The Collaborative Work of Producing Meaningful Shots in Mobile Video Telephony

Christian Licoppe
Department of Social Science
Telecom ParisTech
46 rue Barrault, 75013
33 1 45818116
christian.licoppe@telecom-paristech.fr

Julien Morel
Department of Social Science
Telecom ParisTech
46 rue Barrault, 75013
33 1 45817108
julien.morel@telecom-paristech.fr

ABSTRACT

In this paper we report on the first study of the uses of mobile video telephony based on the collection and analysis of naturally occurring mobile video telephony. We show how a characteristic feature of mobile video telephony, which makes it differ from any other kind of mediated interaction, is that: a) the participants may orient the camera at will to shoot almost any feature within their environment; and b) what they actually show at a given moment may be (and usually is) inspected by the recipient for its relevance to the ongoing interaction, and is produced with an orientation towards such scrutiny. A specific concern of mobile video call users at any time is therefore what they should or should not show. We demonstrate how a partial solution to that problem is the reliance on a particular (full) portrait-like 'talking heads' format as an expected default mode for interaction in mobile video calls. Finally, we discuss the implications, for design, of such an empirically grounded understanding of the specific practical concerns of mobile video telephony users.

Categories and Subject Descriptors

H5.m [**Information interfaces and presentation**]: (e.g. HCI). Miscellaneous

General Terms

Human factors.

Keywords

Mobile phone, video telephony, video mediated communication, mobility, conversation analysis, privacy

1. INTRODUCTION

Since the development of third generation (3G) mobile networks, service providers have made mobile videophony services available. This marks a 'mobility turn' in videophony, which was initially developed for professional applications [13] – though with limited success [10] – and then recently extended to larger audiences and different uses with online services (via webcams and instant messaging or VoIP services that enable video communication). In these settings video communication relies on fixed cameras (orientable to some extent) rather than handheld devices which can easily be oriented at will and in any direction with one hand. The uses of videoconference systems and media spaces have been studied extensively since the early 1990s, with several different orientations: a) investigations into how video links support distributed team collaboration and informal professional meetings [8, 5, 2, 1]; b) analysis of the problems raised by video communication in the organization of interactions, such as the turn-by-turn organization of talk [21], 'frailty of the interaction frame' [4] or the difficulties raised by pointing in 'fractured ecologies' [14]; c) analysis of the tension between using video calls for interaction between 'talking heads', versus showing relevant features of the environment, or 'video as data' [17], or the consequences of video images providing access to a shared field of interaction [9, 23, 16]. These studies suggest that showing things may be more relevant and useful to collaboration than showing people in video-mediated settings. Most were carried out in professional settings, with the experimental study of home video communication in Biarritz (France) being one of the only exceptions [3, 4].

Mobile video telephony marks a 'ubiquitous computing' [22] turn in the field of video telephony in two senses at least: a) users can engage anytime and anywhere – at least theoretically – in video telephony-based interactions; and b) the use of handheld communication devices allows them to orient the camera in (almost) any direction and to show any feature in their current environment to a remote recipient. However, making video telephony ubiquitous has not made it a market success yet. Mobile video calls are still a new and emergent practice with few users. Although this type of service has been offered for two or three years, there has been little research on the uses of private mobile video calls. An exception is a recent study based on interviews and diaries which showed that 50% of calls were for 'small talk' (i.e. social calls), 28% were to show something and talk about it, and 22% were to achieve a particular goal such as coordination or practical arrangements [18]. To our knowledge, there has not been any study of the way participants manage mobile video interactions, based on the recording and analysis of actual mobile video calls.

The aim of this article is to provide such an analysis, based on a collection of naturally occurring mobile video calls. We show how a characteristic feature of mobile video telephony, which makes it differ from any other kind of mediated interaction, is that: a) the participants can orient the camera at will, to shoot almost any feature in their environment; and b) what they actually show at a given moment may be (and usually is) inspected for relevance by the recipient with respect to the ongoing interaction,

MobileHCI'09, September 15 - 18, 2009, Bonn, Germany.
ACM 978-1-60558-281-8.

and is produced with an orientation towards such scrutiny. A specific concern of mobile video call users at any time is therefore what to show. We demonstrate how a partial solution to that problem is the reliance on a particular (full) portrait-like 'talking heads' format as an expected default mode for interaction in mobile video calls. Finally, we discuss the implications of our type of analysis for the design of mobile video services and an understanding of how they are used.

2. METHOD

As a first step in the collection of naturally occurring mobile video calls, we analyzed discussions on mobile video telephony on forums, and interviewed 25 users about their mobile video telephony practices. We then engaged in the construction of a corpus based on the capture of mobile video calls. We used a functionality offered on a few mobile phone models which provide an audio and video plug to allow for the recording of mobile screen activity plus outgoing audio flux. This is the case of the Nokia N93 and N95 phones that we used in this study. This method provides a much neater image of mobile phone screens than what can be obtained with other methods such as having users wear video glasses. However, initially designed on TV sets to allow screen captures, and transposed onto mobile phones, this functionality does not enable the recording of incoming audio flux which we needed to record. We first connected such phones to DV recorders, and then connected the recorder to an additional microphone (users were asked to place the microphone not far from their cell phone, and to refrain from using their earphones) to record the incoming talk. We then gave users the whole set of apparatus, as shown in Figure 1. An unavoidable consequence of this was that the subjects would not use their usual phone for the duration of the study.

