ReCALL **21**(2): 186–205. 2009 © European Association for Computer Assisted Language Learning
doi:10.1017/S0958344009000184 Printed in the United Kingdom

Multimodal student interaction online: an ecological perspective

THERESE ÖRNBERG BERGLUND
Department of Language Studies, Umea University, 901 87 Umeå, Sweden
(e-mail: therese.ornberg@engelska.umu.se)

Abstract

This article describes the influence of tool and task design on student interaction in language learning at a distance. Interaction in a multimodal desktop video conferencing environment, FlashMeeting, is analyzed from an ecological perspective with two main foci: participation rates and conversational feedback strategies. The quantitative analysis of participation rates shows that as far as verbal interaction is concerned, multimodality did not have an equalizing effect in this context, contradicting previous research on multimodal student interaction. Additionally, the qualitative analysis of conversational feedback strategies shows that whereas some multimodal strategies were employed, the students did not manage to fully act upon the communicative affordances of the tool, as the feedback ratio during and after the often long broadcasts was relatively low. These findings are related to task and tool design and the article discusses how design improvements in these areas might result in a more constructive language learning ecology.

Keywords: Language learning ecology, task and tool design, multimodal desktop video conferencing, equalization, common ground, conversational feedback

1 Introduction

Distance education has long followed the correspondence model, and it is only in recent years that synchronous audio interaction between students has become a more common component of language learning at a distance. This development makes it possible to give feedback in real time, which might result in a high degree of social presence and high levels of participation. However, providing tools for synchronous interaction does not automatically result in an efficient and constructive interaction setting, as many different factors may affect the interaction taking place (cf. O'Dowd & Ritter, 2006; Hauck, 2007; Hauck & Youngs, 2008).

In the current article, the focus is on interaction among students of English at a distance participating in discussions in a multimodal desktop video conferencing platform called FlashMeeting. The aim of the article is to investigate how student interaction is oriented to and affected by the complete ecology in which it is situated, with a main interest in the relevance of tool and task design.

https://doi.org/10.1017/S0958344009000184 Published online by Cambridge University Press

Two discussion sessions were analyzed in order to assess the influence of tool and task design on interactional patterns. The analysis was conducted in two steps: first, quantitative data were used to investigate participation rates, with the aim of seeing whether multimodality in fact supports equalization; that is, whether the opportunity to choose a preferred mode of interaction ensures even participation rates (cf. Warschauer, 1996; Vetter & Chanier, 2006); and second, coherence creation through multimodal conversational feedback strategies was qualitatively analyzed, following Common Ground theory (Clark & Brennan, 1991).

2 Background: theory and previous research

2.1 Affordances, conventions and socio-cultural approaches to learning

Approaching an online environment from an ecological perspective means acknowledging the influence of the complete environment on the activities taking place there. One way of analyzing the influence of the environment is by focusing on its affordances, that is, the options for interaction that the environment provides for the participants residing in it, and especially those options that are acted upon by the individual (Gibson, 1977, 1979). Some affordances will be of specifically high relevance when communicating, influencing how we can express and perceive communicative actions (cf. Gaver, 1996). In addition, when interacting with others, specific protocols for how to deal with the communicative affordances of the environment develop, and by analyzing interactional patterns, these conventions can be detected (cf. Hutchby, 2001).

An ecological perspective has also been adapted in research on language learning (cf. van Lier, 2004; Leather & van Dam, 2003), and here it has been argued that a situated approach, taking various aspects of the learning environment into account, is fruitful when considering how best to support language learners. This is in line with socio-cultural theories of learning, where the learner is seen as situated in a specific culture and where learning takes place through interaction with the environment, including artefacts and other human beings (Vygotsky, 1986; Säljö, 2000). Further, from a socio-cultural perspective on learning, communicative affordances are, in fact, also affordances for language learning.

One way in which communicative affordances influence interaction patterns is demonstrated in the ways in which conversational feedback can be delivered. Not least in distance education these types of cues can become crucial, as lack of feedback might add to the feeling of isolation which physically dispersed students might experience. A high level of engagement, contrarily, can help create a sense of social presence (cf. Tammelin, 2004; McIsaac & Gunawardena, 1996; Gunawardena & Zittle, 1997) in spite of physical distance, which in turn might encourage more active participation.

2.2 Common Ground Theory

Common Ground Theory (Clark & Brennan, 1991) provides further insight concerning the importance of conversational feedback in the collaborative project of communicating. When we communicate, the theory poses that we need to continuously reaffirm that common ground has been reached, that is, we need to give

https://doi.org/10.1017/S0958344009000184 Published online by Cambridge University Press

Table 1 *Eight factors influencing grounding (adapted from Clark & Brennan, 1991)*

Co-presence	A and B share the same physical environment.
Visibility	A and B are visible to each other.
Audibility	A and B communicate by speaking.
Contemporality	B receives at roughly the same time as A produces.
Simultaneity	A and B can send and receive at once and simultaneously.
Sequentiality	A's and B's turns cannot get out of sequence.
Reviewability	B can review A's messages.
Revisability	A can revise messages for B.

and receive indications that a previous message has been understood, but also how it has been evaluated. This is accomplished through the process of *grounding*. Here, it is not enough to seek for negative evidence, but we always also look for positive evidence that current contributions are being understood. Clark and Brennan identify three types of positive evidence: *acknowledgements*, *relevant next turns*, and *continued attention*.

By *acknowledgements*, they mean *back-channelling cues*, *continuers* and *assessments*. Also non-verbal acknowledgement, such as head nods are included in their account. The type of *relevant next turns* which, according to Clark and Brennan, are most easily identified are so called *adjacency pairs*, where the production of a first pair part sets up a normative expectation of the production of a second pair part (Schegloff, 1968). Examples include question – answers, request – response and invitation – acceptance. However, other types of utterances are also linked, as it is claimed that conversation generally consists of coherent sections. *Continued attention*, Clark and Brennan argue, is the most basic way of revealing positive evidence, and gaze is mentioned as an important tool for indicating that one is paying attention. Both acknowledgements and relevant next turns are also of importance in this context.

