

Observation et capture de données sur l'interaction multimodale en mobilité

Moustapha Zouinar, Marc Relieu

France Telecom R&D
38-40 rue du Général Degaulle
92794 Issy-Les-Moulineaux, France
{marc.relieu, moustafa.zouinar}@rd.francetelecom.com

Pascal Salembier, Guillaume Calvet

GRIC-IRIT
118, rte de Narbonne,
31062 Toulouse cedex, France
{salembier, calvet}@irit.fr

RÉSUMÉ

Dans cet article, nous présentons un dispositif d'observation et d'enregistrement permettant de recueillir des données sur l'usage de systèmes multimodaux en mobilité. Ce dispositif articule deux prises de vues, une prise de vue « utilisateur » et une prise de vue « externe » du contexte, avec un recueil automatisé de données multimodales sur l'interaction en temps réel.

Mots clés

Mobilité, systèmes mobiles, observation de l'activité, recueil de données d'interactions, multimodalité.

Abstract

This paper describes a method and a data collection infrastructure for capturing data about the use of a mobile multimodal system in the field, under realistic conditions. The infrastructure includes instrumentation worn by the user to capture his perspective while moving, a means of capturing a wider view of the context and a multimodal data logger.

Categories and Subject Descriptors

H.5.2 [Information Interfaces and Presentation]: User Interfaces – *Evaluation/Methodology*.

General Terms

Experimentation, Human Factors.

Keywords

Mobility, observation of activity, interaction data capture, multimodal systems.

INTRODUCTION

Le développement croissant des systèmes mobiles (PDA, téléphone portable, etc.) pose de nouveaux problèmes méthodologiques et techniques dès lors qu'il s'agit d'étudier empiriquement l'usage de ces systèmes. La mobilité des utilisateurs implique un changement d'environnement et des déplacements qui rendent le recueil de données sur l'interaction avec le système ou sur le contexte dans lequel elle se produit plus difficile à mettre en œuvre que dans le cas standard des activités postées. Parmi les rares études qui ont été consacrées à ce problème de l'étude empirique de l'interaction avec des systèmes mobiles [1], deux approches, reposant sur des positionnements épistémologiques différents, peuvent être distinguées. Une première approche vise le développement de méthodes expérimentales pour étudier l'interaction en

laboratoire. Sous-tendus par la recherche d'un contrôle des variables pouvant affecter l'interaction, les travaux menés dans le cadre de cette approche tentent par exemple de mettre artificiellement l'utilisateur en situation de mobilité en lui demandant par exemple de marcher sur un tapis roulant [10] [1]. L'hypothèse sous-jacente à cette approche est qu'il est donc possible de recréer en laboratoire des situations de mobilité « naturelle » de l'utilisateur. De ce point de vue, la mobilité est essentiellement envisagée sous l'angle de ses manifestations physiques ou attentionnelles [1]. Un tel réductionnisme pose le problème classique de la validité écologique des données recueillies. En effet, la relation entre la mobilité et le contexte n'est pas posée comme essentielle. Or un déplacement propose à l'acteur des changements plus ou moins importants du contexte, aussi bien dans sa dimension physique (ambiance lumineuse ou sonore, objets présents, etc.) que socio-culturelle (lieux, personnes, situations, etc.). Même si elles n'ont pas pour autant d'incidence automatique sur le déroulement de l'activité, ces variations sont susceptibles de devenir pertinentes pour l'activité et même en orienter le cours. En outre, l'interaction homme-machine génère elle-même son propre contexte d'accomplissement (historique) et sa dynamique, centrée autour de la réalisation d'une activité donnée (recherche d'information, écriture de messages, etc.). La mobilité pose donc des problèmes complexes d'articulation et de combinaison entre différents types de contexte et d'activité. Une autre voie de recherche d'orientation naturaliste met l'accent sur la conception de dispositifs permettant de réaliser des observations en situation naturelle, c'est-à-dire hors laboratoire (par exemple [6], [9]). Cette approche s'efforce de préserver le plus possible le caractère naturel de la mobilité et par conséquent du contexte. On gagne ainsi en pertinence écologique ce que l'on perd en contrôle de la situation. Cependant l'approche naturaliste rend plus difficile le recueil de données sur l'interaction avec ces systèmes puisque l'utilisateur peut évoluer dans des environnements différents. Les méthodes de recueil qui sont utilisées pour étudier l'interaction en situation fixe ne sont plus adaptées dès lors que l'utilisateur se déplace avec le système. L'un des principaux problèmes posés par l'approche naturaliste est donc celui de déterminer les moyens à mettre en œuvre pour étudier empiriquement et en situation l'usage des technologies mobiles. Dans cette perspective, nous avons élaboré un dispositif de recueil de données sur l'usage de systèmes en situation de mobilité, dans le cadre d'une étude sur l'interaction multimodale.