Subjects were able to avoid recording any calls they did not wish the researchers to have access to. Moreover, to be able to analyze actual video calls, authorization had to be obtained from both parties to the call. Subjects therefore had to be recruited in pairs accustomed to interacting together on a regular basis via mobile video calls. This also meant that all our pairs of subjects were closely related: couples or pairs of close friends. We were able to recruit eight pairs of users and to build a corpus of about 100 mobile video calls in this way.

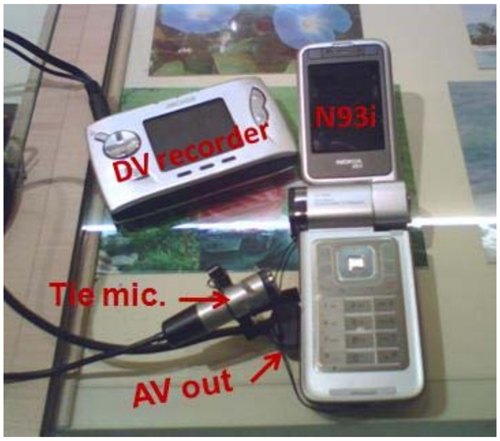


Figure 1: The recording apparatus. In this configuration, the synchronization of outgoing and incoming audio fluxes is made in the DV recorder.

3. TYPOLOGY OF CALLS

In our study we had two resources for typifying calls: first, the way subjects described them in the interviews; and, second, the reasons provided for the calls by the callers, during the conversation. These do not necessarily match. When subjects describe the kinds of call they make, they tend to frame calls within recognizable 'communicative genres' [19]. When they provide a reason for calling, during the call itself, it is shaped to fit the ongoing talk-in-interaction. Moreover, our subjects rarely used mobile video calls for conversations focused on practical arrangements, so that the third category was under-represented. We therefore found it useful to provide a refined typology of calls, consisting of four main categories.

Keeping in touch, through 'small talk'. This generic type of call constituted roughly 50% of our corpus and covered two main sub-categories: 'phatic calls' characteristic of 'connected presence' [11, 12], i.e. short video calls between subjects who are often in touch in many ways, and whose main function is to maintain the relationship rather than to talk about something specific (a kind of video 'wink'); and video calls between intimate participants experiencing a temporary separation (for instance for professional reasons, holidays, etc.). In this case the mobile video is used as a way to maintain visual contact in situations in which opportunities for face-to-face encounters have become scarce. In this configuration, the relevance of a mobile video call might be assessed with respect to the availability of other communicative resources such as computer-based video communication (webcams).

Showing things to talk about. This covers about 30% of calls in our corpus and three sub-categories can be distinguished:

- visualizing people and pets on special occasions (newborn babies, sharing social gatherings with distant friends, etc.)

- visualizing things (for instance, recent devices and appliances, cars, etc.)

- visualizing scenes, i.e. showing one's environment (a new flat, the state of a child's room, a holiday context, etc.).

Collaboration. In such calls, which comprise about 10% of our corpus, the visual component of mobile video calls is exploited for collaborative endeavors as varied as discussing the purchase of an item of clothing, getting a new computer to work, choosing a decorative color, etc.

Mobile video telephony-related calls. This category, about 10% of our corpus, concerns calls in which mobile video telephony itself or its uses are discussed within the calls: ‘demo’ calls, discussions of the interest of mobile video calls, of how to make them, etc. This category may be transitory and related to the emergent character of mobile video telephony. It was nevertheless present to a significant extent in our sample.

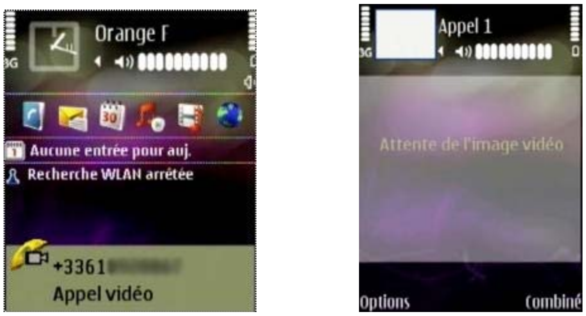
While such typologies remain useful, they do not account very well for the interactional consequences of the crucial feature of mobile video telephony, namely the possibility to reorient the camera at any time. This makes it possible briefly to show something which has become relevant in a conversation initiated mainly for small talk, thus blurring the distinction between the two first categories. Moreover, all these typologies conceal the continuous work of adjusting the camera orientation to the ongoing interaction. Showing such ‘camera work’ requires a fine-grained analysis of actually occurring mobile video calls.

