Zeitschrift: Bulletin suisse de linguistique appliquée / VALS-ASLA

Herausgeber: Vereinigung für Angewandte Linguistik in der Schweiz = Association suisse de linguistique appliquée

Band: - (2012)

Heft: 96: L'espace dans l'interaction sociale = Der Raum in der sozialen Interaktion = Lo spazio nell'interazione sociale = Space in social interaction

Artikel: Caméras en interaction : le travail collaboratif des monstrations visiophoniques

Autor: Morel, Julien / Licoppe, Christian

DOI: https://doi.org/10.5169/seals-978568

Nutzungsbedingungen

Die ETH-Bibliothek ist die Anbieterin der digitalisierten Zeitschriften auf E-Periodica. Sie besitzt keine Urheberrechte an den Zeitschriften und ist nicht verantwortlich für deren Inhalte. Die Rechte liegen in der Regel bei den Herausgebern beziehungsweise den externen Rechteinhabern. Das Veröffentlichen von Bildern in Print- und Online-Publikationen sowie auf Social Media-Kanälen oder Webseiten ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. Mehr erfahren

Conditions d'utilisation

L'ETH Library est le fournisseur des revues numérisées. Elle ne détient aucun droit d'auteur sur les revues et n'est pas responsable de leur contenu. En règle générale, les droits sont détenus par les éditeurs ou les détenteurs de droits externes. La reproduction d'images dans des publications imprimées ou en ligne ainsi que sur des canaux de médias sociaux ou des sites web n'est autorisée qu'avec l'accord préalable des détenteurs des droits. En savoir plus

Terms of use

The ETH Library is the provider of the digitised journals. It does not own any copyrights to the journals and is not responsible for their content. The rights usually lie with the publishers or the external rights holders. Publishing images in print and online publications, as well as on social media channels or websites, is only permitted with the prior consent of the rights holders. Find out more

Download PDF: 11.12.2025

ETH-Bibliothek Zürich, E-Periodica, https://www.e-periodica.ch

Caméras en interaction: Le travail collaboratif des monstrations visiophoniques

Julien MOREL
Télécom ParisTech
Département SES
46, rue Barrault, 75634 Paris Cedex 13, France
julien.morel@telecom-paristech.fr

Christian LICOPPE
Télécom ParisTech
Département SES
46, rue Barrault, 75634 Paris Cedex 13, France
christian.licoppe@telecom-paristech.fr

The purpose of this paper is to look at how participants in video-calls (on mobile terminals or laptops) manage to launch sequences in which one participant shows some details of his or her environment to the other. The practice of getting away from the 'talking heads'-configuration – which is oriented to as the default configuration in video communication – relies on a specific kind of interactional work validating the relevance of the video-mediated 'showing', to manage potential sense-making problems induced by transitory shots produced as the camera is being moved, and to recognise the first relevant image of the 'showing' sequence. By analogy with story-telling and extended turns-at-talk in conversation, we will show that prefaces constitute an effective resource in managing the change from the 'talking heads'-configuration to a first relevant image (related to what is supposed to be shown). Video images-in-interaction are scrutinised with respect to their relevance to the ongoing interaction, this analysis extending beyond the openings of 'showing' sequences. The person who shows and his or her recipient continuously adjust what is being shown to what is being said and vice versa. While the showing person is in charge of the camera, we will show that his or her recipient actively collaborates in the production of the shots.

Keywords:
Video communication, Skype, mobile phone, video-as-data, multimodality, sequence shot, video in interaction

1. Introduction

Cette recherche met en évidence une évolution significative de certaines pratiques en vidéocommunication sur téléphone mobile et PC. Initialement utilisés sur des ordinateurs fixes puis portables, les logiciels de vidéocommunication (Skype et MSN principalement) se répandent désormais sur de nombreux terminaux tels les téléphones mobiles, les smartphones et les tablettes. Il s'ensuit une sorte de "tournant mobile" (Urry, 2007) dans la mesure où les utilisateurs sont susceptibles de passer ou recevoir des appels vidéo en situation de mobilité, de se déplacer pendant les échanges, et d'orienter aisément la caméra pour montrer leur environnement. Les premiers systèmes de vidéocommunication étaient quant à eux relativement

© 2012 Centre de linguistique appliquée
Université de Neuchâtel

encombrants, généralement associés à un lieu dédié (domicile, travail), et les participants devaient se positionner devant une à plusieurs caméras fixes (Heath & Luff, 1992; De Fornel, 1994). L'une des premières études sur les usages de la vidéocommunication sur téléphone mobile (O'Hara, Black & Lipson, 2006) stipule que près d'un tiers des appels sont passés pour montrer quelque chose. Cette proportion est certainement sous-évaluée car elle n'inclut pas les monstrations qui émergent dans le cours des échanges, c'est-à-dire celles qui ne sont pas liées à la raison d'appel. Les occasions sont pourtant fréquentes (Morel & Licoppe, 2009), que ce soit dans nos corpus vidéo d'échanges visiophoniques enregistrés depuis des ordinateurs portables et des téléphones mobiles, et dans une moindre mesure pour les PC fixes équipés d'une webcam filaire.

Les nouvelles technologies seraient ainsi susceptibles de tenir enfin certaines promesses, en particulier concernant la capacité des interactants à distance à partager leurs environnements (Relieu, 2007). Dans des contextes professionnels en particulier, l'intérêt de la "video-as-data" a été précocement mis en relief: le fait d'enregistrer un cours d'activité et de le rendre simultanément visualisable pour les participants favorise l'émergence d'un cadre interactionnel partagé et la collaboration à distance (Nardi et al., 1996; Whittaker, 2003). Dans une telle configuration, les participants articulent vidéo et parole de sorte qu'elles soient pertinentes pour l'interaction en cours (Mondada, 2007a). Les échanges visiophoniques par Skype ou mobile présentent une telle multimodalité. En premier lieu, la réorientation de la caméra vers l'environnement est dans la grande majorité des cas préfacée par des séquences conversationnelles ratifiant l'engagement dans une période de monstration. En second lieu, les images produites par le montreur sont examinées par le destinataire, ce dernier introduisant alors des évaluations, des requêtes, contribuant à la progression et l'organisation de la prise de vues et de la conversation. Nous montrerons comment la réalisation de monstrations relève d'un processus collaboratif dans lequel les images produites et les séquences conversationnelles sont entrelacées, et plus précisément co-élaboratives. Les images produites ne sont pas simplement une ressource ou un type d'action à partir desquels les participants interagissent verbalement. Les composantes visuelles et verbales sont réflexivement liées et contribuent à la structuration et l'organisation spécifique de la "vidéo-en-interaction".

L'objectif principal de cet article est de décrire l'organisation multimodale, collaborative et systématique des séquences de monstration visiophonique de l'environnement "non humain". Pour l'essentiel, il s'agit d'objets et de portions d'espace filmés de façon unique (un seul plan fixe par exemple) ou faisant partie de séquences monstratives plus étendues et souvent complexes en terme de réalisation (alternance de plans fixes et de panoramique). Comment

les participants décident-ils de s'engager dans des séquences de monstration de l'environnement? Comment organisent-ils les prises de vues? Quel type de relation lie la "vidéo-en-interaction" et le "parler-en-interaction"? Face à quels types de problèmes pratiques se trouvent-ils lorsqu'ils souhaitent isoler une composante particulière de l'environnement ou pour le montrer dans son ensemble, tout en disposant du champ limité d'une caméra?

