In search of the virtual interaction order: investigating conduct in video-mediated work meetings

Christoph Meier
DFG-Forschungsprojekt "Telekooperation"
Justus-Liebig-Universität Giessen
Institut für Soziologie
Karl-Glöckner Str. 21/E
35394 Giessen
Fed. Rep. Germany

Giessen, 28.02.1998
Abstract

While the possibility to conduct "face-to-face" interaction on the basis of video conferencing technology has been with us for a while now, in what ways such interactions differ from the "real thing" is still not clear. As the (mostly laboratory based) studies of video mediated interaction have provided at times conflicting results, an alternative research strategy is proposed: making audiovisual records of such interactions the starting point for research and subjecting them to a qualitative analysis. When the aim is to get at the specific quality of video-mediated interaction, concepts which were proposed by Goffman for the investigation of participation and involvement in face-to-face interaction (e.g. "footing" and "region") appear to be particularly fruitful. The analysis of a short episode of a videoconference (based on an audiovisual record) shows how participants establish different participation frameworks by means of bodily orientation and voice quality. Also, it emerges that there are activities like e.g. laughing that cannot be easily pursued jointly across an audiovisual link and thus result in a temporary splitting up between the two sites. What emerges more generally from this analysis is that qualitative analyses of technically mediated interactions rather than any other approach provide opportunities for discovering their specific qualities.

1 "Virtual" face-to-face interaction

When persons come together for activities like having breakfast, car-pooling to the place of work, business meetings, jointly drafting a contract, operating on a patient or sharing a game of billiards, they are involved in what Goffman has termed encounters or face engagements. Encounters typically involve a physical coming together, an "eye-to-eye ecological huddle" providing the opportunity to perceive the others perceiving oneself, a mutual openness for communication and heightened mutual relevance of actions, the maintenance of a single focus of cognitive and visual attention and the sustained and intimate coordination of actions (Goffman 1961:18; 1964:64; 1983:3). The possibility of engaging another person face to face while being many miles away has been with us ever since the - commercially unsuccessful -
introduction of "picturephone" video-teleconferencing services by AT&T in 1970 (cf. Hart et al. 1995:396). Meanwhile a coffee can be shared with colleagues working at a different company site via a "video window" in a common room (cf. Fish et al. 1990), business meetings can be held via videoconferencing, texts can be jointly drafted and edited by means of computer conferencing and surgical specialists can monitor and intervene in operations proceeding in far away operation theaters (cf. Nardi et al. 1993). The issue, in what ways these technically mediated face-to-face (work) encounters are different from the "real thing" (i.e. direct and unmediated face-to-face interaction) is, however, not yet resolved. The relevance of this question not only derives from an interest in the extent to which direct face-to-face interaction can be substituted by video-mediated interaction (i.e. an interest in saving time and money for businesses and administrations). It also derives from an interest in how trust and collegiality can be developed and sustained in spatially distributed organizations. Finally, the relevance of this question derives from an interest in face-to-face interaction as the basic site of sociality - what Goffman has termed the interaction order and what Luhmann has referred to as simple social systems - and its technical extensions. In this paper, I start out by drawing attention to some problems with respect to research on video-mediated interaction. I will then propose an alternative research procedure and discuss some of the issues that surface when making audiovisual recordings of authentic work interactions in their natural context the basis of research. Following this, I present an episode taken from a videocoference that is subsequently analyzed with an eye to concepts derived from Goffman such as "participation framework" and "region".

2 Problems with existing research on video-mediated interaction

Early forms of video-mediated face-to-face interaction (e.g. closed circuit television made use of by teaching hospitals) have been employed and made a topic of research since the early 1970s (cf. Short et al. 1976 and Hiemstra 1982 for early overviews). Since that time research on video-mediated face-to-face interaction has been emanating from rather diverse academic fields: communication studies, social psychology, organization studies and the investigation of computer supported cooperative work (CSCW). Not surprisingly, the issues pursued differ: what is
communication and what are the functions of vision in it? (e.g. Kawalek 1997; Whittaker/O'Conaill 1997); how do small groups function and how does the social presence afforded by a video channel affect the mutual perception of team members? (e.g. Short et al. 1976); what is the role of information processing and decision making in organizations and how are both affected by the use of video technology? (e.g. Daft/Lengel 1984; Birrell/Young 1984); how can collaborative work be supported by means of computer hardware and software and what is the role of video in that? (e.g. Mantei et al. 1991)

Most of the work done on video-mediated interaction has followed the model of experimental laboratory research. The communication technology (e.g. telephone, videoconferencing systems) is assumed to restrict or distort the communicative cues available and therefore interactions conducted on the basis of these technologies are compared to individuals or groups working in face-to-face conditions (Culnan/Markus 1987:423). The time a group needs to solve a set problem, the quality of the decision taken, the way other persons are perceived, the (in-)equality of participation obtained - these and other aspects of the interaction are operationalized and measured in various ways (e.g. Short 1976; Hiemstra 1982; Sellen 1997). A major problem with this kind of work is not so much its limited external validity but the sometimes converging, sometimes conflicting results that have been obtained - both regarding the processes of interaction and their outcomes (see tables 1 and 2 and also Finn 1997).

Table 1: Some conflicting results with respect to the interaction processes

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>in face-to-face (f-t-f) interaction almost two times as many speaker changes occur compared to video-mediated communication (vmc)</td>
<td>there is no difference between f-t-f and vmc regarding the number and length of turns or their distribution among participants</td>
<td>vmc features a higher number of utterances which are shorter compared to f-t-f interaction</td>
</tr>
<tr>
<td>(Cohen 1984:296)</td>
<td>(Sellen 1992:56)</td>
<td>(Färber 1993:47; Weinig)</td>
</tr>
</tbody>
</table>
Table 2: Some conflicting results with respect to interaction outcomes

<table>
<thead>
<tr>
<th>problem solving groups achieve better results in f-t-f interaction</th>
<th>there is no difference in time needed to solve a problem between f-t-f interaction and vmc</th>
<th>problem solving groups achieve better results in vmc</th>
</tr>
</thead>
</table>