4. THE ‘TALKING HEADS’ INTERACTION FRAME

At all times in a mobile video call the participants may orient the camera in any direction within the constraints of permitted or comfortable body movements (such as hand rotations and flexed arm movements), and provide their co-participant with many kinds of shots of their surroundings, within the technical constraints of the mobile camera (such as its aperture width). Both participants are generally aware of this possibility. One of their practical concerns is therefore to provide images that are meaningful and relevant to the ongoing interaction on a moment-by-moment basis. How do they practically manage that? How do they use this powerful but potentially confusing interactional resource which is specific to mobile video telephony?

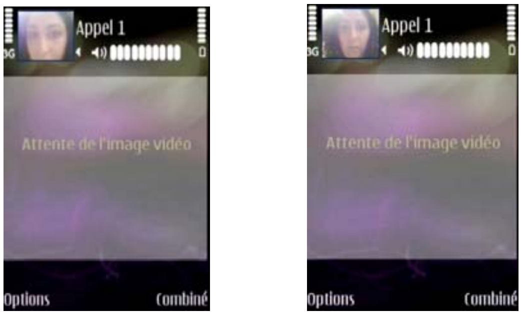
4.1 Preparing to be Seen: a Pre-connection Sequence in a Mobile Video Call

A typical example of the kinds of data that our recording methods provide, and the way they may be used to visualize the pre-connection sequences that precede a mobile video call, is shown in Figures 2 a-d.



a) b)

Figure 2: a) The state of the screen just after the number has been composed. b) The screen that appears just afterwards and can be read as a signal that the mobile video call is proceeding properly.



c) d)

Figure 2: c) The control image of the caller which appears very rapidly after screen b). The screen remains that way for a few seconds, which gives the caller time to adjust her head position. Screen d) occurs just before the image of the call recipient appears.

The adjustment of the caller’s position in the control image between Figure 2c) and Figure 2d) is particularly significant from an interactional perspective. It shows that the caller orients towards the production of a particular type of camera shot at connection: a portrait-like headshot with the whole of the head visible on the screen (in portrait mode). Such mutual orientation, which is also observed in interpersonal fixed video calls (but then usually in a ‘head and shoulders’ framing) is a way of reproducing, in video settings, the kind of ‘eye-to-eye ecological huddle’ [6] and maximization of the transaction segments overlap [7] which characterizes face-to-face interaction. The medallion portrait-like ‘talking heads’ mode which we observe in mobile video calls appears however as a specific trade-off between such interactional orientations and the particular constraints of mobile videophony as an embodied practice. Because the video communication device is handheld in that case, it is more comfortable to hold the mobile phone with a flexed arm. At that distance, and with camera apertures fixed by design, headshots are the best way to satisfy mutual orientation and maximization of transaction segments overlap for extended periods of talk. However the kind of preparatory work observed here suggests that the provision of a proper headshot also constitutes a normative expectation of participants. The following conversation will provide more evidence of this.

4.2 Noisy Environments and the Difficulties of Producing a ‘talking heads’ Interaction Format

In the following mobile video call the caller is in a bus and has called his girlfriend for a chat (Type 1 in our classification). However, there is a lot of ambient noise and she cannot hear him properly. This is marked in the first part of the conversation by the occurrence of many repair sequences and uncompleted adjacent pairs. When the transcript starts they are engaged in a localization sequence in which the caller (B) tries to determine where the call recipient (A) is.

1. B: où ça:: à (.) Nation:

where:: at (.) Nation: ?
2. A: (0.5) nan nan:: (.) ah j'vois qu'ton nez
(0.5) nah nah:: (.) ah I only see yer nose
3. A: (.) c'est trop bon: hu hu hhhhhh hu hu hu
(.) it's too good: hu hu hhhhhh hu hu hu
4. (0.5)
5. A: hu hu hu hu >hu hu hu< °hu hu hu° hum hu hu
6. (inaudible)

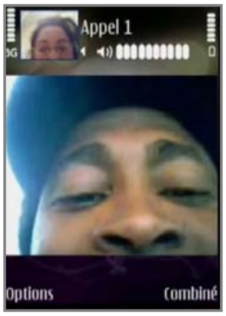


Figure 3a: The caller's image (Line 1-5)

To enable her to hear better, the caller tries to move his mouth close to the microphone of his mobile phone. Each times he does this his face gets close to the camera, so that on her screen she sees a distorted close-up of his face or of part of his face (as in Figure 3a). They have been doing this for some time and have treated it as a joke. In Line 2, for instance, she interrupts the localization sequence – although they have not yet come to an agreement – to remark on the fact that she sees only his nose, which is therefore something that may be treated as noticeable and mentionable. She marks this as something funny by her laughs and appreciative remarks (Line 3-5). Hence, this type of image breaks with some shared expectations about how to conduct a proper mobile video call. The caller treats her remark as a pre-request or a pre-invitation to reframe his image, which he does during his utterance in Line 7 (Figure 3b). This suggests that the correct image to present is a full portrait, and that such a form of communication between 'talking heads' is a collaborative accomplishment.