CRITÈRES DE CONCEPTION DU DISPOSITIF

Une perspective située "ouverte"

Notre objectif est d'étudier empiriquement l'interaction multimodale en mobilité, par l'observation et l'analyse de l'activité de l'utilisateur dans un environnement non artificiellement recréé, en recourant à l'enregistrement de celle-ci et à celle du système utilisé. La mobilité entraînant un changement de contexte, il est également nécessaire de recueillir des données sur ce contexte, dans la mesure où il peut influencer l'interaction avec le système. Par exemple, un changement de modalité peut être effectué pour faire face à une modification soudaine d'ambiance sonore ou lumineuse. Le contexte « pertinent » n'étant pas définissable a priori dans son intégralité, surtout lorsque l'utilisateur se déplace dans un environnement « ouvert », il nous faut en même temps être en mesure pouvoir identifier a posteriori les éléments contextuels qui affectent réellement l'activité de l'utilisateur, ainsi que la nature de cette relation.

Cette orientation méthodologique impose cependant plusieurs contraintes qu'il faut prendre en compte. Une première contrainte est que le dispositif doit être portable et suffisamment autonome pour suivre l'utilisateur dans les différents environnements qu'il est susceptible de traverser au cours de son déplacement, et le gêner le moins possible que ce soit physiquement (par exemple, le poids ou l'encombrement) psychologiquement, ou socialement (vis-à-vis d'autrui). Une première méthode consiste à demander à un observateur d'effectuer ce suivi tout en filmant de près l'utilisateur. Cette méthode présente cependant plusieurs inconvénients. Un premier inconvénient est le sentiment de gêne que la présence très proche d'un observateur équipé d'une caméra peut provoquer chez l'utilisateur. Une deuxième limite tient à la difficulté de filmer simultanément les actions de l'utilisateur et les événements système pertinents (interface, modalités en entrée ou en sortie utilisée). Ces actions sont souvent rapides et l'observateur doit constamment s'ajuster aux comportements de l'utilisateur, par exemple en déplaçant la caméra ou en jouant sur le focus de celle-ci, afin de pouvoir « capturer » ses actions ou l'interface du système. Ces ajustements, qui peuvent s'ajouter à la fatigue physique inhérente à la tenue prolongée de la caméra, nuisent à la qualité des données obtenues. Enfin, cette méthode de recueil achoppe sur l'indétermination a priori de la localisation des éléments contextuels pertinents pour l'usage. En effet, le preneur de vues est confronté au dilemme suivant : soit il privilégie un plan centré sur la manipulation du dispositif mobile, et exclut les éléments contextuels accessibles à un cadrage plus large (par exemple les interactions avec d'autres individus co-présents [13]), soit il choisit un plan large, et présélectionne ainsi des facteurs contextuels globaux, au détriment des logiques d'interaction avec le dispositif technique. Ainsi se reproduit, au niveau des choix "techniques" de prises de vue, le grand écart entre, d'un côté, une perspective d'inspiration plutôt sociologique, qui privilégie la prise en compte des interactions non médiatisées avec des tiers co-présents, et, de l'autre, une orientation plus

traditionnellement "IHM", qui se centre sur l'interaction d'un utilisateur avec l'objet technique.

Enfin, dans l'idéal, le dispositif devrait permettre de réaliser des observations sur le long terme afin de capter les évolutions dans les pratiques de l'utilisateur, sans poser des problèmes de tolérance (physique et sociale) du port du dispositif sur une période prolongée. Le dispositif doit par ailleurs faciliter le stockage et le traitement de la masse importante de données qui peut être générée. Ce point ouvre sur les questions relatives aux moyens nécessaires à un codage rapide et efficace de corpus importants [7].