Pour répondre à ces questions, nous exploiterons des corpus de communications vidéo enregistrées depuis des ordinateurs fixes et portables (via Skype), et des téléphones mobiles.[1] Dans un premier temps, nous présenterons les traits généraux de l'organisation multimodale et séquentielle des communications vidéo, tout en rappelant que la configuration "têtes parlantes" est dominante et qu'elle se présente comme le format "par défaut" des communications vidéo. Lorsque deux participants sont engagés dans une communication visiophonique, ils tendent en effet à positionner leur visage au centre de l'écran, co-produisant une "écologie œil à œil" (Goffman, 1963) se rapprochant du face à face. Dans un second temps, nous montrerons comment l'engagement dans une séquence monstrative est méthodiquement et collaborativement organisé, en particulier au moyen de préfaces initiées par le montreur ou de demandes formulées par le destinataire. Dans un troisième temps enfin, nous examinerons la façon dont les participants structurent les séquences monstratives, en montrant la façon dont ils coopèrent pour la structuration de ce qui s'apparente à un plan-séquence. Si le montreur maitrise pratiquement les prises de vues, le destinataire contribue activement à leur réalisation.

2. L'organisation multimodale et séquentielle des communications vidéo sur Skype et téléphone mobile

Les monstrations de l'environnement sont rendues possibles par la mobilité des caméras (webcam manipulable) ou celle des terminaux (ordinateur portable, téléphone mobile). Ceci autorise les participants à montrer leur environnement lorsqu'ils jugent qu'une telle action est pertinente. Il y toutefois une organisation distincte quant à la production de telles prises de vues relativement à ces cinq caractéristiques des appels vidéo:

- Dans les ouvertures et les clôtures des communications, les participants se positionnent de façon quasi systématique en configuration "têtes parlantes", en cherchant à maximiser la visualisation de l'entièreté de leur visage.

---

1 Au total, une vingtaine de personnes ont constitué ces corpus (150 communications au total), au moyen de logiciels de capture vidéo d'écran ou grâce à la sortie vidéo dont sont pourvus certains téléphones mobiles (voir Morel & Licoppe, 2011 pour une présentation du système de captation vidéo).

- Que ce soit sur Skype ou téléphone mobile, les participants co-produisent principalement cette formation interactionnelle (Kendon, 1990) tout en s'engageant régulièrement dans des périodes plus ou moins étendues de monstration satisfaisant des besoins variés de visualisation distante et partagée de l'environnement.
- La configuration "têtes parlantes" se présente comme le format "par défaut" des échanges visiophoniques sur mobile et Skype, et s'il n'y a rien de pertinent à montrer, cette configuration est attendue.
- Les images vidéo ne cessent d'être produites et inspectées en vertu de leur pertinence pour l'interaction en cours, avec une attention particulière pour les images de transition dans la mesure où le passage de la configuration "têtes parlantes" à la production d'une première image de monstration (un plan fixe montrant un objet par exemple) implique une réorientation de la caméra générant des images plus ou moins intelligibles.
- Dans les communications multipartites, les participants s'orientent vers le fait que celui qui filme doit mettre à l'écran le locuteur en cours ou celui identifié comme étant le prochain. Les participants organisent le "cadre interactionnel visiophonique" (De Fornel, 1994) pour générer des cadres de participation appropriés aux activités en cours. Le travail de production d'une image pertinente est lisible (et lu) comme une interprétation de l'organisation de la conversation en cours.

Nous avons montré que ces caractéristiques n'étaient pas des composantes isolées les unes des autres, mais qu'elles participaient à l'élaboration d'ensemble de la vidéo-en-interaction (Licoppe & Morel, à paraître). Elles peuvent être dérivées d'une orientation vers une maxime unique et fondamentale, *mets le visage du locuteur actuel à l'écran*. Ceci explique pourquoi l'arrangement interactionnel en têtes parlantes est traité comme le format "par défaut": s'il n'y a pas de raison de montrer quelque chose de particulier, les participants s'orientent vers cette maxime en positionnant leur visage au centre de l'écran. Cela explique aussi pourquoi les ouvertures se font presque toujours dans ce format: avant qu'une raison d'appel ait été introduite, il n'y a pas lieu de montrer quelque chose en particulier, et la configuration "têtes parlantes" est attendue par les participants. Une autre conséquence est que les participants sont continûment responsables des images qu'ils produisent, et pas seulement dans les ouvertures. Quand l'un d'entre eux décide de montrer autre chose que son visage, les nouvelles images sont analysées par le destinataire. Se posent alors des problèmes d'intelligibilité, quant à la nature des images produites (sont-elles compréhensibles?) et la façon dont elles contribuent à l'interaction en cours (sont-elles pertinentes?).

3. Commencer à montrer l'environnement: l'organisation séquentielle des ouvertures des monstrations visiophoniques

L'engagement dans une séquence de monstration est dans la plupart des cas préfacé par des séquences conversationnelles. L'abandon du format "têtes parlantes" implique un travail interactionnel spécifique permettant de ratifier une orientation vers une visualisation partagée de l'environnement, puis l'engagement dans un cours d'action spécifique. Les images monstratives constitueront en effet une nouvelle ressource à partir de laquelle les participants structureront et orienteront collaborativement la communication. Au préalable, ils ont à faire face à deux types de problèmes pratiques. Comment introduire puis ratifier l'engagement conjoint dans ce type de séquence? Comment déterminer quelle image débute la période, tout en limitant les troubles potentiels liés à la réorientation de la caméra qui génère des images peu intelligibles? Dans cette partie, nous montrerons que l'introduction de séquences de préface constitue une méthode routinière et efficace pour l'engagement dans une séquence monstrative, initiée par le montreur ou plus rarement, le destinataire.

3.1 L'organisation séquentielle des monstrations initiées par celui qui montre: le travail de préfaçage

L'engagement dans une séquence de monstration de l'environnement est méthodiquement et collaborativement organisé, et il exhibe des caractéristiques séquentielles reconnaissables. Lorsqu'un participant est sur le point de montrer quelque chose, il fait face à un problème pratique analogue à celui d'un conversant souhaitant s'engager dans une narration. En suivant l'analyse séminale de Sacks, l'engagement dans une narration implique une suspension de l'organisation usuelle de prise de tour, et du type d'obligation d'écouter qu'elle intègre. Le narrateur gère généralement cela via l'introduction de séquences de préface (Sacks, 1992; Jefferson, 1978) qui agissent de manière méta-pragmatique en orientant l'interaction en cours vers un format conversationnel particulier. La préface offre également au destinataire potentiel une place pour ratifier l'engagement vers un tel développement. La conception de la préface constitue de plus une ressource pour le destinataire, pour reconnaître quand l'histoire est arrivée à un terme possible. Un problème séquentiel similaire se pose à un conversant vidéo sur le point de montrer quelque chose, avec un problème pratique additionnel quant à la production de la première image monstrative. D'une part, la permutation de la caméra (dans le cas des téléphones mobiles) et/ou sa réorientation, génèrent en effet des images de transition (de quelques dixièmes à une dizaine de secondes) pouvant affecter l'attente d'intelligibilité des images ou constituer une source de trouble (en particulier lorsque leur durée est importante). D'autre part, il s'agira ensuite de déterminer quelle image débute la séquence monstrative,

tant en vertu du cadrage que de la capacité du destinataire à l'identifier comme étant la première image de la séquence.