Moreover, it is claimed that both the purpose of the interaction and the medium used will affect grounding techniques, and the authors present a model for analyzing the influence that mediation has on the possibilities of reaching common ground. This model consists of eight factors which in different ways affect the grounding process. These are summarized in Table 1.

In addition, Clark and Brennan identify eleven different types of costs that vary depending on which constraints apply in any given situation. These costs are: *formulation costs*, *production costs*, *reception costs*, *understanding costs*, *start-up costs*, *delay costs*, *asynchrony costs*, *speaker change costs*, *display costs*, *fault costs* and *repair costs*.

The model for constraints and costs in relation to grounding has often been applied in research within the fields of Computer-Supported Collaborative Work (CSCW) and Computer-Supported Collaborative Learning (CSCL) (cf. Veinott, Olson, Olson & Fu, 1999; Fussell, Kraut & Siegel, 2000).

2.3 Multimodal online interaction

When using multimodal tools for online interaction, the way the different modes combine result in new types of constraints and affordances.

https://doi.org/10.1017/S0958344009000184 Published online by Cambridge University Press

Multimodality is defined by Kress and van Leeuwen (2001: 20; quoted in Hauck & Youngs, 2008) as "the use of several semiotic modes in the design of a semiotic product or event, together with the particular way in which these modes are combined – they may for instance reinforce each other […], fulfill complementary roles […] or be hierarchically ordered." These possibilities for combining modes of meaning making may result in an "orchestration of meaning" (Kress, Jewitt, Osborne & Tsatsarelis, 2001; quoted in Hauck & Youngs, 2008).

The multimodal dimensions of interaction have been studied by several scholars and from different perspectives (cf. e.g. Kress, 2000; Norris, 2004; Goodwin, 2000). Norris (2004) introduces the notion of *modal density* to account for the ways in which different modes interrelate when we communicate. Participants in conversation are able to conclude levels of attention through modal density, she argues, and this can be achieved either through a combination of different modes (*modal complexity*) or through emphasis on one specific mode (*modal intensity*).

The affordances of the specific system used here, FlashMeeting, have been investigated by the developing team at the Open University in the UK. For example, they have published on knowledge mapping in relation to the various data collected during meetings (Okada *et al*., 2007). The most relevant results in relation to the current article can be found in their research concerning participation rates and mode choices in different types of meetings. Here, they have used innovative visualizations to illustrate how participation patterns may alter depending on the purpose of the interaction (Scott *et al*., 2007). In another article, participation rates and roles are compared with users' own perception of peer-to-peer learning in FlashMeeting. Here, the authors have access to a large set of longitudinal data, which shows that participants are able to maintain "symmetrical support" (Scott *et al*., 2008, forthcoming). Their work also includes investigations of speech acts, lexical analysis and emotion identification (Binti Abdullah *et al*., 2008, forthcoming), as well as research into how the tool can be used for language learning in the 'Proteach Italia' project (http://flashmeeting.open.ac.uk/research/language-teaching.html).

Previous research on language learning in multimodal online environments has often been concerned with audiovisual tools (an extensive review is available in Hauck & Youngs, 2008). Some examples of particular relevance here are Hampel's (2006) findings concerning task adaptation depending on the affordances and constraints of the tool employed, and Vetter and Chanier's (2006) findings concerning participation rates. The results here showed that multimodality in audiographical communication in fact led to more equal participation rates, in that in groups where participants contributed unevenly in audio, those who used the audio channel the least were the most active in the text chat. Whether the same is true for the interaction taking place in the desktop video conferencing environment employed here remains to be seen.

3 The ecology of the student sessions

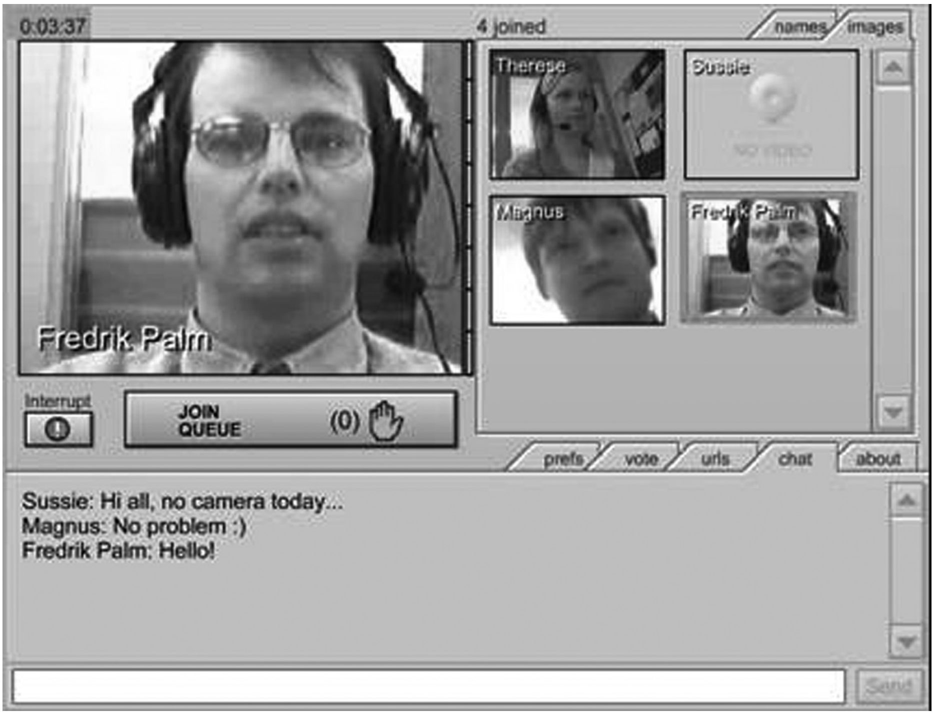
In this section, the setting of the study is presented. First, an introduction to the tool is provided (section 3.1), and this is followed by a description of the tasks that the students were given (section 3.2). The section concludes with an account of the methodology used when gathering and analyzing the material (section 3.3).

https://doi.org/10.1017/S0958344009000184 Published online by Cambridge University Press

3.1 Tool design

The desktop video conferencing platform employed in this investigation is called FlashMeeting (http://www.flashmeeting.com). FlashMeeting supports both voice and text interaction. In FlashMeeting there is a built-in turn-managing device through which people "broadcast" in order to speak and have the opportunity to "raise their hands" (a hand with a number, indicating position in the queue will appear in the corner of the video image thumbnail) and line up to take the floor if someone else is already broadcasting. The material qualities of the system itself allow for only one speaker at a time. However, it is possible to interrupt, by pushing a button which has been provided for this specific purpose.