7. B: quoi ! h.h.h.
what ! h.h.h.

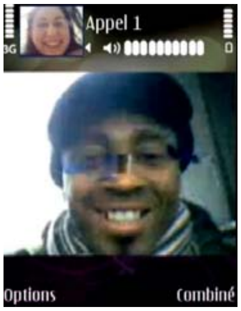


Figure 3b: The caller's 'corrected' image

8. B: ça te fait marrer: et tu (.) éh mumu tu vas voir:
it makes you laugh and you (.) hey mumu you'll see:
9. A: hu hu hu !
10. B: ah ah ah ah hhhhhh et un p'tit (.) ah::: p'tit nez:
ah ah ah ah hhhhhh and a little (.) ah::: little nose
11. (0.2)
12. A: uh: hh uh: hhh well (.) et lucien je
uh: hh uh: hhh well (.) and lucien I
13. B: allô::::
hello::::
14. (0.1)
15. A: j'te rappelle tout à l'heure hein
I'll call you back later hey
16. B: ouais:
yeah:
17. (1.1)
18. B: comment ?
what ?
19. (0.2)
20. A: J'TE RAPPELLE TOUT A l'heure yeu:
I'LL CALL YOU BACK later yeu:
21. (1.2)
22. B: pourquoi ?
why?

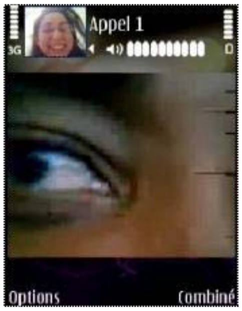


Figure 3c: His playful distortion of his head shot (Line 23)

23. A: (0.2) béh parc'que en visio comme ça c'est nul huhuhu
(0.2) *bah because in video like this it's useless huhuhu*
24. A: j'te vois pas là: c'est °bon° hh j'vois qu'ton oeil
I don't see you there: that's °it° hh I only see
your eye

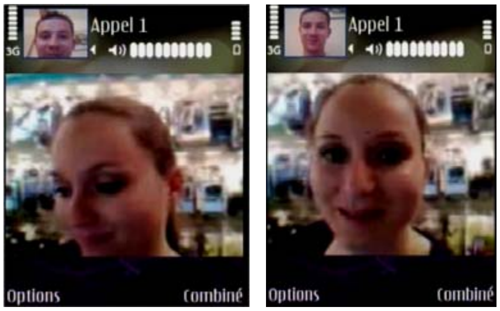
The caller then exaggerates the distortion of his image. The fact that this is a game is signaled by his announcement in Line 8 ('you'll see') and his laughs. After their joint appreciative laughs, Lines 9-12, his verification that they can hear one another ('hello' in Line 13) is treated by her as an opportunity to initiate closing arrangements. When he asks why, he also produces an exaggerated close-up, which orients a potential answer towards the visual inadequacy of their exchange. While she might otherwise have argued that they cannot hear one another properly, she follows his (visual) cue and states instead that mobile video calls are useless when one can see only an eye (a kind of close-up which he playfully exaggerates here, but which before that was a collateral effect of his trying to compensate for the ambient noise). So not only is there a strong normative expectation that makes the portrait-like head shot the proper mode of interaction in mobile video calls, with respect to which both participants orient; the norm itself is forceful enough to constitute a justification to close the call if it does not allow the accomplishment of this interactional frame.

4.3 The Impropriety of Incomplete Portraits

The following excerpt starts after the exhaustion of a previous topic. After a pause, the call recipient remarks on the fact that the caller has been presenting a portrait with a missing chin (see Figure 4a) in the entire opening part of the mobile video call (Line 3, repeated in Line 5).

1. A: bon:
well
2. (0.2)
3. B: et pourquoi tu coupes ton menton quand tu m'appelles ?
and why do you cut off your chin when you call me: ?
4. A: (0.3) comment: ?
(0.3) what: ?
5. B : (1.4) pourquoi tu coupes ton menton (.) quand tu m'appelles
why do you cut off your chin (.) when you call me?
6. A: (0.6)] [oh bah chai pas:
oh um I dunno
7. A : (0.2) tiens
(0.2) *here*
8. (0.7)
9. A: tu m'vois là ?
you see me there ?
10. B: (1.4) bah:: oui:
(*1.4) bah:: yes:*
11. (0.3)
12. A: bon (le)=
well (the/)=
13. B: = c'est quand meme mieux d'avoir la tête en entier
=*it's sure better to have the whole head:*
14. B: hein sinon euh
uh (.) otherwise: er

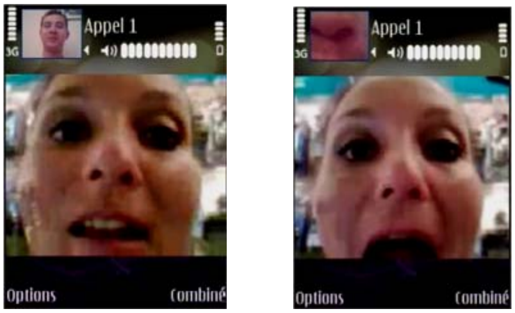
Just before answering (Line 6), the caller starts to reframe his image and produces a proper portrait, fully accomplished at the end of his utterance in Line 8 (Figure 4b), which calls for confirmation that this new frame meets his girl-friend's expectations. This shows he has treated her question not as a request for information or for an account, but as a 'pre-request' for a reframing of his headshot as a proper portrait. She confirms that she is satisfied (Line 11) and then even states the rule: it is better to provide a proper portrait with a full face (Line 13).



a) b)

Figure 4 a) Initial headshot of the caller with 'missing chin'. b) After the caller has readjusted the frame into a portrait showing full face.