L'introduction d'une séquence de préface permet d'annoncer que le participant va montrer quelque chose, et de lever l'ambiguïté possible des images transitionnelles. Comme pour les récits, cela offre une place au destinataire pour valider un tel mouvement, et la formulation lexicale de la préface constitue également un moyen pour 1) fournir au destinataire des ressources interprétatives pour pouvoir déterminer quelle est la première image pertinente (par rapport au projet de monstration), dont la production marque le moment où l'attente d'intelligibilité visuelle devrait être rétablie, et 2) annoncer le type de prises de vues qui sera réalisé.

L'extrait suivant, tiré du corpus de visiophonie sur téléphone mobile, nous permet d'identifier une séquence à partir de laquelle un montreur potentiel initie une orientation vers la réalisation de prises de vues montrant l'environnement de l'appelant (A, image témoin).

Transcription 1: j'te fais visiter l'appart ?

```
1.    A  non (.) t'as eu raison hein
```

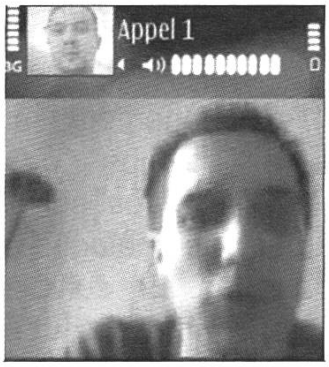


Figure 1: période "têtes parlantes"

```
2.         0.45
3.    A  bon bi(e)n c'est bien
4.         0.96
5.    A  bon
6.         (.)
7.    A  j'te fais visiter l'appart
8.         0.93
9.    B  comment ?
10.        0.06
11.   A  j'te fais visiter l'appart
12.        0.68
13.   B  ouais: vas y (.) fais moi visiter.
```

En configuration têtes parlantes, les participants ont ratifié (l.11-13) un changement d'orientation de la caméra d'A, et toute image reconnaissable du mobilier ou d'une pièce de l'appartement par exemple, peut constituer un "montrable" correspondant à une telle orientation. Ceci offre également au

destinataire des ressources interprétatives pour déterminer quand quelque chose va être montré, et quand débutera la visite. Cependant, le passage des têtes parlantes à une première image de monstration n'est pas immédiat, et les participants ont à gérer différentes étapes de transition.

```
14.        0.04
15.    A   allez go
```

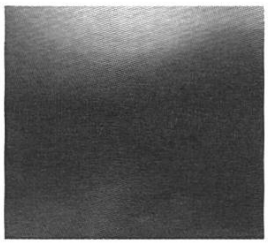


Figure 2: l'image fixe de transition lors de la permutation de caméra

```
16.        0.28
17.    A   att[ends
18.    B     [(inc)
19.        0.22
20.    A   attends
21.        1.80
22.    A   ho::p
23.        1.03
24.    B   (rasé)  *aujourd'hui
25.        0.22
```

Figure 3: première prise de vues environnement (plan fixe avec légers mouvements de caméra)

```
26.    A   alors tu vois quelqu'chose là (.) c'est bon ?
27.        0.90
28.    B   ouais (.) c'est bon (0.1) mais t'es pas rasé là
29.        0.15
```

L'activation de la caméra secondaire du mobile (, l.15) fait apparaître une image de fond d'écran (Figure 2), et s'il ne s'agit pas d'une image monstrative, A introduit néanmoins dès son apparition l'orientation vers le projet de visite ("allez go"). Cependant, A introduit ensuite plusieurs "attends" (l. 17 et 20) pouvant projeter la réalisation prochaine d'une première image pertinente (l'image fond d'écran reste affichée plusieurs secondes, jusqu'à la ligne 25), et viser également à bloquer des séquences conversationnelles initiées par B (l. 18, l. 24, l. 28) portant sur l'apparence de son interlocuteur. La demande de vérification de l'image (l. 26) qui fait suite à l'apparition d'une première prise de vues "environnement" (au niveau de l'étoile de la ligne 23; Figure 3), révèle

une étape préparatoire additionnelle typique des échanges visiophoniques (voir par exemple, De Fornel, 1988, 1994; Mondada, 2007b). Dans le cas présent, la formulation de la demande de vérification ("tu vois quelque chose là?") projette une réponse "oui/non" quant à la réception de l'image, et si pendant cette séquence une portion de l'environnement d'A est visible, il ne s'agit pas encore de la première image du projet de visite. La seconde partie de paire de la demande de vérification (première partie de l. 28) confirme la réception correcte de l'image, et devrait permettre la réalisation de la première image monstrative. Cependant, B poursuit son tour de parole pour délivrer une évaluation (cette fois-ci totalement compréhensible) sur l'apparence d'A, et d'autres séquences conversationnelles peuvent différer l'engagement dans le projet de monstration.

```
30.   A   h.h. non:: bah tu sais moi l'week end euh:::*
31.       j'm'en fous (0.2) j'm'en fous mon rasoir i
32.       marche que la s'maine (.) i marche pas *le week
33.       end
```

Figure 4: décadrage, plan fixe mur

```
34.       (.)
35.   B   par contre ça coupe vachement::
```

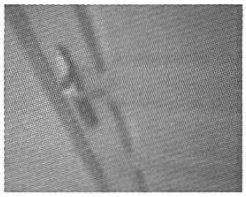


Figure 5: plan fixe poignée (retour)

```
36.       0.25
37.   A   hein ?*
38.       *0.81
39.   B   ça n'arrête pas de couper
```

Figure 6: plan fixe mur/sol

```
40.       0.06
41.   A   ouais ouais (.) ça::: apparemment ça::::: déconne
42.       un peu là (.) (mais bon)* on va essayer d'faire
```

43. *euh:: avec*
44. *0.38*
45. A *donc ça c'est l'entrée::

Figure 7: la première image de la séquence de monstration

46. 0.790
47. B la porte d'entrée (.) blindée:: ?

C'est le cas avec les justifications apportées par A sur son apparence (l. 30-33), puis avec les séquences d'évaluation de la qualité du signal audio (l. 35-43), dont la faiblesse génère d'ailleurs une séquence de réparation (l. 37-39). Conjointement, on observe un relâchement de la contrainte d'intelligibilité des images. Après le plan fixe initial offrant une vue partielle d'une porte (Figure 3), la caméra semble "errer" dans l'environnement, montrant successivement un mur, une poignée de porte ou encore le sol (Figure 4, Figure 6). Compte tenu des thèmes de conversation sur l'apparence ou les conditions de communication, il n'y a, d'une certaine façon, pas de cadrage "correct" ou "approprié" de l'environnement à produire, et un retour aux têtes parlantes serait même particulièrement bien adapté au premier type de développement. Que ce soit dans sa dimension conversationnelle ou filmique, il n'y a plus à ce stade de progression du projet de monstration. Elle reprend avec la dernière unité de construction de tour (Sacks, Schegloff & Jefferson, 1974; dorénavant UCT) des lignes 42-43 ("on va essayer d'faire euh:: avec"), qui réalise un mouvement subtil vers un début possible de la séquence monstrative: a) l'UCT est susceptible de mettre un terme aux séquences d'évaluation, b) le destinataire est associé ("on") à un cours prochain d'activité, et c) un mouvement de caméra est initié pendant l'UCT (période italique). C'est à partir de ces ressources multimodales qu'une première image monstrative est produite juste après ce tour (Figure 7).