These conflicting findings have to do not only with rather different environments being compared to telephone and face-to-face interaction: e.g. group videoconferencing in dedicated studios, desktop videoconferencing at the workplace or long term connections between offices known as media spaces (e.g. Dourish et al. 1994). They also have to do with different system specifications (analog or digital video, high quality video or low quality video, half duplex or full duplex audio, signal delay or no signal delay), the kinds of groups observed (ad hoc student groups recruited for the experiment vs. natural work groups), the kinds of tasks that have been involved (problems contrived by an experimenter vs. natural tasks in the natural work environment) and the way data have been collected and analyzed (Finn 1997:9-17). While it has been proposed to treat these and other aspects (e.g. reward structure or the level of trust among participants) as independent variables that have to be controlled for in future experiments (Olson/Olson 1997), I am skeptical as to the possibility of controlling for two dozen or more such variables at a time. Furthermore, proceeding this way does not answer a further puzzle: the self reports of those participating in these investigations and the objective measures of interaction processes and interaction outcomes obtained at times do confirm each other while at other times they do not (Sellen 1997:98-99; Isaacs/Tang 1997:195).
When researchers have not based their investigations in the laboratory but have observed the use of video in authentic work environments, they have mostly looked at those persons actually developing computer systems with video support. While these studies tap the long term experience of users and provide findings on how users adopt a technology to their needs, there are other drawbacks: evidence is often anecdotal, the work environments described are rather special and there is a lack of reflection on how the work context (e.g. rivalry between project development groups and pressures to be loyal to one's own product) affects the use of technology (Sellen 1997:102-3). Sellen and Harper comment that "We have been, in effect, looking at ourselves" and continue that "The extensive body of experimental literature effectively stands alone and in want of complementary empirical studies of [ordinary, C.M.] 'people at work' with video" (1997:226). How such an investigation of the way ordinary users make use of video systems might proceed and what may be obtained in terms of results occupies the body of this paper.

3 Investigating video-mediated interaction

As has already been pointed out, much of the research on video-mediated interaction has taken place in the laboratory or else focused on the long term experiences of system developers. One reason why field studies of ordinary users are comparatively rare probably has to do with the difficulty of obtaining access to authentic work settings for the purpose of studying video-mediated interaction. This difficulty is compounded by the fact that the study of ephemeral interaction processes demands audiovisual documentation as opposed to such received research strategies as interviewing or observing with paper and pencil. I will expand on this claim a little.

Of course ethnographic procedures such as (participant) observation can be employed and users can be interviewed as to their experiences and impressions. However, when it comes to reconstructing properties and structures of (video-mediated) interaction processes such established procedures pose several problems. A limitation particularly relevant with respect to self-report data is that the details of a particular social situation are of interest to those involved (and thus accessible to introspection) only up to a point. The details of just how we ask for a drink, invite
In search of the virtual interaction order

somebody else to continue with his story or formulate a decision usually are "uninteresting" to us and remain "seen but unnoticed" (Garfinkel 1967:33). In addition to this problem, there are other properties to observational and interview data that limit their usefulness for an investigation focusing on properties and structures of interaction processes:

i) they are derived from retrospective, post-facto processes of condensing, coding, summing up or interpreting a prior action, process or event;

ii) in the resulting descriptions, the original action, process or event is transformed in accordance with requirements of established communicative genres (e.g. a report focussing on the "relevant" facts);

iii) such data are - in all their detail - designed for and shaped by the context of their generation and use; elicited descriptions are just as much designed with an eye to the here and now of the interview as field notes taken and codes employed represent knowledge, hypotheses and intuitions current at time of writing them down (Bergmann 1985:305-307).

An alternative is to penetrate the skin of an organization, inject a probe and actually document how collaborative work is performed in situ. In the case of our research, the probe is our videorecording equipment. The audiovisual recordings thus generated provide us with a real-time version of participants' actions and allow us to analyze authentic and contextualized processes of work and interaction in much more detail than any other procedure would afford.

Audiovisual recordings are sometimes made use of in laboratory-based research as well. However, this mostly involves coding procedures and subsequent comparisons of telephone, video-mediated and face-to-face interaction on measures obtained for e.g. the number and length of turns, the frequency of backchannels, interruptions or explicit handovers or the frequency and distribution of utterance types such as "work related" (e.g. proposals) or "socioemotional" (e.g. apologies) (see for example O'Conaill/Whittaker 1997 and Weinig 1997). One problem with such coding schemes is that the materials under investigation tend not to be looked at in any more detail than is required to allocate labels to instances. Another problem is that coding schemes mostly presuppose a fixed meaning for items independent of the local context they are embedded in. This again may lead to prematurely aborting an
investigation and explication of the interactional quality of particular items (cf. Psathas 1995:8 and Schegloff 1987:216). In contrast to this line of research, I do not consider coding procedures as a necessary or even profitable step in an investigation of video-mediated interaction. Rather, I believe that relevant properties of video-mediated interaction have to be discovered inductively: through getting familiar with the tape recordings, repeatedly viewing particular episodes, making initial noticings, preparing collections of instances of a potential phenomenon and preparing paper-based transcripts that can literally be compared side by side are some of the steps involved.

Making video recordings the basis of one’s research on video-mediated interaction leads, however, to new issues that have to be taken care of. Among these are gaining access to the field, the possibility of modifying the situation by introducing recording equipment, technical difficulties in getting the information needed for analysis on tape, the question of the objectivity and comprehensiveness of such recordings and the relative advantages and disadvantages of using a tripod plus fixed framing or alternatively moving about with the camera. As I have discussed these issues in a separate report (see "Arbeit als Interaktion: Videodokumentationen als Voraussetzung für die Untersuchung von flüchtigen Telekooperationsprozessen" at [http://www.uni-giessen/~g31047, Arbeitsberichte](http://www.uni-giessen/~g31047), I will limit myself here to a brief mention of two aspects: gaining access and technical issues.

3.1 Gaining access

Often no attempt is made to obtain permission for audiovisual recordings of authentic work processes. This is probably due to either fear of being rejected immediately or fear of losing whatever rapport has already been gained with representatives of an organization. With respect to the studies undertaken by our group, the likelihood of eventually gaining permission for such videorecordings was an important criterion in pursuing access to various organizations. Of course we were not granted permission to do so right away. However, in negotiating access with representatives of various organizations, we have made clear from the very beginning that audiovisual recordings are an indispensable component of our research procedures and that
eventually we would want to make such recordings. With respect to the study about videoconferencing at "Technics" reported on here, contact was pursued during a "trial period" of three months (during which I was allowed to sit in on these meetings). Then my desire to prepare audiovisual recordings was again discussed with the members of the group observed and I was granted permission to do so.