15. A : ah bah
16. (0.3)
17. A : si [tu l'dis
if [you say so:
[X
18. (0.2)
19. B : ça va] axel
is it] okay axel
Y**]**
20. A : (0.3) huhun:::



a) b)

Figure 5: a) Her playful distortion of her face starting at X b) His collaborative response produced at Y, in which only his nose.

The caller responds by a kind of hedging agreement ('if you say so', Line 17). This prompts his girl-friend to display the potential consequences of improper portraits by putting the mobile phone closer to her face so that only the central part of it is visible (Figure 5a). Because it comes just after her statement of the rule and the caller's half-hearted agreement to it, this reframing, which occurs between the X and Y marks in the transcript (Line 17-19) is marked as playful. She explores the limits of what might constitute a proper frame and head shot, so as to show, by contrast, that a cut-off face is an obviously improper format for the ongoing interaction. She is clearly using the possibility to change the camera frame as an interactional resource (here to playfully make her point). The caller then plays the game by distorting his own image so that only his nose is visible (Figure

5b), which is a way of producing an alignment with her game and her expectation of a talking head without having to go back on his previous half-hearted agreement.

We have shown how participants oriented normatively towards the relevance of mutually providing a (full) portrait-like 'talking heads' video image format. Though audiences were not previously familiar with on-screen headshots (recall movie audiences' startled reactions to the introduction of headshots by D.W.Griffith in the early twentieth century), current media such as TV have made that kind of camera shot familiar and characteristic of the media rendering of intimate face-to-face conversations. It is also remarkable that in a corpus of TV ads on mobile video calls, which we obtained, participants were always depicted in a similar mode of interaction (except of course when one of them was featured using the camera phone to show something to the other).

5. 'SHOW AND TALK' OR 'VIDEO AS DATA'

5.1 Showing One's Environment as a Routine Interactional Resource

The following conversation takes place between a ten-year-old child and his older sister. It starts as a 'phatic' call with no stated reason other than an orientation towards nurturing a close relationship between siblings which relies and supports a form of 'connected presence' [11, 12]. The relevance of leaving the 'talking heads' interactional format, showing parts of the environment and talking about them, emerges from the interaction itself and is contingent on the way it unfolds, so that this particular call belongs equally to our two first categories. This is one of the main reasons why any neat typology of mobile video calls is impossible.

1. A: d'accord. (.) mais t'es où là (0.2) t'es chez maman ?
okay (.) but where are you now (0.2) you're at Mum's place?
2. B : (0.3) à la l/
at the
3. (1.0)
4. A : t'es dans ta chambre?
you're in your room ?
5. B : (0.7)
6. A : ta chambre ?
your room?
7. B : (0.1) ouais:!
yeah !
8. (0.4)
9. A: fais voir comment t'as range (.) mais allume
show me how you have tidied it up (.) but turn on
10. A: les lumières parce que j'vois rien
the lights because I can't see anything
11. (1.8)
12. A : t'as rangé ou pas ?
have you tidied it or not?
13. B : (1.9) bah:: ouais un peu:
bah:: yeah a bit:

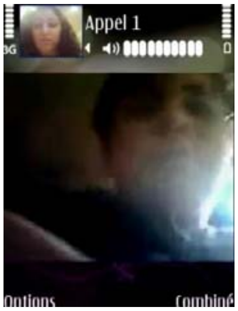


Figure 6: The video image at the start of Line 14.

14. (3.2)
15. A: ah ouais d'accord (.) c'est ta ch/ uh ! hhhh
ah yeah okay (.) this is your roo/ uh ! hhhh
16. A: pourquoi tout est par terre comme ça:
why is everything on the ground like that
17. B : (1.6) pace que j'ai joué aya: aya:::::: hier
because I was playing ya: ya::: yesterday
18. (0.8)

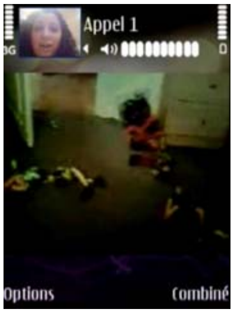


Figure 7: The call recipient reorients the camera so as to show his room, producing this image for her sarcastic appreciation in Line 21.

This sequence shows that reorienting the camera to show something relevant is always an available option. After the caller has ascertained that he is in his room she requests that he show his environment. He does not do so immediately because she initiates a repair sequence about the lights to improve the quality of the image, and then a question-answer pair asking confirmation of the fact he has tidied up the room. It is only in the long silence afterwards that he changes the camera orientation to show his environment in a way which has been made relevant by her question (i.e. his room, shown to support an assessment of its tidiness). He shoots it in such a way (showing the ground and the way it is littered) that it elicits an explicit assessment from his sister (Line 15). They collaborate to produce an assessment of his environment which is now shared visually to some degree, in line with previous studies suggesting that shared visual access to workspaces supports collaboration [23].