Après un bref silence (l. 44), le cadrage est stabilisé sur la porte d'entrée (Figure 7, qui correspond également à l'endroit d'où débute une visite en situation de coprésence) et c'est à ce moment précis qu'A reprend la parole pour nommer l'élément montré. L'association plan fixe/désignation constitue un procédé courant maximisant l'intelligibilité des images produites, et le silence qui suit offre ensuite au destinataire une place pour introduire des évaluations, des requêtes, etc. Dans le cas présent (l. 47), il rend compte de son identification de la porte d'entrée, puis formule une question sur l'un de ses attributs possible ("blindée?"). Ainsi, la structuration des prises de vues et celle du "parler-en-interaction" sont co-élaboratives, et de sorte qu'une image particulière puisse être identifiée comme la première d'une séquence de

monstration de type "visite". Le destinataire contribue à déterminer quelle image débute la période monstrative, et il peut ensuite initier d'autres séquences conversationnelles projetant le maintien, l'ajustement, l'abandon, etc. de la prise de vues en cours.

L'extrait présenté révèle le type de travail interactionnel nécessaire pour commencer à montrer quelque chose dans un appel vidéo: la production de séquences de préface, dont la première partie offre une ressource pour déterminer une première image pertinente en relation avec le type de monstration qu'elle projette. La préface limite également les problèmes interprétatifs des images de transition entre le positionnement "têtes parlantes" et la première prise de vues de l'environnement. La suspension de l'attente normative concernant l'intelligibilité et la pertinence de la vidéo-en-interaction est particulièrement saillante lors de l'introduction de demandes du type "attends", pouvant conjointement viser à bloquer des développements thématiques. Un participant sur le point de montrer son environnement se trouve ainsi face à un problème pratique proche de celui d'un conversant souhaitant s'engager dans un récit, auquel s'ajoute celui de l'intelligibilité des images. Le travail vidéo requis ici pour abandonner le format "par défaut" en "têtes parlantes" et l'arrivée à une première image de monstration suppose une suspension du voir incorporé dans l'organisation séquentielle de la "vidéo-en-interaction". Les séquences de préface constituent un moyen efficace et routinier pour une telle orientation, qui reste vulnérable aux aléas de connexion.

3.2 Les monstrations initiées par le destinataire

Dans la majorité des cas, c'est le montreur qui initie l'engagement vers une période de monstration à partir de différents types de préface (proposition, annonce, etc.). Examinons à présent le cas où c'est le destinataire qui introduit le projet. Quels types de ressources et procédures utilise-t-il pour obtenir une vue de l'environnement? Comment le montreur produira-t-il des images correspondant au projet initié par le destinataire? L'extrait étudié est issu du corpus de communications visiophoniques sur téléphone mobile. La jeune femme de l'image témoin est l'appelante (A), et l'enfant (B) est son neveu. Ce type de lien, l'âge respectif des participants et leurs savoirs communs ne sont pas neutres quant aux façons dont débute le projet.

Transcription 2: fais voir comment t'as rangé

```
1.    A   d'accord. (.) mais t'es où là (0.2) t'es chez
2.        maman ?
```

Figure 8: la configuration initiale en "têtes parlantes"

```
3.          (0.3)
4.     B    ah ouais:
5.          (1.0)
6.     A    t'es dans ta chambre ?
7.          (0.7)
8.     A    (ta chambre ?) ((écho))
9.          (0.1)
10.    B    ouais !
11.         (0.4)
12.    A    fais voir comment t'as rangé:(.) mais allume
13.         les lumières pa(r)ce que j'vois rien
14.         (1.8)
15.    A    t'as rangé ou pas ?
16.         (1.9)
17.    B    bah:: ◉ouais::: un◉ peu:
18.    (1.1)
```

Figure 9: premier plan fixe (sol)

```
19.    A    ah  ouais d'accord (.) c'est ta ch:/ uh ! hhhh
```

La transcription commence au moment où l'appelante (A, l. 1) délivre une demande de localisation rapidement suivie d'une question intégrant une réponse candidate, le tout après un premier développement thématique. Une fois cette localisation confirmée (l. 4), A reprend la parole et poursuit son investigation sur la localisation plus précise de l'enfant, à l'échelle d'une pièce. L'inférence est confirmée par B (l. 10) et A s'auto-sélectionne de nouveau au tour suivant pour introduire une demande de monstration de la pièce dans laquelle se trouve justement B, et plus précisément pour vérifier si la chambre est rangée. Son tour se termine par une demande d'amélioration des conditions d'éclairage, le tout projetant la réalisation prochaine d'une première image de monstration. Cependant, un silence de près de deux secondes suit,

et le comportement non verbal de l'enfant ne laisse pas présager une orientation vers la production d'une telle image (manipulation du mobile, mouvement de caméra par exemple). C'est pourquoi A reformule sa question (de type oui/non; l. 15) concernant l'état de rangement de la chambre, le fait de ne pas avoir rangé pouvant constituer une raison de ne pas montrer l'environnement. La réponse apportée par l'enfant (l. 17) est hésitante, positive puis atténuée, et une brève inspection visuelle de la chambre est réalisée pendant ce tour (👁période👁). C'est pendant le silence de la ligne 18 qu'un mouvement de caméra est initié et aboutit rapidement à la réalisation d'un premier plan fixe permettant de visualiser le sol. La technique du pivotement du mobile réduit considérablement la durée des images transitionnelles, et la première image monstrative s'avère tout à fait appropriée au projet d'inspection: les jouets éparpillés sur le sol suscitent une évaluation ironique et amusée du destinataire (l. 19), qui rend ainsi conjointement compte du caractère intelligible et pertinent de la prise de vues.

Cet extrait nous permet d'identifier un autre type de travail préparatoire d'un projet de monstration initié par la destinataire. Il n'est pas possible de déterminer si ce projet était lié à une orientation précédant l'appel ou si sa pertinence émerge dans le cours de la communication. Quoiqu'il en soit, l'appelante dispose de savoirs quant à l'entretien de la chambre de l'enfant, et le projet nait de séquences de localisation dont l'issue est originale. De façon progressive (t'es où? > maison? > chambre?), l'appelante vérifie des pré-conditions de localisation, qui rendent compatible l'introduction d'une demande de monstration. Si la première image monstrative n'est pas produite immédiatement après la demande, elle est réalisée beaucoup plus rapidement que dans le premier extrait analysé. D'une part, le pivotement du mobile réduit à quelques dixièmes de seconde la durée des images de transition, et d'autre part, l'évaluation qui est faite du premier plan fixe rend compte de son intelligibilité et de sa pertinence pour le projet de monstration. Que la séquence monstrative soit auto- ou hétéro-initiée, on observe un travail de préfaçage qui a) ratifie une orientation commune vers un projet de monstration, b) limite le trouble potentiel des images transitionnelles (même de durée brève) et c) prépare à un certain type de prise de vues. Par contre, dans le cas où c'est le destinataire qui nourrit un projet de monstration, il peut être amené à vérifier si la localisation du montreur est compatible. Ce type de problème pratique ne se pose pas dans la configuration inverse.