The participants in FlashMeeting are represented by video images, which are shown in a bigger format when one is broadcasting, and otherwise will appear in thumbnail format on the right-hand side of the broadcasting area. These thumbnail images are arranged according to the order in which you log into the conference and are updated, but with a time lag. The broadcasting image is updated in real time, but depending on internet connection and computer capacity this image will also sometimes lag behind noticeably. On the image you also find the name of the participant. In addition, participants have graphical emoticons and votes at their disposal, which when they are activated by mouse clicks will appear in the corner of the thumbnail of the image of that person. In order to access these pre-programmed cues, participants choose the vote tab in the lower half of the FlashMeeting window. Other tabs to choose from include the text chat tab and a tab for shared URLs.



Sussie: Hi all, no camera today...
Magnus: No problem :)
Fredrik Palm: Hello!

Fig. 1. Screen shot from FlashMeeting.

Table 2 *Factors influencing grounding in FlashMeeting on a mode level*

	Broadcasting	Text	Thumbnail video	Pre-programmed emoticons and votes
Co-presence	–	–	–	–
Visibility	+	–	+	–
Audibility	+	–	–	–
Contemporality	+	+	+	+
Simultaneity	–	–	+	+
Sequentiality	–	–	–	–
Reviewability*	–	+	–	–
Revisability	–	+	–	–

* Everything happening during the meeting can be reviewed afterwards in the recorded version of the session. However, only text messages can be reviewed during ongoing interaction.

The different modes available can be mapped onto Clark and Brennan's (1991) model for constraints on grounding.

The combination of the different modes leads to different types of affordances on the tool level. For example, during multimodal interaction, simultaneity is possible during broadcasts, in that text and cues in the video thumbnail images can be used to deliver feedback. On the other hand, the simultaneity of the use of pre-programmed cues reported in the Table 2 may be distorted during multimodal interaction, since participants need to switch between tabs in order to access different modes.

It should be mentioned that the recordings analyzed were captured during the spring semester of 2006, and since then improvements have been made to the FlashMeeting system. For example, in newer versions some emoticon buttons are also available in the text chat tab. In addition, FlashMeeting now includes a shared whiteboard and other features.

3.2 Task design

In an attempt to allow for student-centred discussions, there was no teacher present in the sessions analyzed. Instead, the teacher got access to recordings of the sessions so that comments could be made later. Questions for discussion, concerning different aspects of cultural studies, were distributed to the students before each lesson, and they were told to find information online to support their claims. Further, the students were told that active participation was needed in order for them to receive their grade.

In preparation for the first discussion session, sixteen questions were distributed, and the students were given the main responsibility for a few questions each. Further instructions told the students to appoint two of the others to give additional comments on the question they had been assigned, and so, everyone should be prepared to reply to all questions. Here are some examples of questions from session 1. The indicated topic is "the cellphone":

- What did people use the cellphone for initially? Give examples (and go online or to a reference book to find information).

https://doi.org/10.1017/S0958344009000184 Published online by Cambridge University Press

- What associations do we have to it today? Here, you may have to think about context as well – we react differently to the sound of a cellphone or a person talking when we are at the bus stop than when we are at the cinema, right?
- In our society today we experience a lot of stress, and many of us are often required to be available around the clock. Can we see how the cellphone fits into this?

After the first session, the teacher made the following complement to the instructions:

> I recommend that you make brief notes of your responses for the discussion, but also keep in mind that since a discussion is a dynamic form of interaction, you may move away from the specific questions and talk about other things related to the topic.

In session 5, the students were given four questions in total, and no instructions concerning specific responsibilities were provided. Here is an example of a question from session 5:

> Like Sweden, Britain is a parliamentary democracy encased in a constitutional monarchy. The American Congress also shares some features of the British Parliament, such as being divided into two parts, or houses – the Senate and the House of Representatives. However, unlike in Britain where the prime minister is chosen from the majority party, the American president is voted upon directly by the electorate. What difficulties do you think this poses for a head of state trying to work with Congressional representatives and attempting to represent an entire nation to the world and to itself?

Before the regular sessions, there was a test session where the students were given instructions concerning the different modes and were encouraged to try them out. In both sessions analyzed, the conference was opened ten minutes prior to the scheduled beginning of the session, and this period of time was excluded from the analysis.

The sessions had five and four participants respectively (pseudonyms used in the following), all with Swedish as their native language. One of the students had lived in Britain for a longer period of time (Emma) and another was living in the US and had lived there for over ten years (Benny). Both Emma and Benny had much previous experience with online communication, whereas Denise was new to this way of communicating. Filip and Adam were both quite technically advanced.

3.3 Methods of gathering and analyzing the material

The student sessions were recorded with the screen capturing tool Camtasia. In each discussion session, a non-student user with no video or audio connected (the author) was present to do filming. The sessions were also automatically recorded, but the thumbnail video images were not included in these recordings, so a more detailed recording was needed. If necessary, technological support would also be provided in audio, but from a separate user account.

The verbal production of the recordings was transcribed according to Jeffersonian transcription conventions, but with a low level of detail (Jefferson, 2004). Each contribution in the different modes analyzed was time stamped, and in addition the

https://doi.org/10.1017/S0958344009000184 Published online by Cambridge University Press

timing of contributions in the visual modes was shown in relation to the audio (abbreviations in parentheses) within the low detail broadcasting transcripts. Beneath each broadcasting transcript the different insertions were further explained. As for the thumbnail video images in FlashMeeting, it was decided to focus on smiles, as these are examples of positive evidence and are relatively easy to detect.