3.2 Technical issues with respect to documenting video-mediated interaction

In order to serve the purpose, audiovisual documentation needs to capture those aspects of the situation and the ongoing interaction that are relevant to participants in designing their conduct. To capture these aspects, some prior understanding of the event studied is necessary. For some events for which participants gather around a table, relevant actions may take place more or less exclusively above the tabletop, with talk, gestures, a sheet with figures or a new haircut at one time or other constituting relevant foci of participants' attention. For other events, relevant actions may also take place below the tabletop, where kicks or nudges might be exchanged.

As we assumed that participants in the videoconferences studied would be confronted with much the same conditions and problems regardless of the particular location they participated at (e.g. being able to hear and see those at the other location only via loudspeakers and a big TV screen, being confronted with a slight delay of sound and image, not being able to pick up some of the things that are uttered in soft voice at the other location), we decided it would be sufficient to document what is going on at one site rather than both sites.
A view of the meeting room at Wesseling, where the videoconferences were recorded. The TV set on the left usually displays a view of the colleagues at Karlsruhe, the one on the right either the picture of the local group as it is transmitted to the far end or else some slide that has been captured with the document camera.

The positioning of our camera and the selection of the frame was guided by the assumption - an assumption that also appears to be underlying the design of the videoconferencing equipment - that whatever is located above the tabletop (torsos, arms, faces, papers, computers) is relevant to the interaction.

Picture 1:
A view of the meeting room at Wesseling, where the videoconferences were recorded. The TV set on the left usually displays a view of the colleagues at Karlsruhe, the one on the right either the picture of the local group as it is transmitted to the far end or else some slide that has been captured with the document camera.

Picture 2:
Picture of the participants at Karlsruhe as they become visible for their colleagues at Wesseling

Picture 3:
Picture of the participants at Wesseling as they become visible for their colleagues at Karlsruhe
As we were concerned to proceed with our investigation in a minimally invasive manner, we searched for ways to avoid cluttering the meeting room with cameras and other recording equipment. We discovered that it was possible to obtain some of the videosignals we deemed necessary from the videoconferencing equipment itself and thus employed only one camcorder\(^4\) and a fairly small microphone in recording these meetings. The different signals were then fed through several video recorders and eventually through a four-in-one image splitter onto a videotape recorder capturing a split image. This provided us with a synchronized picture of what is going on in the meeting room at Wesseling and what is visible to participants in Wesseling at any moment (see schematic 1). All recording equipment was positioned behind the videoconferencing system in order not to be too conspicuous.
Schematic 1: The synchronized views and their origin
4 Videoconferences at "Technics"

The videoconferences studied occurred at "Technics", the German subsidiary of a multinational corporation. The group of approximately one dozen high level managers observed is called "services group meeting" (SGM) and meets on a biweekly basis. Until about January 1996 they used to meet at either Karlsruhe, the German headquarters, or at Wesseling, the largest of several other company sites. Thus, more or less half the group had to fly to the other site at regular intervals in order to participate. Eventually it was decided to invest in videoconferencing equipment and now these meetings are regularly held as videoconferences. They usually begin around 8:30 in the morning and last somewhere between 12:30 to 2:00 p.m. In addition to these meetings, the group meets face-to-face once every three months for a two-day workshop. Activities in the videoconferences include a review of the minutes and short statements as to the current status of projects and issues, the discussion of current financial performance and possible measures, reports and presentations by either members of the group or invited guests, as well as discussing and taking decisions on issues ranging from advertising campaigns to work schedules. While most members of the SGM fairly regularly attend meetings either at Karlsruhe or Wesseling, some members participate at Karlsruhe one time and at Wesseling the next - depending on where they have business to do.

4.1 The data

Proceeding along the lines argued for in this paper, the kinds of materials one ends up with for study consists of hours and hours of videorecordings - in this case approximately 25 hours. Looking at these can be overwhelming and intimidating. Where should one start with an analysis? One possibility is to pursue a "top down" approach and search for something like a recognizable episode that can then be repeatedly looked at in an effort to identify its constituent components (Erickson 1982:218). Another possibility is to begin by analyzing the turn-taking arrangements as the basic organization of any form of interaction (Schegloff 1987:208). The tack taken here was different again: familiarizing ourselves with the recordings, at one of
Around 8:30 a.m. the connection between the two rooms has been set up and participants have been gathering in the videoconferencing rooms at both locations. Johannes, the leader of this management group, has greeted the participants and made a short pep talk mentioning some of the challenges lying ahead (this being the very first meeting after the Christmas/New Year break). He has also welcomed a new member to the group (Heiner, located at Wesseling), with whom he then engages in some talk about a meeting they both attended. In response to this, Heiner also gives his impression on how the meeting went and this is where the transcript sets in (an
overview of the transcription convention employed is provided in the appendix). Picture 4 provides an initial view of the scene and the names of participants. According to company policy, participants address each other by first name and either "Du" or "Sie", depending on their mutual standing. Johannes is only partly visible as the picture-in-picture mode of the videoconferencing equipment is activated.

1 ((parallel zur Äußerung von Heiner Geflüster))
   ((whispering in parallel with Heiner's utterance))

2 Heiner  ne gÜte:: ( - ) ausgangsposition jets hA::on.
   are:: ( - - ) in a good position to stAr:: from.

3 Carla?  <<pp>(                     )>>
   <<pp>(                     )>>

4 (- - -)

5 Johannes  okeh.
   all right.

6 (- - - -)

7 <<f>SUper. (.hh) so. e dann fang wer An,>
   <<f>gREat. (.hh) okay. em now let's begIn,>

8 (- - -)

9 e: zunächst e:: (.) damit wer das
   em first of all em (-) so that we

10 nich  wieder  zu  spt  machn
   do n't  do  that  late  again
Johann turns towards his colleagues in Karlsruhe and remains in that position for some time.