Lorsqu'une première image monstrative a été produite, puis reconnue et traitée comme telle par le destinataire, les participants ont ensuite à organiser la structuration conjointe des prises de vues et du parler-en-interaction. Notons que cette relation multimodale et co-élaborative ne caractérise pas systématiquement l'entièreté d'une séquence monstrative: les participants

s'engagent parfois dans des développements thématiques sans lien avec les images préalablement et actuellement produites.[2]

4. Le travail collaboratif de la caméra et de sélection des composantes pertinentes de l'environnement

Mais intéressons-nous à présent au cœur des séquences monstratives, et aux façons dont les participants structurent les prises de vues en vertu de différentes séquences conversationnelles. Ils se trouvent d'une certaine façon confrontés à la production d'un plan-séquence susceptible d'exploiter différentes prises de vues (plans fixes, panoramiques, vue large/rapprochée, etc.), en temps réel, en une seule prise, sans aménagement préalable de l'environnement, et sans que les "dialogues" soient préparés à l'avance. Comment montreur et destinataire organisent-ils cette vidéo-en-interaction? Comment définissent-ils les composantes pertinentes du champ de la caméra? Peut-on identifier des formats séquentiels reconnaissables contribuant à l'organisation du plan-séquence? Les catégories "montreur" et "regardeur" sont-elles appropriées pour caractériser les contributions respectives des participants? Pour répondre à ces questions nous exploiterons trois extraits d'enregistrements de communications visiophoniques montrant ce travail collaboratif et multimodal de la vidéo-en-interaction orientée vers la monstration de l'environnement. Nous nous attacherons en particulier à identifier des modalités de collaboration dont certaines révèlent des appariements catégoriels occasionnés au cours des échanges (Sacks, 1992; Watson, 1997), comme montreur-spectateur voire réalisateur-spectateur.

4.1 Inspecter un espace

La suite de l'échange visiophonique entre la jeune femme et son neveu permet d'identifier un positionnement particulier du destinataire (A) quant aux images produites, et lié à la façon dont le projet a débuté: la monstration actuelle est une "vision instruite" (Mondada, 2003) dans le but d'inspecter un espace déterminé. Dès lors, le projet implique que le montreur produise une forme de documentaire à partir duquel le spectateur doit pouvoir fonder son jugement.

Transcription 3: fais voir comment t'as rangé (suite)

```
19.   A   ah  ouais d'accord (.) c'est ta ch:/ uh ! hhhh
20.       pourquoi tout est par terre comme ça:
21.       (1.6)
```

---

2 Généralement avec pour conséquence un maintien du dernier plan fixe produit ou un relâchement de la contrainte d'intelligibilité du cadrage (prise de vues "flottante" évoquée dans le premier extrait vidéo étudié).

22. B pace que j'ai joué aya: *ya:::::: hier.*

Figure 10: deuxième plan fixe bureau

23. (0.8)

Figure 11 : retour au plan fixe initial sol

Ce jugement est rendu explicite dès l'évaluation (l. 19) du premier plan fixe (sol), et il est suivi d'une question projetant une justification du désordre régnant dans la chambre. Pendant ce tour, on observe l'initiation d'un panoramique horizontal gauche (dorénavant PHG, italiques) se terminant par un second plan fixe (bureau; Figure 10), stabilisé pendant la délivrance de la seconde partie de paire (l.22).

24. A *d'accord (.) okay. (.) t'as chang(è) la/ (.) t'as*
25. *rangé la chambre* de ta sœur ?

Figure 12: plan fixe meuble

26. *(1.4)*
27. B *ouais: mais j'crois qu'elle l'a dérangée (.)*
28. *j'ai pas vu:=*
29. A *=ah bon ?=*
30. B = attends (0.4) j'vais t'dire ça tout d'suite

Figure 13: le retour en têtes parlantes

Un nouveau panoramique est initié pendant la délivrance de ce tour, et se termine par un retour au plan fixe initial. Un panoramique horizontal droite (PHD) suit (l. 24), au moment où la spectatrice s'engage dans un tour de parole où elle manifeste dans un premier temps l'épuisement de la pertinence de la monstration en cours ("d'accord (.) okay."; l. 22) et dans un second, elle introduit une requête visant à déterminer si l'enfant a "également" rangé la chambre de sa sœur. C'est à la fin de cette requête qu'un nouveau plan fixe est produit (Figure 12), suivie d'un PHG (l. 27) accompagnant la délivrance de la seconde partie de paire (l. 27-28) et dont la dernière image est de nouveau celle du plan fixe initial. Le premier symbole 📱 marque le début d'un mouvement de caméra aboutissant à un retour en position tête parlante (second 📱).

Cette courte séquence permet d'identifier différentes modalités de participation du montreur et du spectateur pour la conception du plan-séquence. Son organisation d'ensemble (alternance de plans fixes et de panoramiques) semble reposer exclusivement sur les choix du montreur. La structure de ses prises de vues peut être schématiquement présentée ainsi (PF pour plan fixe):

👽 =>[PF1 => PHG => [PF2 => PHD => [PF1 => PHD => [PF3 => PHG => [PF1 => 👽
[sol [bureau [sol [meuble [sol

Le plan fixe initial fonctionne comme une image clé, à partir de laquelle a) peut débuter la séquence monstrative si le destinataire la traite effectivement comme pertinente, et b) s'articulent les autres prises de vues (plans fixes, panoramiques). Ce procédé associé aux panoramiques sont des méthodes courantes et efficaces pour restituer un environnement dans son ensemble, ainsi que pour mieux localiser les portions d'espace révélées par le champ de la caméra. Cependant, le documentaire correspond à un projet d'inspection initié par la destinataire, dont certaines des contributions verbales sont dès lors susceptibles d'orienter le plan-séquence: c'est potentiellement le cas avec les premières évaluations, qui peuvent rendre pertinent le changement de prise de vues, et la destinataire contribue à mettre un terme à la séquence

monstrative. Dans le cas présent, le montreur est libre de produire les images correspondant au projet de monstration, et cet exercice est ici réussi.

4.2 Faire un tour de l'environnement

Examinons à présent un autre style de travail collaboratif et des séquences conversationnelles distinctes à partir desquels se structure la visite d'une autre chambre. L'appel a été passé et reçu sur des ordinateurs portables; l'appelante est la cousine de l'appelé, et ce dernier vient d'emménager à l'étranger.

Figure 14: la configuration initiale en têtes parlantes

Transcription 4: fais voir ta chambre déjà

```
1.      B :  donc (.) faudra que j'te montre à côté: mais là
2.           chuis chuis: (.) j'ai un peu la flemme (inc)
3.           chuis pieds nus (.) fait froid au sol=
4.      A :  =fais voir ta chambre déjà
5.           (0.7)
6.      B :  ra:::::: (inc)
7.           [          ]
8.      A :  fais juste tourner ✋1 ton ordi
9.           (0.5)
10.     B :  (inc) (voir ✋2 pas) grand chose
11.          (0.5)
12.     B :  ssss *ça c'est:: euh la:: c'est donc le mur de:
13.          not chambre
```

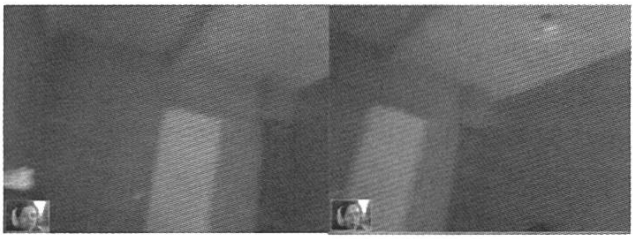
Figure 15 : premier pln fixe mur, *l.12-15*

```
14.          1.34
15.     A : d'accord: elle* est bleue
```