The following example serves to introduce the transcription method. Here, we see how some of the modes available might be employed in the same interactional sequence, as Emma reacts to what Filip is saying both by smiling in the video thumbnail image and by using pre-programmed emoticons. Note also how Emma's acknowledgements are noticed by Filip during his contribution.

Example 1: Illustrating the transcription method; Multimodal reinforcement – Session 1

0:44:02 – 0:44:32 Filip broadcasting
eh well I think today it's more than (.) getting pregnant I think today it's more of a (.) luxury (Ev) vacation (Ev) (.) ((laughs)) (.) (Ev) eh (.) it's a time to (.) the the most luxury vacation you have in your life you have on your honeymoon (.) you go to some romantic or exotic place (.) could be cold or warm but (.) and there yeah you have fun and (.) so on

0:44:09 Emma video
smiles

0:44:09 Emma video
emoticon

0:44:12 Emma video
changes to winking emoticon

Based on these basic data, figures concerning participation rates were calculated. In the analysis of conversational feedback mainly qualitative methods were employed, and the data was analyzed for positive feedback, following Clark and Brennan (1991).

4 Results

This section begins by summarizing general data concerning the discussion sessions. We then turn to results concerning participation rates (section 4.1), and this is followed by a section dealing with conversational feedback (section 4.2). Here, Clark and Brennan's (1991) discussion of *positive evidence* is taken as a basis for the analysis.

Table 3 shows general data about the analyzed sessions. As previously stated, the analysis begins when the actual discussion starts, and here we can see that the analyzed time is somewhat shorter in session 1 than in session 5. This, together with the fact that there was one participant less in session 5, should be kept in mind when comparing the figures in Table 3.

It is interesting to note that there are many more text messages in total in session 5 than in session 1. One reason for this is that while waiting for the discussion to begin, those students who arrive early use the text chat for socializing (this section has been excluded from the current analysis). Table 3 also shows that there are fewer and longer broadcasts in the analyzed section in session 5 than in session 1, whereas

https://doi.org/10.1017/S0958344009000184 Published online by Cambridge University Press

Table 3 *General data concerning analyzed sessions*

Session 1		***Session 5***	
Total time	01:05:40	Total time	01:16:20
Analyzed time	00:52:42	Analyzed time	01:00:17
Broadcasting (analyzed)	00:46:36	Broadcasting (analyzed)	00:54:49
Silence (analyzed)	00:06:06	Silence (analyzed)	00:05:28
Total figures		***Total figures***	
Broadcasting turns	136	Broadcasting turns	100
Text messages	57	Text messages	110
Figures for analyzed material		***Figures for analyzed material***	
Broadcasting turns	111 (128/hour)	Broadcasting turns	91
Text messages	32 (37/hour)	Text messages	34
Line-ups	35 (31% of broadcasts)	Line-ups	35 (38% of broadcasts)
Interruptions	–	Interruptions	–
Mean length of turns	00:00:25	Mean length of turns	00:00:36
Mean length of pauses	00:00:04	Mean length of pauses	00:00:03

almost the same number of text messages are sent in both analyzed sections. In addition, there is less silence in session 5. It can further be noted that in comparison with focused face-to-face interaction, in this material turns are quite long and so are the pauses in between turns.

4.1 Participation rates

In the instructions for session 1, the students were told to assign the next speaker to give further comments to their questions. The students did not follow these instructions, but instead left the floor open or asked a question without selecting a specific addressee. The only occasions when someone was appointed by the others to take the floor was when Benny was asked questions relating to his expertise on the United States. Table 4 shows participation rates in broadcasting, text and visual modes in session 1. In the total figures, the number of instances per hour has been calculated in order to allow for comparison with participation rates in the hour-long session 5, summarized in Table 5.

When comparing participation rates in the two sessions, it is clear that they do not alter very much: Emma and Benny are the most active in text and audio, whereas Denise is the least active in these modes, but is seen to smile in the thumbnail video representation among the most in both sessions.

Adam does not use text very much in either session, but uses emoticons and votes instead. Furthermore, in session 5 there might be a correlation between his more frequent text chatting and less frequent use of emoticons and votes. Emma, who used some emoticons in session 1, does not use any in session 5. This indicates that the usage of emoticons and votes does not catch on within this group. Of course, we also need to keep in mind that in order for an emoticon to be used it has to be relevant in

Table 4 *Participation rates – Session 1*

		Participant					
		Adam	Benny	Denise	Emma	Filip	Total (per hour)
Broadcasting	Total nr of broadcasts	28	26	13	29	15	111 (128)
Session 1	% of total nr of broadcasts	**25**	**23**	**12**	**26**	**14**	**100**
	Total broadcasting time	00:08:05	00:15:16	00:04:48	00:12:34	00:05:53	00:46:36
	% of total broadcasting time	**17**	**33**	**10**	**27**	**13**	**100**
Text chat	Total nr of text messages	4	9	1	9	9	32 (37)
Session 1	% of total nr of text messages	**13**	**28**	**3**	**28**	**28**	**100**
	Total nr of words in text	7	33	1	79	9	129 (149)
	% of total nr of words in text	**5**	**26**	**1**	**61**	**7**	**100**
Visual cues in thumbnail video image	Thumbnail smiles	16	3	14	12	10	55 (63)
Session 1	Emoticons	3	–	–	6	–	9 (10)
	Votes	5	–	–	–	–	5 (6)

https://doi.org/10.1017/S0958344009000184 Published online by Cambridge University Press

Table 5 *Participation rates – Session 5*

		Participant					
		Adam	Benny	Denise	Emma	Filip	Total
Broadcasting Session 5	Total nr of broadcasts	18	26	11	36	X	91
	% of total nr of broadcasts	**20**	**29**	**12**	**40**	X	**100**
	Total broadcasting time	00:12:38	00:19:49	00:04:32	00:17:50	X	00:54:49
	% of total broadcasting time	**23**	**36**	**8**	**33**	X	**100**
Text chat Session 5	Total nr of text messages	8	15	2	9	X	34
	% of total nr of text messages	**24**	**44**	**6**	**26**	X	**100**
	Total nr of words in text	24	90	2	74	X	190
	% of total nr of words in text	**13**	**47**	**1**	**39**	X	**100**
Visual cues in thumbnail video image Session 5	Thumbnail smiles	7	1	16	8	X	32
	Emoticons	3	–	–	–	X	3
	Votes	1	–	–	–	X	1

https://doi.org/10.1017/S0958344009000184 Published online by Cambridge University Press

the local context. As the discussion in session 5 is often of a serious nature we might also expect fewer emoticons here.