However, these early studies were mostly focused on fixed views of joint workspaces. A key property of mobile video telephony is that the camera can be oriented in almost any direction at any time. When the caller requests an explanation for the mess he has shown (Line 16), he has already moved the camera so that he is showing another part of his room, where his clothes are hanging – a frame whose relevance to her question is at best equivocal. This shows that there is a specific skill to shoot what is relevant at a given time for the ongoing talk-in-interaction: a very particular kind of 'director's skill.

5.2 The 'talking heads' Frame as a Default Mode of Interaction

Because the production of a given shot has to be interactionally relevant, it often leads to repair sequences and makes explicit various kinds of expectations regarding the relevance of what is featured with respect to the ongoing interaction. Showing something to the other party in mobile video calls appears to be a sensitive collaborative endeavor, as shown in the next excerpt, which occurs later in the same mobile video call.

19. A : fais voir comment tu l'as rangée: éh ! j'vois pas là:
a)
show me how you've tidied up hey ! I can't see here
20. (0.3)
21. B : (inaudible)
22. (0.1)
23. A : t'as mis ton doigt:
you got your finger on it
24. B : t'inquiètes
don't worry/
25. (0.1)
26. A : voi::là
here::.
b)
27. (1.2)
28. A : ah::::::: t'as super bien rangé
ah::::::: you've tidied up really well
29. (1.7)
30. A : bon ça va (0.1) c'est nickel
well it's fine (0.1) it's neat
31. (3.0)
32. A : bon (.) ça va (0.1) nickel
well (.) okay (0.1) neat
33. (0.5)
34. A : d'accord
okay
35. (1.5)
36. A : c'est bon raphael
that's enough Raphael
37. (2.3)
38. A : rafi c'est beau:!
rafi that's beautiful !
39. (1.[4)
[X
40. B : oui (.) c'est beau ici
yes (.) it's beautiful here
41. (1.0)
42. A : tu rentres (.) >tu vas chez mamie<
you get back (.) >you're going to Granny's<
43. A : là ou pas (.) parce qu'i pleut hein :
now or not (.) because it's raining isn't it?

The sequence starts by an explicit request for a change of frame to assess the cleanliness of the room he is supposed to have tidied up (Line 19). The production of a proper frame appears as a collaborative accomplishment, involving frame construction-oriented repair sequences: the comment on the presence of the finger is treated as a pre-request to remove it (Lines 23-24). The construction of an acceptable frame is achieved by her ratifying that he has provided a proper frame (Line 26).

Figure 8 a) The image produced in Line 25, which elicits her pre-request. b) The image he then produces and which she eventually ratifies in Line 26, thus marking the end of the frame repair sequence.

This frame-oriented repair sequence clearly shows how the production of a proper camera shot is a joint interactional accomplishment in mobile video calls. After that collaborative achievement of a proper frame, the caller produces a positive assessment of the state of the room, which also provides further evidence that a camera shot suitable to that purpose has indeed been produced (Line 28). This particular sequential organization is also reminiscent of a question-answer-assessment format which has been shown to be characteristic of classroom interactions [15]. Here we have a request (to show the room) + a non-verbal answer (production of a proper camera shot with inserted repair sequence) + an assessment (of the state of the room). It shows how the possibility to orient the camera at will may constitute a resource for collaborative teaching at a distance, combined with conventional pedagogy-oriented conversational sequences.

Figure 9: Successive camera shots produced by the brother between turns 30 to 38, and which his older sister treats as irrelevant to the ongoing talk-in-interaction.

After that initial assessment the boy carries on scanning the room with his mobile phone (see Figure 9), while his sister produces several turns-at-talk to which he does not seem to respond, either in words or through non-verbal moves. She first provides a second assessment 'well it's ok (.) it's neat' (Line 32), which she repeats in a slightly abbreviated form, thus indicating that she expects a next action on his part. She goes on providing three closing-oriented token turns which become increasingly personal and loud (Lines 36-38), and which therefore take the form of an emergent summons [20], displaying her growing impatience with his apparent lack of response.

The turn in Line 38 is particularly significant in such a progression. The term of address (first name), introduced for the first time at the end of the previous turn, is uttered at the start of the term in a familiar diminutive form, marking the fact that the girl definitely expects something of her brother, then and there, on the basis of a familiar asymmetric relationship (older sister and younger brother). The loudness marks an intensification of the summons. Finally, she transforms the topic-pre-closing token in the previous turn (‘c’est bon’ i.e. ‘*it’s okay’*) in the virtually homophonic expression ‘c’est beau’ or ‘it’s beautiful’, uttered with a slight smile and ironic pitch. This works as an ironic appreciation of the frame, based on the notion that if something is shown for a long time on screen it must have a particular relevance, such as esthetic value. The irony lies in the fact that this cannot be said of the boy’s haphazard shots of the room. It seems to be lost on him, as he flatly acknowledges the comment by repeating it (Line 40).