Après plusieurs développements thématiques (non transcrits) en configuration "têtes parlantes", B introduit un projet de monstration de son nouvel appartement (l. 1-3). Présenté comme une forme d'obligation, il apporte deux types de justification pour en reporter la réalisation, le segment "faudra que j'te

montre à côté:" projetant certainement la monstration de la pièce principale de l'appartement. Rapidement (=; l. 4), A prend la parole et réajuste le projet en introduisant une demande de monstration se limitant à la pièce dans laquelle se trouve B. La demande et l'item "déjà" finissant le tour réalisent une forme de compromis entre le projet auto-avorté du montreur et sa capacité à le débuter néanmoins ici et maintenant. Le "cri réponse" (Goffman, 1978) qui suit rend compte d'une orientation plutôt négative du potentiel montreur vers ce nouveau projet, et c'est via une nouvelle demande (l. 8) suggérant une méthode "économique" de prise de vues que le destinataire parvient à obtenir l'initiation de la séquence monstrative. Trois éléments rendent compte de cette progression: a) l'initiation d'un mouvement de main (✋1) jusqu'à la saisie de l'ordinateur (✋2), b) un mouvement de caméra (période italique; l. 10-12), accompagné d'un tour de parole introduisant une réserve non spécifiée quant à la pertinence de la monstration à venir.[3] L'image qui débute la séquence est un plan fixe de quelques secondes montrant une partie de la chambre (Figure 15). De façon notable, on observe que dès la première image (*; l.12) d'un nouveau plan, le montreur s'engage dans un tour de parole pour le qualifier et renforcer ainsi l'intelligibilité de la monstration. Cette dernière est interactionnellement réalisée lors de l'introduction d'une évaluation de la part du destinataire (l. 15), et un certain type de réception ("d'accord:") rendant possible la progression du plan-séquence. C'est en effet pendant ce tour, que débute un PHD (*l.15*) qui va cependant poser problème.

16. (0.3)



Figures 16-17: les premières images du PHD

17. A : j'vois:: donc la lampe maintenant (.)

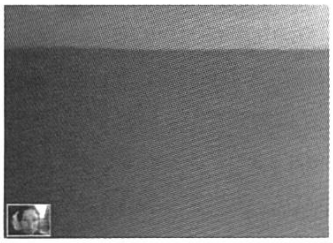


Figure 18: mauvais cadrage (plafond)

18. >tu m'fais voir l'plafond< mic

---

3 On ne peut déterminer en effet si B anticipe sur des problèmes de visualisation de la pièce ou s'il considère par exemple qu'il y a peu à voir.

19. 0.*61

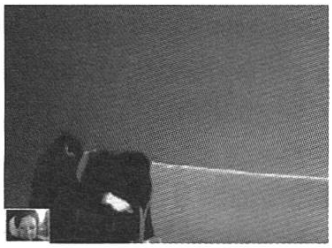


Figure 19: cadrage lit, bord gauche

20. A : baisse baisse
21. (1.*05)
22. B : *l*à: ?

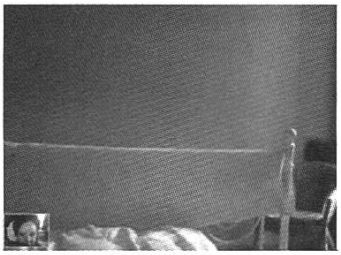


Figure 20: cadrage lit, bord droit

23. A : (0.4) ça c'est l'bout d'ton lit* *j'imagine*
24. *(0.2)*
25. *B : okay*

La mise en relief du caractère non pertinent du PHD débute certainement lorsque le destinataire indique voir "la lampe maintenant" (l. 17). L'intonation suggère une légère ironie, et compte tenu des caractéristiques banales de la lampe montrée, elle est certainement jugée non digne de mention visiophonique. Il est intéressant de noter comment ce tour prend en compte la technique de réalisation du film: le segment "je vois" allongé et le marqueur temporel "maintenant" sont liés à l'effet de révélation produit par le panoramique. Cette technique de prise de vues a aussi pour effet de générer des évaluations-réceptions a posteriori du destinataire relativement aux images produites: lorsque B met en relief sa visualisation de la lampe, le cadrage actuel donne à voir une portion de mur et le plafond (Figure 18); lorsqu'elle stipule voir le "plafond",[4] c'est lors d'un mouvement transitionnel de caméra qui préface un retour prochain à un cadrage "correct" (l. 18); et lorsqu'elle introduit la nouvelle instruction "baisse baisse" (l. 20), c'est après le rétablissement d'un cadrage pertinent (plan fixe d'une partie du lit; l. *19-21*; Figure 19).

---

4 La mention d'un cadrage montrant le "plafond" et le terme d'adresse terminant le tour manifestent cette fois-ci officiellement un problème de cadrage et de pertinence des images dont le montreur est en pratique responsable.

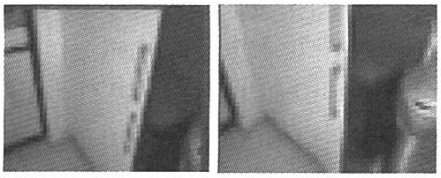
Le destinataire joue ainsi un rôle particulièrement actif pour la réalisation du plan-séquence, en qualifiant constamment les éléments montrés et de sorte qu'ils soient rendus pertinents ou non pertinents pour l'interaction en cours. Selon le même principe, les tours de parole évaluatifs qui font suite aux plans fixes (l. 15, l. 23-25) et le fait qu'ils incorporent des indicateurs tels que "d'accord" et "okay", marquent l'épuisement de la pertinence des images produites, et c'est pendant la délivrance de ce type de tour ou juste après que des changements de cadrage sont réalisés. Ces contributions du destinataire constituent une ressource à partir de laquelle le montreur conçoit le plan-séquence. Si les préfaces des séquences monstratives peuvent projeter certaines manières de tourner ("faire un tour", "faire visiter", "montrer une pièce"), elles ne déterminent pas les méthodes collaboratives mises ensuite en œuvre par le montreur et le destinataire pour la production d'images intelligibles et pertinentes de la séquence monstrative. Dans le cas présent, cette dernière aboutira à un panoramique de 360°, entrecoupé de plans fixes, de mouvements transitionnels de caméra, cet accomplissement étant fragilisé par de mauvais cadrages chroniques.

4.3 Visiter l'environnement

Lorsque les périodes monstratives sont étendues, le destinataire alterne généralement entre les deux modes de collaboration étudiés supra, tantôt spectateur, tantôt co-réalisateur. Si le montreur tend à produire les prises de vues jugées appropriées au projet ratifié de monstration, par ses évaluations, ses requêtes ou ses instructions, le destinataire contribue étroitement à l'organisation des prises de vues. Cette alternance des formes de collaboration caractérise en réalité de nombreuses séquences monstratives, et mérite à ce titre une dernière analyse d'une communication visiophonique passée et reçue sur téléphone mobile. Le fragment étudié fait suite à l'initiation du projet de visite de l'appartement examiné dans la seconde partie. Après l'inspection de l'entrée, le montreur se rend dans la cuisine et débute un inventaire filmique de l'ensemble des composantes de la cuisine, au moyen de plans fixes successifs révélant l'électroménager, les meubles, le plan de travail, etc. Ce type de projet tend à positionner le montreur en tant que réalisateur, choisissant les prises de vues pertinentes pour montrer l'ensemble des éléments de sa cuisine. Cette caractéristique est saillante au début de la transcription, mais le destinataire peut intervenir dans le cours de cette réalisation, et en modifier le déroulement pour a) orienter l'interaction vers certains éléments isolés des images et/ou b) favoriser leur conversion en "objets de la conversation" (Relieu, 2002).