4.2 Conversational feedback

The results presented in this section show how the affordances of this multimodal tool and the task design influenced strategies concerning *acknowledgement*, *relevant next turns* and *continued attention* in the two sessions analyzed.

4.2.1 Acknowledgement. In the analyzed material, broadcasting is rarely used only to deliver feedback: no more than three broadcasts in session 1, and one broadcast in session 5 have been coded as only consisting of feedback. This indicates that in order to begin broadcasting, the students usually have longer turns planned and that giving acknowledgement only is not seen as a strong enough reason to take the floor. Example 2 shows one of the exceptions, as Benny gives a short reply to Emma's suggestion in broadcasting.

Example 2: Audio used for feedback only – Session 5

00:35:15 – 00:35:22 Emma broadcasting
maybe we'll do both medias as well as the (Dv) survival one is media-ish as well (.) or

00:35:16 Denise video
Big smile

00:35:27 – 00:35:27 Benny broadcasting
Sure

Whereas it is impossible to give simultaneous audio feedback in this platform, acknowledgement can be given via other modes instead. Almost half of the text messages analyzed in each session are responses to what is being said during broadcasting (see Example 3).

We also find examples of the video mode being employed to deliver feedback in both sessions analyzed, but more feedback in the thumbnail video images has been identified in session 1 than in session 5. All participants can be seen to smile in the thumbnail images, but as was seen in Tables 4 and 5, some smile more than others. Only two of the participants in session 1 (Adam and Emma) use emoticons, and in session 5, only Adam uses these cues. Often the person using the emoticon can be seen to smile in the thumbnail video image at the same time, so the thumbnail smiles and the emoticons mutually reinforce each other (see Example 1).

Only one participant, Adam, uses the votes. This is in response to questions posted in the audio, and with one exception the votes are all used in connection with questions concerning the practical continuation of the session. In the following example, Adam first votes "yes" during the audio broadcast, and then reinforces his vote in audio while also explaining his response.

Example 3: Multimodal feedback in Session 1

0:43:30 – 0:43:36 Filip broadcasting
eh I think that answered both the seven and eight question (.) or?

https://doi.org/10.1017/S0958344009000184 Published online by Cambridge University Press

```
0:43:37 – 0:43:47 Emma broadcasting
(Av) well yeah (had) my answers to those in basically the same thing as
history and (.) well why it started is basically the same (Al) (Bt)

0:43:37 Adam video
votes yes

0:43:47 Adam lines up

0:43:47 Benny text
yes I think we can go to 9

0:43:47 – 0:43:57 Adam broadcasting
Yeah and I believe Benny was eh (.) touching the the today's ehm (.)
eh (.) indication of the term
```

The next section deals with coherence between broadcasting utterances. Here, it is important to keep in mind that some of the seemingly incoherent utterances do receive feedback in other modes, as shown in the examples above. However, this is far from always the case.

4.2.2 Relevant next turns. According to Clark and Brennan (1991), a relevant next turn does not have to explicitly link to the previous one. Nevertheless, here it is hypothesized that in this specific context clear links are of particularly high relevance for three reasons:

1) the lack of possibilities for simultaneous audio feedback,
2) the unusually long turns at talk, and
3) the language learning setting.

Here, the investigation concerns the co-creation of what henceforth will be referred to as *strong local coherence*, indicating that there is a clear connection between adjacent broadcasts. This category includes explicit acknowledgement, adjacency pairs and utterances displaying strong links through cohesive devices such as *anaphoric reference*, *lexical repetition, clarification requests* and *conjunction*.

Example 4 illustrates strong local coherence between two adjacent utterances, being both a relevant next turn, as the second part in an adjacency pair, and including clear acknowledgment.

Example 4: Strong local coherence – Session 5

```
00:22:15 – 00:22:38 Denise broadcasting

eh I didn't know if I followed you Emma there but I thought this national
tests were something that (.) were sent away to be corrected by other (El)
teacher not the ordinary teacher but in your case she (.) she looked it
over (.) on ((smiles)) on the (test so) (.) for you (.) is that correct?

00:22:28 Emma lines up

00:22:38 – 00:22:56 Emma broadcasting
```

https://doi.org/10.1017/S0958344009000184 Published online by Cambridge University Press

```
yeah that's exactly what I mean (.) while we were taking the test she
would walk around (.) look at it and (Dv) basically say well this is wrong
think about this thing instead and then you'll get it right so we got a
lot of hints which meant that we did better than we really should have

00:22:44 Denise video
small smile
```

More than half of the utterances in session 1 are linked by strong local coherence, and there is a slight increase in the feedback ratio in session 5. However, a striking number of broadcasts do not clearly link to the previous one. One factor to keep in mind is that, as Clark and Brennan (1991) state, acknowledging cannot continue "ad infinitum", but at certain points it is relevant to begin new threads. For example, broadcasts which respond to a question asked in a previous broadcast might be considered the potential end of a thread. However, since most of the broadcasts here not only comment on the previous one, but also introduce new information, one would expect some kind of acknowledgement.

Many of the utterances which do not display strong local coherence are linked in other ways. For example, often the contributions relate directly to the main task, sometimes positioning the statement in relation to the current context through connectives and misplacement markers. Here, coherence is created on a global level above the immediately local one, as illustrated in Example 5.