Before that, however, during the silence that follows the girl’s last summons (at position X in the transcript), he changes the camera’s orientation and produces a headshot, thus returning to the ‘talking heads interaction format’ (Figure 10) This proves to be a satisfying answer to the caller’s summons, for she then moves on to another topic (Line 42).

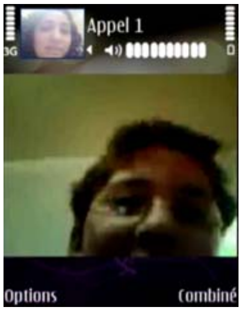


Figure 10: The return to a headshot in the middle of Line 39 at position X in the transcript.

This shows retrospectively that the action expected by the caller (as attested by the intensification of the summons), was a return to the ‘talking heads’ mode of interaction. The ‘talking heads’ format therefore appears as the default mode of interaction in mobile video calls when there is nothing to show that is or might be relevant or made relevant with respect to the ongoing interaction.

6. DISCUSSION AND IMPLICATIONS FOR DESIGN

We have presented here the first analysis of mobile video calls based on naturally occurring mobile video conversations. It shows that, contrary to media richness or social cue models, mobile video telephony cannot be treated as talk plus image. Nor can it be straightforwardly compared to video telephony in settings with fixed or semi-fixed camera orientation. One crucial feature of mobile video telephony is that participants may orient their handheld devices in almost any direction and at any time during the interaction, to provide the other participant with camera shots of their environments. We have shown how camera shots were mutually produced with a view to inspection for meaningfulness and relevance in the ongoing interaction (and they actually are, as we have shown in the last example). At any moment during the call, the mobile video telephony users face a specific (with respect to other forms of mediated communication) practical problem: what to show, and why, considering that what they show becomes available to the scrutiny of the recipient, and liable to be assessed for relevance. This involves particular communicative skills in which users appear at first glance as mundane directors and critics for the images they respectively provide and are shown to watch. But such competence is also interactional in its essence: producing a proper camera shot appears as a collaborative accomplishment, tightly woven into the unfolding interaction.

An important resource in that respect is the culturally reinforced standardization of some mobile video telephony interaction formats (for example in ads showing uses of mobile video telephony). We have identified and analyzed one of those, a portrait-like ‘talking heads’ format in which participants interact through headshots featuring their full face (and almost only their face). This may be an adaptation of a mode of interaction already observed in video telephony and media spaces in which participants are mutually oriented to the screen and appear to talk ‘through it’. Replacing a ‘head plus torso’ shot by a portrait-like close-up allows adjustment to the embodied constraints (keeping a flexed arm for comfort) and technical constraints (visual field as determined by fixed camera aperture) which are particular to mobile video telephony and its reliance on handheld mobile communication devices. We have shown how this interaction format is used as a default mode of interaction (when there is nothing relevant to show, and at the start of the calls where the relevance of showing something is yet to be collaboratively ratified). We have also shown how the production of a proper portrait with a full face constitutes a normative expectation on the part of the participants in the call. Whereas previous research on fixed video telephony suggested that showing the participants’ image might be less useful to collaboration than mutually sharing a significant part of the environment [6, 23], our case study shows that in mobile video telephony there might be good interactional reasons for bad collaborative resources, that is, showing people rather than things.

This has several implications for design. First, it is important to complement traditional user-centered studies of communicative and collaborative practices based on interviews, diaries and observation, with detailed analysis based on the recording and analysis of actual interactions as provided in this study. This is necessary to show how camera shots constitute a moment-by-moment joint practical accomplishment in mobile video calls. Second, it is important for designers to understand mobile video telephony on its own (rather than from a disengaged perspective in which it is one communication technology among others), and to identify the specific concerns of participants in the practical management of mobile video interactions. As we have seen, foremost among these concerns is the joint production of interaction-relevant camera shots.

This may be impaired by noisy environments, in which it is impossible to sustain a conversation between ‘talking heads’ because one has to move the communication device closer to one's mouth. Improving the position of the cell phone microphone would help with respect to that particular problem. More importantly, the flexibility of image production is tied to the camera aperture and embodied constraints such as the average distance between a user's face and hands. Allowing a choice between a few pre-determined apertures, or introducing a zooming function would greatly increase the resources for producing relevant shots.

It might also help to solve the longstanding 'presentation of self' problem, to which the failure of personal video telephony has been attributed in part, and which is also salient in mobile video calls [18]. Participants feel under pressure to present a proper face during the call (depending also on the type of social relationship that he or she has with the caller). For instance, an incoming video call at the wrong time might prove embarrassing to the recipient – much more so than a simple phone call. Previous authors have proposed to solve that problem by providing 'one-click' easy-to-use switches between audio and video mode that allow users to manage their visibility on an on-off basis [4]. If we combine our results showing the importance of providing a relevant image all the time, which in turn becomes a resource for the ongoing conversation, and analyses of the uses of various types of pictures in Web 2.0 site profiles, the possibility to change the camera's angle or to zoom away would allow participants to modulate the recognizability of their features. Likewise, rather than relying on on-off audio-video switches, providing the users who decide to block the video mode with the possibility to send a fixed image or video sequence of their choice would offer them resources for articulating personally chosen images and talk-in-interaction in new ways. This would definitely take into account the singular properties of mobile video telephony as a medium of interaction.