Transcription 5: j'te fais visiter l'appart (suite)

1. A : *donc là tu vois (.) t'as congel en bas:::



Figures 21-22: plan fixe congélateur

2. 1.1*8
3. A : le frigo [là:::]
4. B : [elle est bien vot poub/ euh/
5. *0.09
6. B : et ouais::

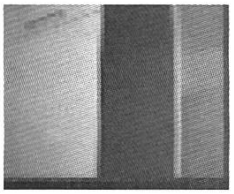


Figure 23: plan fixe réfrigérateur

7. 0.71
8. A : au d(e)ssus t'as un placard pour euh* du:.
9. bordel *(0.2) là*

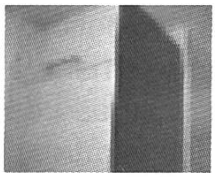
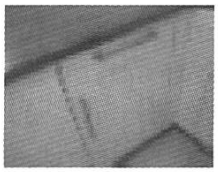


Figure 24: plan fixe meuble

10. A : euh::: là encore des placards:: ici *(0.3) hop



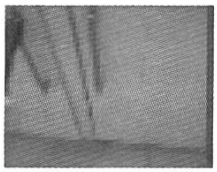


Figure 25-26: plans fixes meubles

11. (0.5) (inc) la machine à laver:::: ?

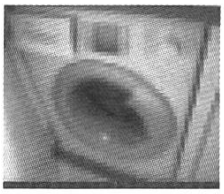


Figure 27-28: plan fixe machine à laver

12. B : ah:::: c'est génial comment c'est caché: (.)
13. alors on a la même alors vraiment ou pas=

14. A : =ah >ouais ouais< c'est la même hein (.) >c'est
15. exactement< la même qu'on a
16. 1.4*4

Des lignes 1 à 9, on observe que le montreur tend à contrôler entièrement les prises de vues, en faisant se succéder des plans fixes de chaque élément de la cuisine, et en les désignant juste avant ou pendant que le cadrage est stabilisé. L'inventaire visuel est réalisé de façon assez systématique, le montreur filmant de bas en haut, de droite à gauche. Le destinataire n'intervient pas à chaque plan fixe, mais un détail retient son attention (l. 4). Une poubelle devient partiellement visible lors du plan fixe "congélateur", et elle suscite une évaluation interrompue mais comprise par le montreur (l. 6). Cette séquence ne modifie cependant pas l'organisation méthodique des prises de vues, le montreur continuant de faire se succéder des plans fixes en nommant à chaque fois les cibles au moment où celles-ci deviennent visibles (l. 10-11 par exemple). L'ouverture d'un placard dans lequel se trouve la machine à laver (Figure 27) fait l'objet d'une évaluation du destinataire puis d'une question (l. 13) visant à vérifier si les participants disposent du même modèle. Le montreur confirme et maintient le plan fixe jusqu'à ce que cette thématisation soit épuisée (au niveau du silence de la ligne 16). Un mouvement de caméra débute alors, et aboutit à un nouveau plan fixe "plan de travail" (Figure 29).

17. A : *voil*à:=
18. B : =elle est bien vot poubelle (.) elle elle/ elle

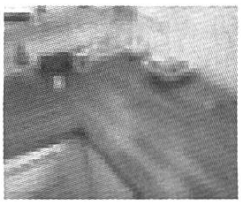


Figure 29: plan fixe plan de travail

19. (.) j'ai vu une poubelle* *en métal aussi*
20. *(.) elle a l'air d'être bien:*
21. *(.)*
22. A : *bah ouais ! c'est une* poubelle magique (.)

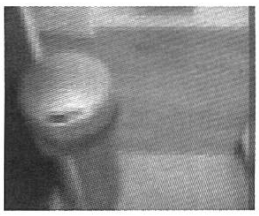


Figure 30: première image du plan fixe poubelle

23. regarde

Figure 31: passage du doigt devant le capteur

24. 1.53

25. B : mais ça coûte une (inc) cette poubelle non ?

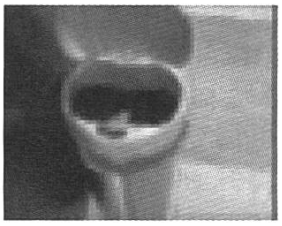


Figure 32: position ouverte

Alors que la méthode d'inventaire visuel systématique se caractérisait auparavant par un couplage plan fixe/désignation de la cible de la part du montreur, le destinataire s'auto-sélectionne (l. 18-21) pour réintroduire (l. 4) l'évaluation concernant la poubelle. Le tour comporte plus précisément une évaluation positive répétée de l'objet, sa matière ("en métal"), et une mention de sa visualisation antérieure ("j'ai vu"). Cette thématisation est prise en compte par le montreur qui pendant la délivrance de ce tour (*l. 19*) initie un mouvement de caméra aboutissant à un plan fixe (l. 22) de l'objet mis en relief par le destinataire. L'évaluation positive d'un objet antérieurement aperçu par le destinataire modifie ainsi l'organisation du plan-séquence puisque le montreur ajuste le cadrage pour le rendre visualisable et disponible pour l'interaction en cours. Du point de vue du destinataire, le nouveau plan fixe favorise une analyse plus poussée de l'objet, et cette posture est renforcée par l'engagement du montreur dans une démonstration d'une particularité de la poubelle, dite "magique" (l. 22). Après l'évocation de cette caractéristique, il instruit le destinataire pour le positionner en tant que spectateur (l. 23), et en fin de tour, il réalise un plan rapproché du couvercle tout en passant le doigt (Figure 31) devant un capteur activant l'ouverture de la poubelle. Le silence qui accompagne cette démonstration (l. 24) permet au spectateur de s'auto-sélectionner et il introduit une question concernant le prix d'une telle poubelle. Cette question débutera un développement thématique assez étendu (non transcrit) quant à son tarif, les magasins la proposant, etc.

Le destinataire peut ainsi modifier la conception du plan-séquence qui se caractérisait jusque-là par un inventaire systématique, basé sur l'association de plans fixes et de tours de parole conjoints déclinant la nature des éléments cadrés. Si cette réalisation est adaptée au projet de "visite" tel qu'il a été initié

et ratifié par les participants, on constate qu'il est constamment "exposé" à différentes interventions du spectateur: par des questions, des requêtes, des évaluations, des descriptions, etc., celui-ci contribue à a) l'intelligibilité même des images produites, b) la conception de chaque plan fixe (quand n'est-il plus pertinent?) et finalement, c) à l'organisation d'ensemble du plan-séquence. Ce rôle actif du destinataire est particulièrement saillant lorsque la mention d'un objet hors champ aboutit à un recadrage le rendant visualisable, et qui devient un objet de la conversation. Ce type de séquence révèle un principe dont nous avons plusieurs occurrences dans le corpus, qui se rapproche de l'obligation de filmer le locuteur actuel dans les communications multipartites: en configuration "têtes parlantes" ou lors d'une période monstrative, toute référence à l'environnement proximal rend pertinent sa mise à l'écran, et constitue une nouvelle ressource pour la structuration et l'orientation de l'interaction en cours.

5. Conclusion

Nous avons examiné l'organisation de séquences au cours desquelles l'un des participants d'un échange visiophonique montre certaines composantes de son environnement. Nous avons montré que de telles séquences sont comprises dans l'organisation plus générale des communications vidéo, où les participants s'orientent vers une maxime fondamentale consistant à *mettre le visage du locuteur actuel à l'écran*. S'il n'y a rien de particulier à montrer, ils suivent cette maxime en se positionnant au centre de l'écran, et les têtes parlantes se présentent comme le format "par défaut" des interactions sur téléphone mobile ou Skype. Tout abandon de ce positionnement est susceptible de générer des problèmes d'intelligibilité et de pertinence des images produites, dont sont continûment responsables les participants. Les images sont en effet constamment inspectées par le destinataire en vertu de leur pertinence pour l'interaction en cours. L'initiation d'une séquence monstrative, qui vise précisément à montrer autre chose que le visage de l'un des conversants, implique d'être réalisée de façon méthodique.