Example 5: Global coherence – Session 5

```
00:17:55 – 00:18:12 Emma broadcasting
(eh) I don't think (.) eh it's (still) here that you get any grades or
something until (.) högstadiet (.) and (.) definitely you don't really
have much (.) that someone will see at home (.) eh you can actually hide the
grades I guess ((smiles))

00:18:23 – 00:19:41 Benny broadcasting
coming back to to the question that was in the in the (.) instruction so
the national eh curriculum and and so forth (.) one thing that worries
us being in a (.) a state that normally comes in the last of eh the last I
mean we we sometimes beat Alaska and sometimes we beat Alabama but we're
down at the bottom [...]
```

In some instances where the immediately following turn is irrelevant in the local context, a response is provided in a later turn instead. For example, a student might line up in response to a topic introduced early in the turn, or even in a previous turn, but the current speaker might shift topics during the ongoing broadcast, or a question might be asked, but if it does not receive its answer in the immediately following turn it might appear in a later turn. This relates to the fact that the students sometimes participate in intertwined threads, where they depend on extended local coherence (similar to coherence in text-based CMC; cf. Herring, 1999; Condon & Cech, 1996). Because of disrupted turn adjacency, coherence is created above the level of single utterances also in this context. One example of this is when Adam talks about private and public schools, and four turns later, Denise links her contribution with his by stating "yes I think that this free school you talked about l- (.) for a

https://doi.org/10.1017/S0958344009000184 Published online by Cambridge University Press

moment ago Adam is increasing in whole Sweden''. Five of the messages in session 1 receive feedback in a later turn and three of the messages in session 5 do so.

Whereas irrelevant next turns are sometimes compensated for through extended local coherence, in both sessions we also find examples of problematic communicative breakdown. For instance, sometimes feedback is elicited but none is given. This happens both in cases where speakers express uncertainty, and where they specifically ask questions which do not receive replies. Example 6 illustrates this, as Emma's question never receives any reply after Benny returns to a previous topic.

Example 6: Communicative breakdown – Session 5

```
00:25:11 – 00:25:27 Emma broadcasting
I can't remember did (they) say that eh (.) one particular school (there)
or male or female did better and (.) eh it it sort of averages out in total
or (.) I really can't remember

00:25:31 – 00:26:14 Benny broadcasting
ie I I want to say though that I like the Swedish system there where
it's not eh (.) eh (.) dependent on the income level or the financial
situation [...]
```

Another type of communicative breakdown is exemplified in long pauses. Example 5 illustrates how there might be a correlation between pause length and coherence strategies. Here we find quite a long pause between utterances, and it might not seem very relevant to the participants to create a link to the previous broadcast. As was seen in Table 3, the pauses are somewhat shorter in session 5 than in session 1, but we find examples of very long and awkward pauses in both sessions. Sometimes these are commented on by the participants, as for example in session 5, where, after a 21-second-long pause, Emma takes the floor again and states: "okay so either I killed the conversation again or it's eh (.) that we (.) feel th- that we have done this topic eh which is it? ((smiles))''

4.2.3 Continued attention. The third type of positive evidence identified by Clark and Brennan is *continued attention*. They argue that an important way of indicating continued attention is through gaze. Even though there are video representations of participants in conversation in the current material, gaze is not as relevant in desktop video conferencing as in face-to-face interaction; apart from when you clearly see that someone is looking away from the computer, which can be taken as negative evidence, it is impossible to use gaze as in face-to-face interaction (cf. Heath & Luff, 1992). Whereas attempts have been made to evaluate levels of participation based on video images (cf. Guichon & Develotte, 2008), the video thumbnail images employed here are not updated frequently enough to allow for strict monitoring of continued attention through this mode. Here, the importance of multimodal cues again becomes apparent, something which the designers of FlashMeeting have kept in mind when pre-programming emoticons and votes.

Another related problem concerns how to know where attention is paid during multimodal interaction. Norris (2004) argues that as analysts we can determine where participants in conversation are paying attention by looking at how they react

https://doi.org/10.1017/S0958344009000184 Published online by Cambridge University Press

to each other's actions. Thus, acknowledgements and relevant next turns are of high importance also here.

In this particular platform, one additional way of indicating that one is paying attention is by lining up to take the floor. This can function as a sign that one has a comment on something which was just said, but as previously mentioned, it might also lead to coherence problems. It could be noted here that the students taking part in this study choose to line up slightly more in session 5 than in session 1 (see Table 3) indicating that they do not see this as a problematic aspect of their interaction.

Building further on the reasoning of Norris, it can be argued that levels of attention here can be identified by investigating how modal density is created; either through modal intensity (high participation rate in a particular mode) or through modal complexity (high level of integration among different modes).

Interaction in broadcasting is a clear example of modal density through intensity. Broadcasting is by far the most prominent and intrusive mode here, and this is also where most of the topical discussion takes place. However, the high frequency of broadcasts which do not exhibit strong local coherence might leave doubt among participants as to whether the others really do pay attention. Further, in both sessions, we find examples of modal intensity in the text chat, as some students participate in separate text conversations. However, these text conversations mainly concern issues relating to technology or to organization and we only find one example of a separate text conversation dealing with the actual discussion topic.

Modal density through complexity can also be created in different ways. For instance, levels of attention are easily detected when people engage in multimodal reinforcement. Examples 1 and 3 above illustrate this phenomenon, and below is another example, as Benny first types a question in text and then reinforces it in broadcasting, while expanding his contribution with further instructions.

Example 7: Multimodal reinforcement – Session 1

```
00:16:25 Benny text
can you raise the volume

00:16:26 – 00:16:41 Benny broadcasting
yeah eh can you raise the volume a little bit if you go to preferences
options or voice there [...]
```

An affordance of text is that it is less intrusive than audio, at the risk of not being noticed. Benny's reinforcement in Example 7 above might indicate that he is uncertain whether the others are paying attention to the text chat. We find a similar example of reinforcement in session 5. However, by analyzing what kinds of topics the text chat is used for in the two sessions, we might conclude that text is given higher prominence in session 5; whereas text responses to broadcasting in session 1 mainly concerned technology or organizational issues, text responses in session 5 more often were elaborations on the topical discussion.

Other instances when complexity might indicate which modes the others are paying attention to are when actions in other modes are being recognized during broadcasting. Example 1 shows one such instance, as Filip reacts to Emma's multimodal feedback in his broadcasting. Example 8 illustrates modal density through

https://doi.org/10.1017/S0958344009000184 Published online by Cambridge University Press

complexity in session 5, as Emma notices Benny's text message during her ongoing broadcast and incorporates it in her contribution.