Johannes: who is going to take the minutes?

Johannes: last time I think we said...
In search of the virtual interaction order

em:

holger war der letzte,
holger was up last,

Johannes
who is ((no sentence-final verb?)) now:

Michele
<<p>lempp>
<<p>lempp>

Johannes
<<p>leh
<pp>ehh>
=el:

Heiner
<<p>eh
<<p>eh

Johannes
lempp
le mpp

?
In search of the virtual interaction order

42 Mich.R ehh hh hh / hh
        ehh hh hh / hh

43 Mich.D << p>heh heh heh | (mi|r ham) =
        << p>heh heh heh (with lempp ) =

44 Johann?

45 Mich.D =<< p>(mit lempp) ei(nen mit) el. >
        = << p> (we have) (one) el. >

46 Johann? =(     ) is still.  
        = (     ) keeps silent.

47 Mich.R << p>heh hehheh>
        << p>heh heh heh>

48 Mich.R =<< p>heh heh heh (heh) . hh>
        =<< p>heh heh heh (heh) . hh>

49 Jürgen? << p>el: em (o pe q)u er>
        << p>el: em (o pe q)u ar>
In search of the virtual interaction order

picture 7

50 ( - )

51 Johann
< ER (wie) R IM:bech.>
<AR (like) RIM:bech.>
<s slowly, with relish >

52 ?
(bach)
(bach)

53 ( - )

54 Johannes
RAYMbech.
RAYMbech.

55 ?
eh(eheh) (eh h(eheh))

56 Mich.R
n:::.
<< ff, dim> lempp war doch oder
no:::.
<< ff, dim> lempp was ((part.)) or

57 Johannes
(          )
(          )

58 Heiner
<< p>(hhehh)>
<< p>(hhehh)>

59 Mich.R
wie war da(h)s;;>
wasn't that so;;>

60 Christoph
<<all>hm|hm>
<<all>hn|hn>

61 Mich.R?
<< p>heheheh>=
<< p>he heheh>=
Johannes?  

Mich.R  .hh war das nich lempp?  
.hh wasn't it lempp?  

Mich.D  RAsemann <<p>kommt noch vor>  
RAsemann <<p> precedes ((part.))  

<<p>rim bach.>  

em RA semann.  

(- )  

Mich.R  wie geht das denn DAmit;  
how would THAT work out;  

?  
(a)( )  
(a)( )  

?  
<< ff>(oh: :)(  
<< ff>(oh: :)(  

?  
<< ff>hah hah hah ha::: .>  
<< ff>hah hah hah ha::: .>  

Heiner  hh hhh(  
hh hhh(  

Carla leans forward and beats on the table with her fists  

Johannes  ( )ie sacht aber (n)ix =  
((pron.)) doesn't say a thing=  

?  
=(  
=(  

Mich.R  e|hh =  
'e|hh =  

Holger  (wenn die)=  
(if she)=  

Heiner points towards the TV set  

((weiter lachen u. stimmen in karlsruhe))
In search of the virtual interaction order

((more laughter and voices in Karlsruhe))

78 Mich.R
= ehh=
= ehh=

79 Mich.R?
= heheheh (eh)
= heheheh (eh)

80 Holger
=(jetz geschickt is sagt) | (die) VO n |
= (is smart she'll now say) VO n =

81 Mich.D
(da) haste (an)

(what a’)

picture 8

82 Mich.R?
= ( )
= ( )

83 Holger
=(rasemann,) ( )=
= (rasemann,) ( )=

84 Mich.D
mein Lieber kollE:ge. (da)=
my dear colLEa:gue. (what) =

Heiner makes a gesture towards Mich.R

85 Holger
= ( )
= ( )

86 Mich.D
= ha ste aber wieder (g)lück gehabt du.
= (a) lucky turn for you.

87 ?
= heh heh heh =

88 ?
= ((husten))
= ((coughing))
Johannes? aber wie Vorsich(tich) ( ) =
how very care(fully) ( ) =

? =ehIH(heh heh)=
=ehIH(heh heh)=

? = (heh)
= (heh)

Mich.D = ich seh das nmllich grade
= i just noticed that

r Asemann steht vor rimbach =
r Asemann precedes rimbach =

Johannes ( ) (der) rimbach =
( ) (def. article) rimbach =

Johannes = (war) vorsischtsch ne?= = (was) careful wasnt he? =

Carla? (o) ( )
(o) ( )

? =eh eheheh
=eh eheheh

Heiner? heheheh

((sniffing/breathing?))

Johann? (was is los,) was is LOS mit dir.
(what’s the matter) what’s the matter with you.

? (heh heh heh) =
(heh heh heh) =

?1 = heh)
= heh)

?2 (eheheheh) ( ehe he heh)

Mich.D << p>(jetzt hab) ich (wieder) einen >=
<< p>(now i have) got one credit >=

? <p>.hhh> =
<p>.hhh> =
In search of the virtual interaction order

Mich.D $=<< \text{p>gut du.}>$

$= <<\text{p>(again) ((pers. pron.)) .>}$

Burkhard ((hohe Stimme, laut aber weit weg von Mikro?))
((high pitch, loud but far away from mike?))

fast LEArning (is das)
(that is) fast LEArning

Burkhard (  )
(  )

Heiner f(h)ast learning.

f(h)ast learning.

Mich.R $<< \text{p>(hih hih| hih)>|}$

$<< \text{p>(hih hih| hih)>|}$

Johannes (  ) also carla,
(  ) so carla,

Johannes es hat dich ( - ) $<\text{p>erWicht,}>$
there is no ( - ) $<\text{p> escape,}>$

Johannes $<< \text{p> (  )}>$

$<< \text{p> (  )}>$

$<< \text{p> (  )}>$

$<< \text{p> (  )}>$

Mich.D $<< \text{p> (da hab ich wieder einen gut)>}$

$<< \text{p> (now i've got one credit again)>}$

Mich.D $<< \text{p> (du.)>}$

$<< \text{p> ((pers. pron.))}$

Johannes (eheh heh heh)
(eh eh heh)

Mich.D $<< \text{p> (  )}>$

$<< \text{p> (  )}>$

Mich.D $<< \text{p> (  )}>$

$<< \text{p> (  )}>$

Mich.R? hh hh

hh hh
In search of the virtual interaction order

Mich.R  Bitte, sorry,

Mich.D  (da) hab ich wieder, (now) i've got one

Carla?  ( ) ( )

Johannes  e dann hab isch noch (.) (n)e frage
          em now i still have (a) question

Mich.R?  ( ) (dhanke) (heh)>
          ( ) (thahnks) (heh)>

Johannes  =gibts noch punkte fr die agenda?
          = are there more items for the agenda?

