appendix A: Excerpt from the multimodal coding chart used to analyse the data

Time	The interaction order (relationship between bodies)	Visual semiotics (the use of images and signs)	Place Semiotics
0-1:45	<i>time</i>	<i>modality</i>	<i>social context</i>
1:45-2:03	<i>space</i>	classroom furniture, displays	school lesson with T
2:03-2:47	confirmed - examining Ss	classroom arranged so both screen focus (S field to right)	S field 5 talks to friend/quiz
2:47-3:03	fills screen with self n prop	S: Behaviour controlled	Lesson - T authority
3:03-3:28	leans to tell T what is written	see school like props: learn	quiz, all Ss engaging, public
3:28-3:46	attempts clarify self n prop	S field with T, CLIP public	public
3:46-5:11	Sa, Closez distance	confusion's attention on T	classroom focus
5:11-6:04	attempt clarify self n prop	field gaze: T, CLIP screen/prop	focus on Skype redirected vision and lang in collaboratio
6:04-6:27	Sa, Closez distance	B field, attention to periphery	focus on T as S tries to question IPF, body passive, T, priat
6:27-8:07	attempts clarify self n prop	centre to periphery	Ill/off screen, Prop on screen
8:07-8:32	Sa, Closez distance	S field periphery, CLIP prop	public
8:32-9:06	attempts clarify self n prop	periphery (T) and screen	public
9:06-10:03	Sa, Closez distance	screen: engaged	field T dispute S, Ra, Vz, M
10:03-10:40	attempts clarify self n prop	off screen activity	public with CLIP S n CLIP S
10:40-11:50	attempts clarify self n prop	interaction through T	individual
11:50-12:15	attempts clarify self n prop	direct interaction w Ss	public
12:15-13:29	attempts clarify self n prop	direct interaction w Ss	public performance
13:29-14:24	attempts clarify self n prop	direct interaction w Ss	public
14:24-14:50	attempts clarify self n prop	direct interaction w Ss	public
14:50-16:20	attempts clarify self n prop	direct interaction w Ss	public
16:20-16:54	attempts clarify self n prop	direct interaction w Ss	public
16:54-17:32	attempts clarify self n prop	direct interaction w Ss	public
17:32-18:28	attempts clarify self n prop	direct interaction w Ss	public
18:28-18:38	attempts clarify self n prop	direct interaction w Ss	public
18:38-19:30	attempts clarify self n prop	direct interaction w Ss	public
19:30-20:06	attempts clarify self n prop	direct interaction w Ss	public
20:06-20:12	attempts clarify self n prop	direct interaction w Ss	public
20:12-20:19	attempts clarify self n prop	direct interaction w Ss	public
20:19-21:14	attempts clarify self n prop	direct interaction w Ss	public
21:14-21:38	attempts clarify self n prop	direct interaction w Ss	public
21:38-22:07	attempts clarify self n prop	direct interaction w Ss	public
22:07-22:55	attempts clarify self n prop	direct interaction w Ss	public
22:55-23:17	attempts clarify self n prop	direct interaction w Ss	public
23:17-23:59	attempts clarify self n prop	direct interaction w Ss	public
23:59-24:32	attempts clarify self n prop	direct interaction w Ss	public
24:32-24:53	attempts clarify self n prop	direct interaction w Ss	public

### Appendix B: Multimodal interactional transcription of speech, gaze and

Time	Screen Shot	A	B	C	D	E	F	G	H
17:35		hello, my name is A and I'm going to ask some questions for you Dominant attention, delivering quiz, subordinate attention on T, gaze vector off screen. Ss side involvement – not properly engaged demand: requires some form of interaction – first move in opening up interaction space in the social world: intimate personal distance: 96 (represented)							
17:40									
17:44		gaze vector of A and B down at question sheet, main involvement with the quiz and what to say next, civil inattention signalled		thank you A Craning head in to fit screen and engage	thank you A exaggerated voice intonation – wishes to catch attention (engage A) away from crowd while	thank you A CD and E all lean forward	thank you A	thank you A	thank you A

Platform event- though all potentially participate (not a watch) seating directed by T to fit people in as participants, others try to fit the frame too, vying for a place

Formally organised seating, uniform angle of body, hands low and on laps, eye gaze angled to right (screen) or to left T

Main attention on screen action, a 'watch' with A and B, posture looks attentive, head up

Main attention on T, gaze vector on her, T: models response

T: thank you