Example 8: Modal complexity – Session 5

```
00:46:25 — 00:47:05 Emma broadcasting
well basically you have (.) eh if you watch that you want to have the biased
view (Bt) if you're (.) aware of it oh okay you pay for (.) ((smiles)) getting
it as well (Bv) (.) so maybe you know what you're (.) buying when you're
buying it (.) and you want that pro-american things because (.) well it
feels right (Av) for you [...]

00:46:29 Benny text
fox is a paid channel

00:46:35 Benny video
Potential nod

00:46:44 Adam video
Smile
```

We only find a few examples like these in the two sessions. Again there is a qualitative difference when comparing the two sessions, as in session 5, textual comments had greater influence on the main discussion, suggesting that text was given higher prominence here. The higher modal density involving text in session 5 might indicate that students here had come to realize that the others did pay attention to this mode.

5 Concluding discussion

The results from this study indicate that the patterns found in the student interaction analyzed here relate to both tool and task design, but also to previous experiences and personal speaking styles. It is important to keep in mind that the analysis is based on a relatively small set of data, yet some conclusions can be drawn.

The interaction in which students engage here in many ways resembles that of moderated discussions, as contributions often consist of long monological turns. In relation to the impact of tool design, we can note that the turn-taking device which governs contributions in the broadcasting mode has a clear influence on the length of turns and pauses here. However, the fact that students have prepared replies to questions in advance and deliver these in order to get their grades also plays an important role. Nevertheless, alterations to the instructions presented after session 1, promoting dialogue rather than monologue by encouraging students to move away from the specified discussion questions, did not seem to have any effect on length of turns since these were slightly longer in session 5.

Further, it was shown that, contradictory to previous research, neither the affordances of broadcasting nor those of multimodality automatically lead to more even participation in the verbal modes; instead these rates seem to depend on other factors such as, for example, language proficiency and previous experiences with online communication. From a language learning perspective, it is discouraging to see that the two verbal modes, audio and text, are predominantly employed by the same

https://doi.org/10.1017/S0958344009000184 Published online by Cambridge University Press

students throughout both sessions. Nevertheless, it is positive that, for example, less talkative students can participate through the video channel by changing facial expressions. This they could not do in an audio conference, which shows that multimodality allows for active participation on different levels. Also, task design might have influenced participation rates, in that no teacher was present to lead the discussion. However, the fact that the students were assigned questions in session 1 but not in session 5 did not influence levels of participation.

The analysis of conversational feedback was based on Clark and Brennan's three types of positive evidence for grounding. As far as acknowledgements are concerned, we can note that pre-programmed emoticons and votes were not used to a great extent in either of the two sessions analyzed. One reason for this might be that in the version of FlashMeeting used here, text chat and emoticons were located on competing tabs. Relating this to Clark and Brennan's discussion of production costs, it can be noted that producing a pre-programmed emoticon or vote has a higher cost than producing a smile in the video thumbnail image; thus, either the cost was too high, or participants thought that thumbnail video smiles were visible enough. Further, the fact that we find more thumbnail video smiles and emoticons in session 1 than in session 5 may be a consequence of the more serious nature of the task in session 5. Apart from task and tool design, also personal speaking styles could have influenced the patterns found here.

With regard to the second type of positive evidence for grounding, relevant next turns, relatively few examples of strong local coherence were found. The turn-taking device might be influential here in that when the line-up function is employed, topics may shift before the person in line appears in the broadcasting window, making it difficult to create a clear link between adjacent utterances. Similarly, the cases of extended local coherence that were identified were clearly influenced by the turn-taking device in that one-way broadcasting sometimes resulted in intertwined threads. However, coherence strategies may also relate to task design, as students were involved in the main task of replying to the posted questions, and coherence was created on a global level.

The final area of investigation concerned continued attention. Here, focus was on how the affordances of the tool influenced the strategies employed to create modal density. In both sessions we found examples of modal density through intensity and through complexity. The higher modal density involving text in session 5 might indicate that students were viewing it as a valuable communication channel on a more equal level to the broadcasting, even though it was not given nearly the same prominence. The task might also be relevant here, as these discussions were framed by the teachers as an oral alternative to the traditional text-based examination.

All in all, in the current material, the discussion climate was often not very supportive, and this article will conclude with suggestions for design improvements which might have a positive influence. One might first consider how task design might be improved. By engaging the students in collaborative tasks, where they have to work together to come up with answers, conversations might become less stilted, and we might find more examples of strong local coherence. Instructions for students could also be improved. Instructors in online classes have an important role in encouraging active participation in all modes, since this will strengthen the possibilities of detecting levels of attention. Speaking without receiving any positive evidence of grounding during or after one's turn can be quite discouraging and awkward, and by

https://doi.org/10.1017/S0958344009000184 Published online by Cambridge University Press

stressing the importance of multimodal feedback, it is possible to foster an affirmative social climate. It might also be valuable to raise awareness of participation rates, both to encourage talkative students to step back, and to encourage less talkative students to take the floor more. If the instructor chooses to not be present during discussions, these issues could be addressed in connection with the sessions.

Furthermore, the findings also provide some suggestions regarding how the tool itself might be improved. First, one way to ensure that participants can contribute more easily in all modes would be to avoid competing tabs. As previously mentioned, FlashMeeting has been further developed since this study was undertaken, and so this first suggestion has already been partly implemented. Second, broadcasting might be a significant advantage when large groups are involved, ensuring clear turn-taking and the possibility for quiet students to take the floor and keep it until the complete message has been delivered. However, it may not be an absolute requirement to control the floor space with smaller groups. Instead, an open floor allowing for simultaneous feedback might be more beneficial for all involved, and allowing both options would be one solution.

References

Binti Abdullah, N. N., Tomadaki, E., Scott, P. J. and Honiden, S. (2008, forthcoming) What Goes on in a Meeting? Empirical Work. *Proceedings of the 30th Annual Meeting of the Cognitive Science Society, CogSci08*.