7. ACKNOWLEDGMENTS

Our thanks to Daniel Bessis and the Region Ile de France, which has funded the project 'Mobiles en Ville' and this particular research. We would also like to acknowledge the collaboration of Thomas Beauvisage, Frederique Legrand and other researchers at OrangeLabs.

8. REFERENCES

[1] Dourish, P., Adler, A., Bellotti, V., & Henderson, A. (1996). 'Your Place or Mine? Learning from Long-Term Use of Audio-Video Communication.' Computer Supported Cooperative Work **5**: pp. 33-62.

[2] Fish, Robert S., Kraut, Robert E., Root, Robert W. and Rice, Ronald E. (1992): 'Evaluating Video as a Technology for Informal Communication.' In: Bauersfeld, Penny, Bennett, John and Lynch, Gene (eds.) Proceedings of the ACM CHI 92 Human Factors in Computing Systems Conference June 3-7, 1992, Monterey, California. pp. 37-48.

[3] de Fornel, M. (1988). 'Contraintes systémiques et contraintes rituelles dans l'interaction visiophonique.' Réseaux **29**: pp. 35-46.

[4] de Fornel, M. (1994). 'Le cadre interactionnel de l'échange visiophonique.' Réseaux **64**: pp. 107-132.

[5] Gaver, W., 'The affordances of media spaces for collaboration', Proceedings of CSCW'92 (Toronto, 1-4 November 1992), ACM Press, pp. 17-24.

[6] Goffman, E. (1963). Behavior in Public Places. New York, the Free Press.

[7] Kendon, A. (1990). Conducting Interaction. Patterns of Behavior in Focused Encounters. Cambridge, Cambridge University Press.

[8] Kraut., R., Fish, R., Root, R., & Chalfonte, B. (1990). 'Informal communication in organizations: form, function and technology'. In Human reactions to technology: Claremont Symposium in applied social psychology. S. Oskamp, & Spacapan, S. Beverly Hills, CA., Sage Publications**:** pp.287-314.

[9] Kraut, Robert E., Fussell, Susan R. and Siegel, Jane (2003): 'Visual Information as a Conversational Resource in Collaborative Physical Tasks'. In Human-Computer Interaction, 18 (1) pp. 13-49.

[10] Lewis A., & Nightingale C., 'The Paradox of Videotelephony. Unconscious Assumptions and Undervalued Skills', BT Technical Journal, 17(1), pp. 47-58.

[11] Licoppe, C. (2004). 'Connected presence: the emergence of a new repertoire for managing social relationships in a changing communication technoscape.' Environment and Planning D : Society and Space **22**: pp. 135-156.

[12] Licoppe, C., & Smoreda, Z. (2005). 'Are social networks technologically embedded? How networks are changing today with communication technologies.' Social Networks **27**(4): pp. 317-335.

[13] Lipartito, K (2003). 'Picturephone and the Information Age: The Social Meaning of Failure', Technology and Culture - Volume 44, Number 1, pp. 50-81.

[14] Luff, P., Heath, C., Kuzuoka, H., Hindmarsh, J., Yamazaki, K., Oyama, S. (2003). 'Fractured Ecologies: Creating Environments for Collaboration.' Human Computer Interaction **18**: pp. 51-84.

[15] McHoul, A. W. (1978) 'The organization of turns at formal talk in the classroom.' Language in Society, 7, pp. 183-213.

[16] Mondada, L. (2007). 'Operating together through videoconference: Members' procedures for accomplishing a common space of action'. In: Hester, S., Francis, D. (eds). Orders of Ordinary Action. Aldershot: Ashgate, 51-67.

[17] Nardi, B., Kuchinsky, A., Whittaker, S., Leichner, R., & Schwarz, H. (1996). 'Video-as-data: Technical and Social Aspects of a Collaborative Multimedia Application.' Computer Supported Cooperative Work **4**(1): pp. 73-100.

[18] O'Hara, K., Black, A., & Lipson, M. (2006). Everyday Practices with Mobile Video Telephony. CHI 2006, Montréal, Québec, Canada, ACM Press, pp.871-880.

[19] Orlikowski, W., & Yates, J. (1994). ‘Genre repertoire: The Structuring of Communicative Practices in Organizations.’ Administrative Science Quarterly **39**(4): pp. 541-574.

[20] Schegloff, E. (1972). ‘Sequencing in Conversational Openings.’ Directions in Sociolinguistics. The Ethnography of Communication. D. Hymes, & Gumperz, J. Cambridge, Cambridge University Press**:** pp. 346-380.

[21] Sellen, A. (1995). ‘Remote Conversations: The Effects of Mediating Talk with Technology.’ Human Computer Interaction **10**(2): pp. 401-444.

[22] Weiser, M. (1991). ‘The computer for the 21st century.’ Scientific American **285**(3 (Sept.)): pp 94-104.

[23] Whittaker, S. (2003). ‘Things to talk about when talking about things.’ Human Computer Interaction **18**(1): pp. 149-170.