Interdisciplinary Institute for Innovation

Patterns of Gaze Switching in the "Naturally-Occurring" Uses of Smartphones in Urban Mobile Settings

Christian Licoppe
Julien Figeac

Working Paper 13-TS-01

June, 2013

Department of Economics and Social Sciences
TELECOM ParisTech
46 rue Barrault, 75013 Paris, France
Email: wpecon@telecom-paristech.fr

Patterns of Gaze Switching in the "Naturally-Occurring" Uses of Smartphones in Urban Mobile Settings

Christian Licoppe
Professeur
LTCI (UMR 5141)
Télécom ParisTech
46 rue Barrault, 75013 PARIS
christian.licoppe(at)telecom-paristech.fr

Julien Figeac
Chargé de recherche CNRS
LTCI (UMR 5141)
Télécom ParisTech
46 rue Barrault, 75013 PARIS
julien.figeac(at)telecom-paristech.fr

ABSTRACT

We report on the development of a method for observing and recording the uses of mobile communications 'on the move', based on the combination of context-oriented recordings made with user-worn camera glasses with mobile screen capture data. We show how this allows the temporal organization of gaze switches (to and away from the mobile screen) to be observed and documented, thus providing crucial empirical information to understand how users actually manage mobile communication as well as other activities in everyday multi-activity settings. We report on the findings of an empirical study of smartphone use in transport situations. Being oriented towards multi-activity appears as a particular form of attunement to the potential sequential implicativeness of events occurring both in the navigation of mobile communication applications or the mobility environment, i.e. as possible occasions to switch the orientation of one's gaze from one activity-relevant field of activity to another.

Keywords:

Mobility, mobile communication, smartphone, mobile phone, car, multi-activity, camera glasses, gaze switch.

INTRODUCTION

The development of increasingly powerful smartphones has led to an increasing use of mobile Internet and social media on the move [20]. However, the study of mobile uses presents some specific methodological difficulties. The adoption of Internet applications has been studied over extended periods by quantitative mining of connection logs [31]. While such methods are very powerful in terms of the scale at which the analysis can be conducted, knowledge about the actual situations of use is poor and has to be inferred indirectly. Qualitative interview-based methods, such as self-reports [11], are richer in terms of context awareness and user experience, but they are also post-hoc narratives with at best only a loose relationship with the fine details of the actual situations in which smartphones are used.

Video recordings made with portable devices enable the development of promising methods because they may 'follow' the user and record his visual experience. This constitutes a powerful way to obtain naturally occurring data regarding how smartphone users manage mobile applications in a way that is sensitive to the context of use and to the other activities they might be involved in at the time. Cameras that follow users along their mobility paths have been used in visual anthropology, where the analyst 'shadows' and visually records the person's engagement in places and orally records her comments to produce a self-centered visual tour [26], and this method has also been applied to the ethnography of the uses of mobile phones in urban mobility settings [33]. Multimodal interaction analysis has focused on car situations (a mobility context in which it is easier to 'follow' users by putting cameras in the car [19, 22]). Some of this work has explicitly looked at the way car users might collaboratively manage phone-related events [4, 9] or GPS-based information [2]. However, such setups are not adapted to other mobile settings. The most flexibly portable video recording devices are those that are carried or worn by the users themselves. Their wearability ensures the possibility of following the users in all sorts of contexts.

There again, different recording methods provide different empirical data and support different types of analytical claims. When the camera is worn only on the torso of the user, it aims to provide a view of the context and a general sense of the user's body, without much access to what he might actually be gazing at or attending to. When the video recording is produced from the user's head, stronger claims are usually made about capturing, more or less accurately, the user's visual experience 'in action', at the cost of transforming him into a kind of 'cyborg', as in Steve Mann's famous experiments. Oculometric methods are most precise because they record eye movements and fixation patterns [32]. They have also been applied to the use of mobile phones [3] and even outside the laboratory [29], but oculometric devices are

not easily portable and may be cumbersome to “in the wild”, particularly for the study of naturally-occurring activities. Another approach is to ask the user to wear a single camera on his head, either on the subject’s forehead [25] or on specially fitted glasses, sometimes described in literature as ‘video glasses’, ‘subcams’ or ‘camera glasses’ [16].