DESCRIPTION DU DISPOSITIF

Ces différentes contraintes nous ont conduits à développer un dispositif mobile, peu gênant au plan physique, relativement discret et permettant de recueillir des données sur l'activité de l'utilisateur, sur l'interaction avec le système et sur le contexte. Ce dispositif est composé de trois modules de recueil (figure 1) : un système de prise de vue large du contexte (vue contextuelle observateur), un système de prise de vue du champ visuel de l'utilisateur (vue « subjective »), et un système de capture automatique des actions de l'utilisateur sur le système et les sorties de celui-ci (capture log).

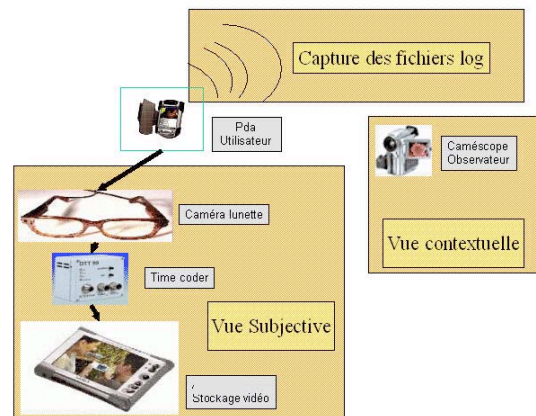


Figure 1. Dispositif d'observation mobile.

Vue contextuelle "observateur"

Pour identifier a posteriori les éléments contextuels qui vont permettre de rendre intelligibles les actions de l'utilisateur, il est essentiel de capturer les informations contextuelles pertinentes pour l'activité. Cette capture peut se faire soit à l'aide de caméras fixées dans l'environnement, soit au moyen d'une caméra mobile. Les vues fixes, quand elles sont possibles au plan logistique, présentent l'inconvénient majeur d'offrir une perspective unique sur l'action et de ne pas pouvoir la modifier. Aussi nous avons choisi de filmer à distance l'utilisateur à l'aide d'un caméscope porté par un observateur extérieur et mobile. En effet, un observateur peut ajuster continuellement la prise de vues à la dynamique de l'activité et du contexte. Cette vue large permet de situer l'utilisateur dans son contexte global d'usage, information qui n'est pas toujours disponible dans une vue restituant le champ visuel de l'utilisateur.

Vue subjective

La capture du contexte n'est pas pour autant suffisante, car elle ne donne pas accès au point de vue de l'utilisateur, notamment lorsque celui-ci porte son attention sur des objets ou interagit avec eux. Pour rendre cet accès possible, nous avons choisi d'équiper l'utilisateur de lunettes-caméra [11] qui permettent de filmer l'environnement présent dans son champ visuel en garantissant ainsi l'accès à l'interaction avec le dispositif, en particulier les opérations manuelles réalisées sur le dispositif (navigation).



Figure 2. Exemple de prises de vue « perspective utilisateur ».

Elle est en outre associée à une prise de son permettant de bénéficier de l'environnement sonore de l'activité : interaction vocale avec les outils testés, interaction avec autrui. Afin de disposer d'une trace la plus satisfaisante (lisible) possible des événements qui se produisent sur et autour de l'interface via les lunettes-caméra, il est nécessaire de disposer d'une focale adéquate et d'orienter la caméra selon un axe différent de celui du regard du sujet. L'enregistrement se fait directement sur un disque dur qui est porté par l'utilisateur.

Capture log

La vue utilisateur (vue subjective) n'est pas toujours suffisante pour capturer les événements se produisant sur l'écran du dispositif, notamment à cause des reflets qui peuvent se produire sur l'écran ou lorsque ce dernier se retrouve hors du champ de la caméra. Pour pallier ce problème, nous avons opté pour le développement d'une plateforme logicielle qui permet de capturer automatiquement des traces "systèmes" de l'interaction. Cette plateforme a été conçue sur la base de besoins définissant les éléments à capturer pour permettre une analyse fine de l'activité des utilisateurs [8]. Outre la capture, cette plateforme dispose également de fonctions de filtrage des traces et d'exports de celles-ci vers d'autres applications facilitant ainsi leur traitement. Le filtrage permet à l'analyste de sélectionner le corpus de données qui l'intéressent et leur format de présentation.