Clark, H. H. and Brennan, S. E. (1991) Grounding in Communication. In: Resnick *et al.* (eds.), *Perspectives on Socially Shared Cognition*. Washington DC: The American Psychological Association, 127–149.

Condon, S. L. and Cech, C. G. (1996) Functional comparisons of face-to-face and computer-mediated decision making interactions. In: Herring, S. C. (ed.), *Computer-Mediated-Communication. Linguistic, Social and Cross-cultural Perspectives*. Amsterdam/Philadelphia: John Benjamins, 65–80.

Fussell, S., Kraut, R. and Siegel, J. (2000) Coordination of Communication: Effects of Shared Visual Context on Collaborative Work. *Proceedings of CSCW 2000*: 21–30.

Gaver, W. W. (1996) Affordances for interaction: the social is material for design. *Ecological Psychology*, **8**(2): 111–129.

Gibson, J. J. (1977) The theory of affordances. In: Shaw, R. E. and Bransford, J. (eds.), *Perceiving, acting, and knowing*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Gibson, J. J. (1979) *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin.

Goodwin, C. (2000) Action and embodiment within situated human interaction. *Journal of Pragmatics*, **32**: 1489–1522.

Guichon, N. and Develotte, C. (2008) Learning to manage multimodal resources for synchronous online language teaching. Paper presentation at *EUROCALL 2008*.

Gunawardena, C. N. and Zittle, F. J. (1997) Social presence as a predictor of satisfaction within a computer-mediated conferencing environment. *American Journal of Distance Education*, **11**(3): 8–26.

Hampel, R. (2006) Rethinking task design for the digital age: A framework for language teaching and learning in a synchronous online environment. *ReCALL*, **18**(1): 105–121.

Hauck, M. (2007) Critical success factors in a TRIDEM exchange. *ReCALL*, **19**(2): 202–223.

Hauck, M. and Youngs, B. L. (2008) Telecollaboration in multimodal environments: the impact on task design and learner interaction. *Computer Assisted Language Learning*, **21**(2): 87–124.

https://doi.org/10.1017/S0958344009000184 Published online by Cambridge University Press

Heath, C. C. and Luff, P. K. (1992) Disembodied interaction: asymmetries in video mediated communication. In: Button, G. (ed.), *Technology in Working Order: Studies of Work, Interaction, and Technology*. New York: Routledge, 140–176.

Herring, S. (1999) Interactional coherence in CMC. *Journal of Computer-Mediated Communication*. http://jcmc.indiana.edu/vol4/issue4/herring.html

Hutchby, I. (2001) *Conversation and Technology. From the telephone to the internet*. Cambridge: Polity Press.

Jefferson, G. (2004) Glossary of transcript symbols with an introduction. In: Lerner, G. H. (ed.), *Conversation Analysis: Studies from the first generation*. Amsterdam/Philadelphia: John Benjamins, 13–31.

Kress, G. (2000) Multimodality. In: Cope, B. and Kalantzis, M. (eds.), *Multiliteracies: Literacy learning and the design of social futures*. London: Routledge, 182–202.

Kress, G. and van Leeuwen, T. (2001) *Multimodal Discourse: The modes and media of contemporary communication*. London: Arnold.

Kress, G., Jewitt, C., Osborne, J. and Tsatsarelis, C. (2001) *Multimodal teaching and learning: The Rhetorics of the Science Classroom*. London and New York: Continuum.

Leather, J. and van Dam, J. (eds.) (2003) *Ecology of Language Acquisition*. Dordrecht: Kluwer Academic Publishers.

McIsaac, M. S. and Gunawardena, C. N. (1996) Distance education. In: Jonassen, D. H. (ed.), *Handbook of research for educational communications and technology*. New York: Macmillan, 355–395.

Norris, S. (2004) *Analyzing Multimodal Interaction. A methodological framework*. New York & London: Routledge.

O'Dowd, R. and Ritter, M. (2006) Understanding and Working with 'Failed Communication' in Telecollaborative Exchanges. *CALICO*, **23**(3): 623–642.

Okada, A., Tomadaki, E., Buckingham Shum, S. and Scott, P. J. (2007) Combining Knowledge Mapping and Videoconferencing for Open Sensemaking Communities. Conference on Open Educational Resources 2007, Logan, Utah.

Säljö, R. (2000) *Lärande I Praktiken: Ett Sociokulturellt Perspektiv*. Stockholm: Prisma Bokförlag.

Schegloff, E. A. (1968) Sequencing in Conversational Openings. *American Anthropologist*, **70**(6): 1075–1095.

Scott, P. J., Tomadaki, E. and Quick, K. A. (2007) The Shape of Live Online Meetings. *International Journal of Technology, Knowledge and Society*, **3**(4): 1–16.

Scott, P., Castañeda, L. J., Quick, K. and Linney, J. (2008, forthcoming) Synchronous symmetrical support: a naturalistic study of live online peer-to-peer learning via software videoconferencing. *International Journal of Interactive Learning Environments*.

Tammelin, M. (2004) *Introducing a Collaborative Network-based Learning Environment into Foreign Language and Business Communication Teaching: Action Research in Finnish Higher Education*. Media Education Publications 11. Department of Applied Sciences of Education, University of Helsinki. Helsinki: Yliopistopaino.

van Lier, L. (2004) *The Ecology and Semiotics of Language Learning: A Sociocultural Perspective*. Dordrecht: Kluwer Academic Publishers.

Veinott, E., Olson, J., Olson, G. and Fu, X. (1999) Video Helps Remote Work: Speakers Who Need to Negotiate Common Ground Benefit from Seeing Each Other. *Proceedings of CHI'99*: 302–309.

Vetter, A. and Chanier, T. (2006) Supporting oral production for professional purposes in synchronous communication with heterogeneous learners. *ReCALL*, **18**(1): 5–23.

Vygotsky, L. (1986) *Thought and language*. Cambridge, MA: The MIT Press.

Warschauer, M. (1996) Comparing face-to-face and electronic discussion in the second language classroom. *CALICO Journal*, **13**(2): 7–26.

https://doi.org/10.1017/S0958344009000184 Published online by Cambridge University Press