Mich.R?  = (heh) ( .hh )>
          = (heh) ( .hh )>

Richard lifts his right hand and looks towards the TV sets

Johannes turns towards the TV sets and remains in that position

Mich.R  ((clears throat))

Richard lifts his right hand and looks towards the TV sets

mit draufpackn e::. ( - - ) es a pe en te
### 4.1.2 A quick tour of the episode

The scene I am concerned with here occurs rather early in one of the regular meetings of the group observed. I will first provide a quick tour of what transpires during that minute and then look in some more detail at particular moments.

In this episode, Johannes initiates a move to officially open the meeting (line 7) and, in turning his body away from the TV sets and towards his colleagues at Karlsruhe, asks who is going to take the minutes (line 12). Participants at both locations begin to shuffle through papers. While those at Karlsruhe are more active in assisting Johannes in reconstructing what went on last time and who is up with taking the minutes (lines 19 through 31), those located at Wesseling remain almost completely silent. Michael D. is the exception, softly uttering the last name of his colleague Richard ("Lempp") located at Karlsruhe (lines 32 and 36). When that name is uttered again at the other location (line 40), several participants at Wesseling begin to laugh (lines 42-43), while at Karlsruhe Richard gesticulates and indicates objection.\(^7\)

While at Wesseling, Michael D. and Michael R. dwell on the apparently funny side of things (lines 43-48), at Karlsruhe, Jürgen resumes going through the alphabet (line 49). As he reaches the letter "R", several of his co-located colleagues bring up Michael Rimbach as a candidate for taking the minutes (lines 51-54). Immediately following this there can be heard some laughter issuing from Karlsruhe (line 55). At Wesseling Michael R. objects (lines 56, 59 and 63) and his colleague Michael D. remarks that "Rasemann precedes Rimbach" in alphabetical order (lines 64-65). Michael R. immediately takes this up (lines 66 and 68) and this is followed not only by loud laughter in Karlsruhe but also by Carla Rasemann beating with her fists on the table in a big display of being found out (lines 69 through 72).

While laughter and amusement continues at Karlsruhe for quite some time, participants at Wesseling engage in more sober talk amongst themselves about what just happened: Holger warns that Carla Rasemann might try to claim that her name is
"von Rasemann" and that she is not up after all (lines 76, 80 and 83); Michael D. repeatedly points out to Michael R. how lucky the latter has been and that he now owes him a favor in return (lines 84+86 and again at lines 104+106, 118+119 and 124+125). Those at Karlsruhe can then be heard to comment on the scene as well - Johannes commenting on the way Michael R. conducted himself (lines 89, 94+95) and (presumably) Johann asking Michael R. for an explanation (line 100). After Burkhard has provided an epigrammatic interpretation of the scene (their's being a "fast learning" [organization]), Johannes formulates a conclusion (line 113 and 115) and shortly afterwards initiates a move back to business (lines 127, 129 and 131).

4.2 Nothing new?

What transpires during this effort of assigning the task of taking the minutes is that participants at both sides try to nominate persons at the other locale and share a laugh when they appear to be successful. While participants seem to be teaming up along the lines of locality, it also emerges that there are limits to the sportsmanship and the "we win, you loose"-approach they display. This becomes clear when, along with Carla's display of "being found out", participants at Karlsruhe share laughter among themselves.

In research on the interaction in videoconferences a tendency for polarization, the emergence of locally based alliances ("us" vs. "them") and an increased likelihood of disagreement and conflict have been observed (Short et al. 1976: 137; Weinig 1997:140, 166). Such observations have been based on the analysis of topically focussed discussions. The episode looked at here, however, precedes the official introduction of a first agenda item and "discussion" proper. It is taken from a phase early on in a meeting where participants are still in the process of getting the event started. Thus, the impression that a locally based teaming up appears to be involved here merely extends previous findings. What is more interesting, however, is that common procedures which have been identified as means of establishing a collectivity or an interactional team apparently are not employed here. Such practices include addressing co-recipients as an association (e.g. "what've you guys been doing"), casting oneself as a spokesperson and employing the pro-term "we" (e.g.
"hey we got good news"), joining the production of an ongoing action (e.g. by completing an utterance some other person has begun), arguing a point in a very similar way or providing a source marker (e.g. "I fully agree with Eva with respect to ..." - cf. Lerner 1993 and Kangasharju 1996). In the scene analyzed here, no such resources appear to be employed by participants: there is neither talk about "we" or "you", nor are there any argumentative alliances openly established. Thus, in order to understand the interactional dynamics as it is played out in this episode, we need to look elsewhere.

4.3 Taking a closer look

4.3.1 Working the participation framework

Just before Johannes asks who is going to take the minutes, he turns towards his colleagues at Karlsruhe and subsequently remains in that position for some time (line 11 and picture 5). Subsequently participants at Karlsruhe take a more active part in reconstructing who is up for the task compared to their colleagues at Wesseling. For students of face-to-face interaction, this comes as no surprise.

In proposing "footing" as a lead concept in the exploration of participation and involvement in an interaction, Goffman (1981) introduces the notion of a participation framework as an improvement on the primitive concept of hearer. While persons in a social situation usually may be expected to be ratified participants, not all of them need be the addressed recipients of a particular action or utterance. Addressed recipients are specifically spoken to (e.g. by means of bodily orientation or address terms used) and an expectation of a response by them is established (Goffman 1981:133). By turning towards the persons gathered around the table at Karlsruhe, Johannes establishes them as addressed recipients and thereby indicates that it is them he expects to respond.