Le re-jeu

Pour analyser les données recueillies à travers les vues contextuelles et subjectives, nous disposons d'un outil de visualisation combinant plusieurs fenêtres vidéo actives simultanément, qui permettent de piloter et de visualiser de façon synchrone les films qui correspondent à ces deux vues (figure 3). Cet outil permet également de réaliser directement à partir de ces fenêtres des codages de l'interaction et de construire des chroniques d'activité. Dans le cas où il y aurait plusieurs utilisateurs qui communiquent entre eux, ce dispositif permet de suivre facilement leur interaction.

Les vues contextuelles et subjectives visent également à nous servir de support pour mener des auto-confrontations

avec les sujets. L'auto-confrontation consiste à recueillir les commentaires des sujets en les confrontant aux traces de leur activité. Elle vise à accéder à des éléments inobservables ou à compléter des analyses d'éléments difficilement observables sur la seule base des données d'observation. L'auto-confrontation peut être conduite depuis différentes perspectives, selon la façon dont on envisage la relation de l'action au discours rétrospectif.

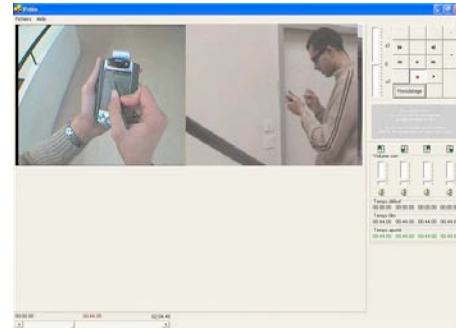


Figure 3. Interface de l'outil de visualisation et de codage des données (version adaptée d'Actogram [4]). Cet outil permet de visualiser les enregistrements vidéo des vues subjectives et contextuelles. Les deux images représentent (de gauche à droite) un exemple de vue subjective et un exemple de vue contextuelle. La partie droite de l'interface permet de piloter les vidéos et de réaliser le codage de celles-ci.

Illustration de l'utilisation du dispositif et limitations

Le dispositif dont il est question ici a été testé en situation, dans différents environnements et avec différents sujets. Du point de vue des critères que le dispositif devait remplir, les résultats obtenus ont été globalement satisfaisants. Cependant quelques limitations non négligeables ont été observées ; nous n'évoquerons que celles relatives à l'utilisation des lunettes-caméra dans la mesure où elles nous paraissent les plus intéressantes.

Limites relatives aux lunettes-caméra : vision centrale

L'utilisation de lunettes-caméra pour capter la zone du champ visuel sur laquelle l'utilisateur porte son attention à un instant donné pose le problème de l'orientation de la caméra (prise de vue), surtout lorsque l'activité de l'utilisateur, comme c'est fréquemment le cas en mobilité, implique d'importants mouvements de son corps et de sa tête. Dans ce cas, la capture de ce qui est effectivement regardé par le biais des lunettes-caméra peut être problématique (figure 4).



Figure 4. Problèmes d'utilisation du dispositif de prise de vue "subjective". Sur la photo de gauche, l'utilisateur fixe alternativement un PDA et un objet dans une vitrine, alors que la caméra subjective ne restitue qu'une vue de la partie

inférieure du présentoir. Sur la seconde photo l'utilisateur fixe du regard le PDA tenu de côté alors que la caméra enregistre ce qui reste dans l'axe de la tête de l'utilisateur.

Une solution à ce problème pourrait consister dans l'utilisation d'un dispositif de suivi du regard (eyetracking). Un tel dispositif peut fournir une description précise des zones "regardées" (fixations oculaires) par un sujet. Des oculomètres portables ont été développés pour enregistrer les parcours oculaires d'utilisateurs sur des interfaces de systèmes mobiles [3] [12]. Pour autant la maturité technique de ces systèmes ne permet pas encore d'envisager de les utiliser dans n'importe quel environnement. Ils se révèlent en outre peu adaptés dès qu'il s'agit de les utiliser dans des lieux publics où ils risquent d'attirer systématiquement l'attention d'autrui, et donc de placer l'utilisateur dans une situation socialement inconfortable. Par ailleurs le problème de la pertinence et de l'interprétation de ce type de données reste posé.

Limites relatives aux lunettes-caméra : vision périphérique
Le fait d'utiliser une focale inadaptée (avec un angle de prise de vue trop étroit) ne prend pas en compte la vision périphérique des utilisateurs ; de ce fait, elle est absente de la trace enregistrée (effet « tunnel »). Or, comme le montrent de nombreux travaux, l'attention périphérique participe également à l'organisation de l'action et de la cognition en situation [2]. Cet écart peut être à l'origine de difficultés de remise en situation si l'on utilise l'enregistrement comme support à la verbalisation à posteriori.