Une telle séquence peut être annoncée, proposée par le "montreur" ou demandée par le "regardeur", ces deux catégories de participation étant à ce stade potentielles. Ceci constitue le premier tour de parole d'une structure de préface, qui offre une place à l'autre participant pour ratifier (ou non) le projet. Dans nos corpus, la seconde partie de paire est quasi systématiquement un accord,[5] et le montreur peut dès lors orienter la caméra vers une composante de l'environnement. Cependant, nous avons vu que cette orientation générait des images de transition sur des périodes plus ou moins importantes, et sans que la préface (offre, proposition, demande, etc.) intègre explicitement une

---

5 Parfois différé par des séquences de réparation, voir Morel & Licoppe (2009: 187-188).

suspension de la pertinence du voir dans la vidéo-en-interaction. Dans le cas où la durée des images transitionnelles se prolonge (liée à un changement de caméra, un déplacement, etc.), le montreur en devenir cherche à bloquer toute séquence conversationnelle et à positionner la première image monstrative comme la prochaine action pertinente. Pour cela, il peut s'engager dans des tours de parole visant à suspendre l'interaction en cours ("attends") ou réitérer le projet de monstration sous la forme d'annonces. Le format de la préface constitue également une ressource pour déterminer quand sera réalisée la première image monstrative, et quel type de prise de vues pourra être mis en œuvre. Ainsi, le projet "faire un tour d'une pièce" rend pertinent un panoramique (même si celui-ci débute par un bref plan fixe); le plan fixe du sol faisant suite à l'instruction "fais voir comment t'as rangé" suffit pour que le destinataire fonde son jugement; le plan fixe "porte d'entrée" débute habilement le projet "visite d'appartement".

Les participants à des communications vidéo sur Skype et mobile comptent sur le fait qu'image et voix constituent des champs sémiotiques mutuellement élaboratifs, générant une sorte de "configuration contextuelle vidéo". Toutefois, au sein de ces communications, les participants sont constitués en tant que paire relationnelle interactionnellement générée, "montreur"/"regardeur", disposant de ressources et de préoccupations asymétriques quant à la production des images et du parler-en-interaction. Le montreur combine parole et voix pour produire des images intelligibles, avec une attention constante quant à leur pertinence pour le projet et à destination du regardeur; ce dernier contribue de façon variée à la structuration du plan-séquence, selon la nature du projet (inspection/visite par exemple), la pertinence des prises de vues (traitement de mauvais cadrages par exemple), et il peut étroitement guider la caméra vers des composantes particulières de l'environnement (via des requêtes, des demandes, etc.). Si les participants interagissent généralement de sorte que l'organisation séquentielle épouse l'organisation des prises de vues au sein des périodes monstratives, ils peuvent néanmoins s'engager dans des développements thématiques distincts des images produites ayant pour conséquence une suspension de la progression du plan-séquence (maintien du dernier plan fixe ou relâchement du cadrage). La production pas à pas des images dans les communications Skype ou sur mobile repose sur un travail collaboratif et un processus interactionnel conjoint. Chaque participant est un "conversant vidéo", agissant constamment pour la conception d'images appropriées et à partir desquelles se structure l'interaction. Le parler-en-interaction et la vidéo-en-interaction ne sont pas des champs sémiotiques indépendants, mais ils sont produits et traités par les participants comme mutuellement élaboratifs: c'est à partir de ces ressources que les participants produisent l'ordre multimodal particulier des échanges vidéo.

BIBLIOGRAPHIE

De Fornel, M. (1988). Contraintes systémiques et contraintes rituelles dans l'interaction visiophonique. *Réseaux, 29*, 35-46.

De Fornel, M. (1994). Le cadre interactionnel de l'échange visiophonique. *Réseaux, 64*, 107-132.

Goffman, E. (1963). *Behavior in public places. Notes on the social organization of gatherings.* New York: The Free Press.

Goffman, E. (1978). Response cries. *Language, 54/4*, 787-815.

Heath, C. & Luff, P. (1992). Media space and communicative asymmetries. Preliminary observations of video mediated interactions. *Human Computer Interaction, 7*, 315-346.

Jefferson, G. (1978). Sequential aspects of storytelling in conversation. In: J. Schenkein (éd.), *Studies in the organization of conversational interaction* (pp. 219-248). New York: Academic Press.

Kendon, A. (1990). *Conducting interaction. Patterns of behavior in focused encounters.* Cambridge: Cambridge University Press).

Licoppe C. & Morel J. (à paraître). Video-in-interaction: "Talking heads" and the multimodal organization of mobile and Skype video calls. *Research on Language and Social Interaction.*

Mondada, L. (2003). Working with video: how surgeons produce video records of their actions. *Visual Studies, 18/1*, 58-73.

Mondada, L. (2007a). Operating together through videoconference: Members' procedures for accomplishing a common space of action. In: S. Hester & D. Francis (éds.), *Orders of ordinary action* (pp. 51-67). Aldershot: Ashgate.

Mondada, L. (2007b). Imbrications de la technologie et de l'ordre interactionnel. L'organisation de vérifications et d'identifications de problèmes pendant la visioconférence. *Réseaux, 144*, 141-182.

Morel, J. & Licoppe, C. (2009). La vidéocommunication sur téléphone mobile: quelle mobilité pour quels cadrages? *Réseaux, 156,* 165-201.

Morel, J. & Licoppe C. (2011). Studying mobile video telephony. In: M. Büscher, J. Urry & K. Witchger (éds.), *Mobile methods* (pp. 164-181). London/New York: Routledge.

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*, 73-100.

O'Hara, K., Black, A. & Lipson, M. (2006). Everyday practices with mobile video telephony. *Proceedings of CHI 2006* (pp. 871-880). New York: ACM Press.

Relieu, M. (2002). Ouvrir la boîte noire. Identification et localisation dans les conversations mobiles. *Réseaux, 112/113*, 20-47.

Relieu, M. (2007). La téléprésence, ou l'autre visiophonie. *Réseaux, 144*, 183-223.

Sacks, H. (1992). *Lectures on conversation.* Oxford: Blackwell.

Sacks, H., Schegloff, E. A. & Jefferson, G. (1974). A simplest systematics for the organization of turn taking for conversation. *Language, 50/4*, 696-735.

Urry, J. (2007): *Mobilities.* Cambridge: Polity Press.

Watson R. (1997). Some general reflections on 'categorization' and 'sequence' in the analysis of conversation. In: S. Hester & P. Eglin (éds.), *Culture in action. Studies in membership categorization analysis* (pp. 49-76). Washington DC: University Press of America.

Whittaker, S. (2003). Things to talk about when talking about things. *Human Computer Interaction, 18/1,* 149-170.

Annexes

Conventions de transcription

°blabla°	parler bas
BLABLA	parler fort
.	intonation descendante
?	contour interrogatif
[	chevauchement de parole
=	énoncés liés
(.)	silence inférieur à un dixième de seconde
1.23	silence exprimé en seconde
:::::	allongement vocalique
blabla	mouvement transitionnel de caméra
* *	balises d'une période de monstration particulière
bl$\boxed{\text{abl}}$a	l'encadré correspond à la période d'une image clé figurant en dessous
((blabla))	commodité de transcription
(inc)	segment incompréhensible