While those at Karlsruhe are engaged in reconstructing who is up with taking the minutes, Michael D. at Wesseling softly utters the name "lempp" (lines 32 and 36). The way he does that deserves mention: he remains absolutely immobile, neither orients towards the TV set nor towards any of the persons gathered around the at his
side, and looks at papers spread out in front of him. Proceeding thus, Michael D. makes an effort not to become visible to his colleagues at Karlsruhe as the speaker\(^9\) of what may either be a projection (who those at the other location are going to hit on) or an insinuation (who they should nominate). When the projected or insinuated item comes forth from the other location, those at Wesseling engage in displays of amusement (lines 42f. and 47f.) - this is further discussed in section 4.3.2.

A little later, when Michael R. has been nominated a candidate for taking the minutes and tries to fend this off (see lines 51-63), Michael D. announces that "Rasemann precedes Rimbach". This is immediately taken up by Michael R., who resubmits this as a question to those at the other site. Michael D., after the first few syllables of his utterance, noticeably reduces his volume (see line 64) and turns to his colleague Michael R., thus treating the latter rather than those at the remote site as addressed recipient(s). Proceeding in that way, Michael D. may be seen to treat Michael R. as "owning" the audio connection to the other side for the moment. Alternatively, and this is the interpretation I lean to, he "assists" his neighbor in getting off the hook while minimizing the visibility of this move vis-à-vis those at the other location.

Michael D. is not the only person designing his talk so as to selectively target meeting participants as addressed recipients even though he makes use of this rather frequently. Listening to the audio track of this episode, it can be noted that participants sometimes speak in a manner that seems to be designed to be heard and focused upon not only by those locally assembled but also by those at the other site. Johannes' opening move (lines 7ff.) is a case in point. The talk by other participants at Karlsruhe which follows this (lines 18-21 and 28-30) is not so designed. While Holger's talk about the possibility of Carla claiming not to be up on the grounds of being called "von Rasemann" (lines 76+80+83) appears to be - in terms of volume, voice quality and bodily orientation - designed for all those gathered around the table at Wesseling, Michael D's insistent remarks about now having one credit (lines 104+106, 118+119 and 124+125.) appear to be designed more specifically for one participant (Michael R.) only. These differences in volume and voice quality are rather difficult to represent on paper as they involve not only one or two distinct levels but rather many fine gradations.
4.3.2 Temporarily splitting up for displaying amusement

Following the already mentioned projection/insinuation by Michael D., we not only find laughter and grinning at Wesseling but also bodily reorientations. Michael D. turns towards his colleague Michael R. and Heiner turns away from the TV set and towards the two of them (picture 6). In doing so, the three of them not only mutually indicate their amusement but also display themselves as momentarily unavailable to activities that might be initiated at Karlsruhe. The same is true of participants at Karlsruhe. While Carla and Arnulf have just withdrawn their gaze from Richard, picture 6 shows Johannes, Johann and Burkhard still oriented towards him. Thus, for a moment of amusement two separate foci of attention have been established. By subsequently re-orienting to the TV sets (i.e. to those located at the other site), participants in Wesseling display themselves available again, while at Karlsruhe Johann (and more briefly) Johannes orient to the TV as if to "touch base" (picture 7).

Something rather similar occurs a few moments later, after Michael R. has nominated Carla as a potential candidate. Again, two separate foci of attention are established along with the displays of amusement and the commenting that is taking place at either side (see picture 8). Just as before, we find several participants at both sides intermittently orienting towards the TV sets. It almost seems as if there is always at least one person keeping track of what is going on at the other site.

It thus appears that participants at both sides establish a local focus of attention during humorous episodes, sharing a laugh with their locally assembled colleagues and displaying their amusement primarily towards them. Also, laughter and having fun do not seem to carry over into the other locale easily. A case in point is when participants in Wesseling share amusement about the successful projection/insinuation by Michael D. (lines 32+36). While Johannes can be seen to be smiling, no audible laughing is coming forth from that side. And when Michael D., by dwelling on the apparently funny side of things (lines 43+45), elicits a further show of amusement by his neighbor Michael R. (lines 47-48), Jürgen (at Karlsruhe), resumes the business of reconstructing who might be up for taking the minutes now (line 49). His starting up in overlap with the still ongoing laughter at Wesseling and his voice
not evidencing any trace of amusement, is indicative of participants at the two locations at this point being slightly out of sync with respect to the interactional mode (serious vs. humorous) they are currently operating in. Similarly, when those gathered at Karlsruhe begin to roar with laughter (lines 70f.), there is only little laughter in Wesseling, which is delayed in its onset and much softer (cf. Heiner at line 72 and Michael R. at lines 75+78).

5 Discussion

We have been looking at an effort to get a videoconference started. What can the analysis of such a short episode tell us about interaction processes in videoconferencing as a form of video-mediated interaction? Following closely what went on in this episode, several observations can be made:

- participants both at Karlsruhe and at Wesseling nominate (or insinuate a nomination of) a person at the other company site (i.e. for them, at the remote site) for the presumably unpopular task of taking the minutes;
- participants can and do produce utterances that are designed for different configurations of recipients and thus establish different participation frameworks; we can observe utterances that are
  a) produced in a manner accessible for participants on both sides and not addressed to any one location in particular (e.g. Johannes at lines 5-10);
  b) produced in a manner accessible for participants on both sides but specifically designed for and addressed to participants at the remote site (e.g. Michael R. at lines 56+59);
  c) designed for and addressed to participants at the local site and thus not/not easily accessible to participants at the remote site (e.g. Holger at lines 76, 80, 83 and 85);
  d) designed for and addressed to only particular persons at the local site rather than all of them and presumably not available to participants at the remote location at all (e.g. Michael D. at lines 104f., 118f. and 124f.);
- aside from bodily orientation, voice quality (i.e. fine gradations of pitch and
loudness) appears to be an important resource in achieving these participation frameworks;

- when participants treat some occurrence as funny, they tend to orient their bodies to members of their own local group; in doing so, they tend to establish separate foci of attention at each site;
- even while such separate foci are established, intermittent glances to the other site occur; somebody almost always appears to position her-/himself so as to be able to keep track of what is going on at the other site;
- laughter and having fun apparently do not carry over into the other locale easily; laughing, which in most situations calls for the finely coordinated participation of several persons in order to come off mirthfully (see e.g. Jefferson/Sacks/Schegloff 1987), appears to be difficult to achieve across an audiovisual link;
- at times participants on both sides appear to be slightly out of sync with respect to the interactional mode (serious vs. humorous) they operate in.