Limites relatives au recueil et au traitement des données
La masse importante de données pouvant être recueillie par le dispositif pose une autre limite d'ordre technique. Outre des problèmes de stockage, le fait que ces données proviennent de différentes sources nécessite un travail relativement fastidieux de synchronisation.

Pour le moment, la plateforme ne permet pas de rejouer l'interaction. L'idéal serait de pouvoir rejouer complètement l'interaction sur un PDA ou un PC afin de disposer du maximum d'informations sur la dynamique de celle-ci. Les fichiers de capture automatique des traces de l'interaction constituent une première étape pour atteindre cet objectif ; il reste cependant à résoudre des problèmes techniques, notamment de synchronisation.

Enfin, de nombreuses questions demeurent sur la réaction des sujets face à une utilisation prolongée de ces lunettes. Des études menées en situation de bureau tendent à montrer que ce type de dispositif peut être porté pendant plusieurs heures sans aucune gêne visible [5]. Mais il reste à démontrer que cela est également vrai dans des situations plus "ouvertes" (espaces publics).

L'identification plus précise de ces limitations et leur traitement constitue une des principales perspectives de notre travail pour améliorer le dispositif d'observation.

REMERCIEMENTS

Ce travail est mené dans le cadre des activités de la plateforme RNRT LUTIN (<http://www.lutin.utc.fr>) et de

l'Equipe Projet Multi Laboratoires STIC-CNRS ELIXIR (<http://www.stic-elixir.org/>).

REFERENCES

1. Beck, E., Christiansen, M., Kjeldskov J., Kolbe, N. & Stage, J. Experimental Evaluation of Techniques for Usability Testing of Mobile Systems in a Laboratory Setting. In *Proceedings of OzCHI 2003*, Brisbane, Australia, 2003
2. Heath, C. & Button, G. (eds) Special Issue on Workplace Studies. *British Journal of Sociology*, Vol. 53, No. 2, 2002.
3. Jameson, A. Usability Issues and Methods for Mobile Multimodal Systems, in *Proceedings of the ISCA Tutorial and Research Workshop on Multi-Modal Dialogue in Mobile Environments* Kloster Irsee, Germany, June, 2002.
4. Kerguelen, A. Quels outils concevoir pour aider au relevé d'observation sur le terrain ? *XXXIIIème Congrès de la SELF, "Temps et Travail"*, Paris, 1998,.
5. Lahlou, S. Observing cognitive work in offices. In N. Streitz, J. Siegel, V. Hartkopf & S. Kosomi (eds). *Cooperative Buildings. Integrating Information, Organizations, and Architecture*. Lecture Notes in Computer Science, 1670, Heidelberg : Springer, 1999.
6. Lyons, K. & Starner, T. Mobile capture for wearable computer usability testing. In *Proceedings of the Fifth International Symposium on Wearable Computers*, Los Alamitos, CA: IEEE Computer, 2001.
7. Mark, G., Christensen, U., Shafae, M. A Methodology Using a Microcamera for Studying Mobile IT Usage and Person Mobility. Paper presented at the Workshop on Mobile Communication, 2001.
8. Merlin, B. Plateforme générique d'observation des interactions. *Rapport technique IRT-Diamant*, 2004.
9. Oviatt, S. Multimodal. System processing in mobile environments. In *Proceedings of the 13th Annual ACM Symposium on User Interface Software and Technology*, ACM Press, 2000.
10. Pirhonen, A., Brewster, S. & Holguin, C. Gestural and audio metaphors as a means of control for mobile devices. In *proceedings of CHI 2002*, 291-29
11. Relieu, M. The « glasscam » as an observational tool for studying screen-based mobile phone uses and management of parallel activities. *International Conference on Conversation Analysis (ICCA-02)* Copenhagen, 17-21 May, 2002
12. Seagull, F.J & Xiao, Y. Using eye-tracking video data to augment knowledge elicitation in cognitive task analysis. In *Proceedings of the Human factors and Ergonomics Society 45th Annual Meeting*, October 8-12, Minneapolis, Minnesota, USA, 2001
13. Vom Lehn, D., Heath, C. & Hindmarsh, J. La découverte interactive des objets de musée. *Alliage: culture, science, technique*, 2003, pp 165-178.