Camera glasses have been claimed to enable the analyst to record the evolution of the subject’s visual field in a situation and capture something of his subjective perspective [17]. They have been used by several groups to study mobile phone uses in natural settings [5, 21, 27], either alone or coupled to log data analysis [24, 33]. But when camera glasses alone are used, they rarely provide a resolution of the mobile phone screen good enough to grasp the finer details of mobile use. It is therefore useful to complement this context-aware device with a screen capture recording, for the latter will provide precise and independent access to the user’s activity on the mobile terminal [28]. Since we are interested in fine-grained data about how users manage the joint demands of mobile communication and mobility, we have decided to combine here the recording produced by camera glasses with the mobile screen video capture apparatus.

The use of complex video methods to record mobile uses has been criticized as being too time-consuming and heavy-handed while providing little new in terms of traditional usability concerns [15]. Such criticism is apt when these methods are compared to laboratory research concerns, but it fails to tackle the effectiveness of such methods to highlight usability problems that are actual user concerns. These problems are tightly tied to the particulars of the real life situation of use, as is the case when using smartphones in public settings. In this paper we will aim to show precisely how the contextual data provided by our dual recording system can document the analysis on the manner in which sequential aspects orient users towards the ongoing situation. These data equally help to identify some local and situation-specific user concerns, while carefully avoiding hyperbolic claims that we would gain access to the user’s subjective experience, or that what the analyst can ‘see’ through the camera glasses video is actually what the user ‘sees’.

The paper therefore makes four distinctive and interlocked contributions, moving all the way from observation to design, and through the construction of relevant phenomena and testable analytical concerns:

a) It describes a method for recording simultaneously mobile screen activity through a video capture system combined with camera glasses to provide data on the involvement of the mobile user with the larger environment.

b) It applies this method to a sample of mobile urban commuters, and it shows that the crucial analytically-relevant phenomenon this method specifically makes empirically observable is the temporal organization of gaze switches, i.e. the way the mobile user's gaze moves towards or away from the mobile phone. By combining the data from the camera glasses and the screen capture recordings, gaze switches can be documented and their occurrence linked to that of various events happening in the mobile interface or the larger situation. Such data can therefore inform us on how users orient to contingencies relevant to mobile communication as well as to other relevant streams of activity (such as mobility) on a 'moment by moment' basis.

c) It shows how such knowledge about the sequential patterning of users' visual involvements opens up to a revision of key phenomena for understanding the organization of behavior in mobile contexts the precise definition of which has proved difficult and which have also proved difficult to pinpoint empirically, such as multi-activity. We will show in particular that with respect to the sequential organization of visual involvements, an orientation towards multi-activity may be defined and observed as a conduct in which events recognizable as breaks in one stream of activity (either with the mobile terminal or with respect to mobility concerns) are seized as transition-relevant, that is as opportunity to switch one's attention from one domain of activity to the other. Conversely, in this sequential perspective, being absorbed in an activity will appear as a tendency to ignore the transition relevance of such occurrences. Because this is empirically observable, it opens new research perspectives on multi-tasking.

d) Finally, we show how this is may be useful to the HCI community, by allowing to think of interfaces specifically in terms of their sequential implications, that is the way they provide a lot (or little) opportunities for the user in terms of transition-relevant points, and which we call here their "texture". Designing for multi-activity settings might therefore favor "rough" interfaces, the use of which will experience many potential transition points she might use as sequential resources to switch involvements.

Recent announcements from the mobile industry have suggested that the next generations of smartphones would have eye-tracking capacities, and might make gaze switches functional [3]. The meaning of gaze switches will then constitute a crucial usability issue, and the method and line of investigation proposed here will acquire an increased relevance, for it provides a way to grasp empirically the situated meaning of gaze switches.