What emerges from this is that even after Johannes has officially opened the meeting, a common interactional space is not established in any unequivocal way. To the contrary: participants can be seen to exploit boundaries to and limits of perception. The technical setup provides them with the opportunity to design actions that are accessible to those located at the remote site only vaguely or not at all. This bears some relation to what Goffman (1959) refers to as a "backstage region" and "backstage activity". A backstage region, according to Goffman, is any place that is "bounded to some degree by barriers of perception" (1959:106). Backstage activity, accordingly, refers to activity that is designed to be accessible only to those persons who belong to the same interactional team\(^\text{10}\) (1959:112). Reformulated for the case of videoconferencing, those who are located on the same side of the audiovisual connection are in a position to perceive actions and events that - despite the high-quality videocamera and microphone - are not/not in the same way accessible to their counterparts at the other site.\(^\text{11}\) However, Goffman's dichotomous distinction of "front" and "back" does not appear to fit perfectly. For example, when softly addressing "Rasemann precedes Rimbach" to Michael R., Michael D. assists his colleague in getting off the hook without officially addressing those located at Karlsruhe. While the latter may indeed see and hear that Michael D. is engaging in some action, it may well be that they cannot pick up just what he is doing. Similarly,
earlier in the episode participants at Karlsruhe can be heard and seen to engage in talk the exact nature of which does not transpire in Wesseling (cf. e.g. lines 18+19 or 28+30). At the same time, however, participants do not challenge these actions as being performed in a deficient manner (e.g. by demanding a repeat so that they can be understood by all). While we do have in our recordings instances where participants from one side ask whether some utterance was meant to be heard by all and, if so, that it be repeated for their benefit, this seems to occur only rarely. Thus it may be profitable to conceptualize actions that are noticeable but perhaps not understandable to those at the other location as occurring on some kind of "interstage" between "front" and "backstage". One kind of observation related to the notion of such an "interstage" is that sometimes actions and activities occurring at the other location are indeed noticed by participants but are not treated as making an action on their part relevant. Thus, one speculation with respect to the interactional organization of video-mediated interaction - a speculation to be pursued in further analyses - is that the technical setup gives participants more leeway with respect to "letting pass" items that would be difficult to treat similarly in direct face-to-face interaction.

6 Conclusion

An investigation of the technical extensions of what Goffman has termed the interaction order most profitably proceeds on the basis of audiovisual records of such encounters in their natural habitat. For getting at the specific quality of video-mediated interaction, the detailed analysis of single episodes has more promise than either eliciting self-reports from participants or else the comparison of interactions conducted in various modes (e.g. in direct face-to-face interaction, via telephone or via an audiovisual link) with respect to measures such as the length and distribution of turns or the frequency of particular kinds of actions like e.g. backchannels, explicit handovers or socioemotional utterances. One may ask, however, what kinds of results can be obtained when proceeding in the fashion argued for in this paper?

Comparing communication in face-to-face interaction and videoconferencing, Weinig (1997:167) concludes that it is the technical medium that splits participants into two
groups. The analysis of the episode presented in this paper has shown that there is no single participation framework operative throughout a videoconferencing event. Participation and participation status are established in the course of particular activities - be that the fending off of an unpopular task, the formulating of a decision or the enjoyment of a good joke. Social space, as it emerges from the analysis pursued here, is not so much a matter of properties of communicative channels per se but rather a matter of the interactional uses to which participants put the resources (including the limitations of technical equipment) available to them. Rather than taking the physical separation into "here" and "there" as a given, we need to look at how participants make use of the technical equipment and the affordances it provides.

The kind of analysis pursued in this paper also leads to relevant questions for further research: just when and how do participants treat their gathering as occurring in a single social space rather than in two separate social spaces? What activities can and routinely are pursued together rather than at the separate physical locations? What are participants' options and practices with respect to participation frameworks they establish? Is there any difference between direct and video-mediated face-to-face interaction when it comes to taking up or letting pass items of conduct? The answers to these questions will get us towards a characterization of the "virtual" interaction order as it emerges in making use of the various technical systems supporting face-to-face interaction at a distance.
**Appendix: Overview of transcription convention**

The transcription convention employed follows the "Gesprächsanalytisches Transkriptionssystem" (GAT) (see Selting et al. 1997).

1) **sequencing**

A: ges tern.  
B: un was is
A: echt?=  
B: =jaja. ich=
B: =weiß auch warum  
was ha'  
(.), (. -), (. - - ), (. - - )
(1,5)  
beginning und ending of overlaps is marked by angles;  
contiguous utterances and contiguous units within utterances are marked by "=";
abrupt cut off
micropause
gap of approximately 0,25/0,5/0,75 seconds
gap of approximately 1,5 seconds

2) **characteristics of utterance realization**

a) pitch and intonation contour

*genau wie wir.*  low falling intonation contour
*genau wie wir;*  falling intonation contour
*genau wie wir,*  falling-rising intonation contour
*genau wie wir?*  high rising intonation contour
/ach \nee.

b) stress/accent

NEIn  all words (including proper names) are set in lower key letters; capital letters represent stress/accent

c) prolongation

:o:ch, ra::dio  prolongation of preceding sound

d) loudness and pacing

<<f>gestern nicht>  forte, loud
<<ff>gestern nicht>  fortissimo, very loud
<<p>gestern nicht>  piano, soft
<<pp>gestern nicht>  pianissimo, very soft
<<all>gestern nicht>  allegro, fast
In search of the virtual interaction order

3) other audible phenomena

-.h, .hhh   audible inhaling
h, hhh     audible exhaling
g(h)en(h)au aspiration or laugh pulse with a word
heheh hah  laughter
((cough)), ((noise)) audible phenomena that cannot be adequately represented

4) interpretive commentary

<nä:: wieso.> (. ) das commentary and range
<offended   >

<un dann wollt er net> <with creaky voice   >

5) visible actions and frame grabs

und gestern hat er visible actions are represented on separate lines and are marked with
|_______| respect to their range
| gesture with right arm

und gestern hat er the point in the interaction, from which frame grabs of the video are
taken, is marked

picture

6) local origin of actions

A: gestern NIch? utterances and visible actions that originate at the remote site are
B: ne; Vorgestern represented with background shading
A: ach.
? : wundert mich net

if phenomena cannot be located beyound doubt, shading is

7) transcriptionist doubt

al(s)o, (sogar) items within single brackets are in doubt
References


Barefoot, John C. & Lloyd H. Strickland; 1982; Conflict and dominance in television-mediated interactions; Human Relations, 35; pp. 559-566


Färber, Berthold; 1993; Videokonferenzen: Einsatzbereiche und Grenzen; Office Management, 41 (4); pp. 45-47


Fish, Robert S., Robert E. Kraut & Barbara L. Chalfonte; 1990; The VideoWindow system in informal communications; In: Proceedings of the Conference on Computer Supported Cooperative Work, Los Angeles, California; pp. 1-11


Goffman, Erving; 1959; The presentation of self in everyday life; New York: Anchor
Goffman, Erving; 1961; Encounters: Two studies in the sociology of interaction; Indianapolis: Bobbs-Merrill

Goffman, Erving; 1964; The neglected situation; American Anthropologist, 66; pp. 133-136

Goffman, Erving; 1981; Forms of talk; Philadelphia: University of Pennsylvania Press

Goffman, Erving; 1983; The interaction order; American Sociological Review, 48; pp. 1-17

Hart, Paul, Lynne Svenning & John Ruchinskas; 1995; From face to face meeting to videoconferencing: potential shifts in the meeting genre; Management Communications Quarterly 8 (4); pp. 395-423


Hiemstra, Glen; 1982; Teleconferencing, concern for face, and organizational culture; Communication Yearbook, Vol. 6, ed. by M. Burgoon; pp. 874-904

Heritage, John; 1984; Garfinkel and ethnomethodology; Cambridge: Polity Press


Kangasharju, Helena; 1996; Aligning as a team in multiparty conversation; Journal of Pragmatics, 26; pp. 291-319

Kawalek, Juergen; 1997; Unterricht am Bildschirm. Der Einsatz von Videokonferenzen in EDV-Schulungen; Frankfurt: Lang

Lerner, Gene H.; 1993; Collectivities in action: establishing the relevance of conjoined participation in conversation; Text, 13; pp. 213-245

Mantei, Marilyn M., Ronald M. Baecker, Abigail J. Sellen, William S. Buxton, Thomas Milligan & Barry Wellman; 1991; Experiences in the use of a media space; ACM; pp. 203-208

factors in computing systems. Bridges between worlds, ed. by S. Ashlund et al. New York: ACM; pp. 327-334


Olson, Gary M. & Judith S. Olson; 1997; Making sense of the findings: common vocabulary leads to the synthesis necessary for theory building; In: Video-mediated communication, ed. by K.A. Finn/A.J. Sellen/S.B. Wilbur. Manwah, NJ: Lawrence Earlbaum; S. 75-91

Psathas, George; 1995; Conversation analysis: The study of talk-in-interaction; Thousand Oaks: Sage


Selting, Margret et al.; 1998; Gesprächsanalytisches Transkriptionssystem (GAT); Linguistische Berichte 173; pp. 91-122

Short, John, Ederyn Williams & Bruce Christie; 1976; The social psychology of telecommunications; London: Wiley

Weinig, Katja; 1997; Wie Technik Kommunikation verändert. Das Beispiel Videokonferenz/ Münster/Hamburg/London: LIT-Verlag


---

1 This paper is based on presentations given at the 2nd international workshop on telework, Amsterdam, September 2-5, 1997 and the 5th Conference on Interaction Analysis, Utrecht, September 17th 1997.

2 For a description of the larger research project this study is part of see "English Version" at http://www.uni-giessen.de/~g31047 and "Strukturen, Dynamik und Konsequenzen telekommunikationstechnisch vermittelter Interaktion - Darstellung eines Forschungsvorhabens" under the heading "Arbeitsberichte" at the same web site.
Proceeding this way one has to be prepared to answer questions as to how those involved in the study can be protected from any adverse consequences. Anonymizing data in reports and disciplined data storage and handling procedures are important here.

This camcorder and the camera that is part of the conferencing system are both barely perceptible in picture 1. The dark spot on top of the left hand TV set is the system camera, the dark spot just to the right is the camcorder we employed.

In order to safeguard the personal rights of those observed, all names employed in this paper are pseudonyms.

The equipment (a PictureTel Concorde 4510 system) was specifically acquired to support the meetings of this group. At the time of recording participants had about twelve months experience using the system. The connection between the two sites consists of two Euro-ISDN B-channels with a bandwidth of 128 kb/sec. The system features full duplex audio and the signal delay between the two sites is approximately 0.4 seconds.

The amusement as well as the objection may be due to the fact that Richard Lempp had been in charge of taking the minutes during these meetings for a long time. Then Jürgen, a newcomer, was charged with this task for several weeks before the group — apparently at a face-to-face meeting which was neither observed nor recorded — decided to rotate this task (see Johannes' utterance in line 15-16).

While this is one of the few observations where several researchers on video-mediated interaction are in agreement, it does not amount to a major finding. In a different context Goffman (1959:92) has observed: "I do not know of any general reason why interaction in natural settings usually takes the form of two-team interplay, or is resolvable into this form, instead of involving a larger number, but empirically this seems to be the case."

With the concept of production format and its various realizations, Goffman (1981:144) has provided an improvement on the rather primitive notion of a "speaker". However, these different production formats need not concern us here.

Goffman refers to a team as a set of individuals cooperating in staging a routine or in jointly maintaining a particular definition of the situation (1959:79, 104). I want to focus, for the moment, on a more restricted sense of team as a collectivity established in interaction. For a discussion of practices employed in establishing an interactional team see Lerner (1993) and Kangasharju (1996).

To provide another example of the creation of a backstage: on one of the very first videoconferences observed, participants in Karlsruhe switched off their microphone when a short presentation in Wesseling began. On the one hand, they thus avoided to interfere with the ongoing presentation (to which they could still listen to) due to any noises resulting from their getting coffee and snacks. On the other hand they thus made their activity (i.e. breakfasting) less conspicuous for those located in Wesseling, especially the presenter.