METHODOLOGY

Analyzing the uses of mobile communication terminals in everyday settings requires synchronized empirical data regarding both the use of the mobile device and the off-screen involvements and activities of the users. What we have done is to combine smartphone screen captures with the recordings of the contexts of use produced by user-worn camera glasses. To capture the screen-based activity, we have used a feature of Android-based smartphones that allows an audio and video connection. We have therefore coupled the smartphone with a light portable A/V recorder (Figure 1), a method initially developed to study mobile video telephony [23].

When the system is on, it provides a recording of the changing mobile phone screen as the user performs various actions upon it, in the form of a video file. Such an approach does not work with iPhones yet, but the situation might change since future versions are announced to be connectable to TV monitors for game applications.

Figure 1. A simple portable set-up to record the audio-video flux on Android-based smartphones

While such a portable set-up may provide rich detailed data on the 'natural' uses of smartphones on the move, it provides very little information on the actual contexts of use. We have asked users to wear camera glasses with a small camera in the middle of the device to obtain a video recording of the users' changing environments that can be synchronized to the mobile device screen capture data. Camera glasses have been used by various authors to gain a 'subjective view' of actors performing various activities [17]. What was interesting to us was that the whole recording apparatus was portable and could be used to obtain data on the uses of smartphones in mobility settings.

(a) (b)

Figure 2. In all images the screen capture appears on the left, and the camera glasses recording on the right (both have been synchronized). In both images the user is sitting and using the smartphone in his lap, which is partly (a) or not at all (b) visible.

However, the temptation for the analyst is to look at what the camera glasses record as the actor's visual perception of his environment. Yet what you get is not what he sees. Because of various constraints, it is often a rather crude approximation of human perception. First, the recording field is only about 40 degrees wide for standard commercial camera glasses, which is much narrower than the human field of vision. So the camera glasses data will not show what is available in the peripheral visual field of the human subjects. Secondly, the visual focal axis of the recording is usually higher than that of the subject. There is an angle between them that may vary, among other things, with the geometrical shape of the subject's face and with the way the glasses are placed on the nose (which may change at two different moments). A consequence of this is that when different subjects are recording with camera glasses at the moment they are sitting and looking at their smartphones, the recording may feature the phone screen only partially (Figure 2a) or not at all (Figure 2b) in the video recording. In spite of all this, camera glasses are easy enough to use and wear to remain an attractive method to gather naturalistic observations on the move, provided we can somehow circumvent some of these limitations.

Although what you get is not what the subject 'sees' and the data has to be treated with care in the analysis, camera glasses may still be used, albeit with a degree of caution. First, the analyst is helped by the small size of the smartphone screen. To look at the smartphone, the user needs to orient his gaze in a rather precise direction, which can be reconstructed in the analysis and confirmed by the screen capture data (which shows what action he is currently involved in on the mobile interface), even if the smartphone is not actually visible on the video data. However, when the user looks away from the phone, it may become more difficult to infer what he might be looking at from the video glasses only (sometimes inferences can

be made). Hence, what camera glasses applied to the uses of smartphones on the move make visible (together with synchronized smartphone screen recording) are gaze switches, that is, when the users look towards or away from their smartphones. Such gaze switches are interpretable as switches between attending to the smartphone (which we may further document from the screen capture) and attending to other meaningful domains in the environment. Such data therefore provide a rich source of information on the way smartphone users may manage multiple involvements in actual public settings.

We will focus here on the joint management by users on the move of mobile social networking applications and everyday mobilities (and more specifically automobilities), in order to demonstrate the potential of our research method. We will gather empirical data on gaze switches towards and away from the smartphone, and we will analyze their temporal patterning. Our research question will then be "Why this gaze switch now?" a) to show that there are some recognizable and understandable patterns in the temporal placement of gaze switches; b) to show that such patterns inform us on the way the temporal organization of the different relevant activities are used as resources in their 'simultaneous' management; c) to draw some implications from this for the design of technologies and applications that might be particularly fitted to multi-activity settings.

FIELDWORK

We have recruited 10 users of mobile social networking applications, which accepted to wear camera glasses and record their mobile phone activities during their daily commuting for a period ranging from a week up to ten days. All of them signed a binding written argument providing us with the rights to use the data for scientific purposes, and under conditions of anonymization. It is important also to note that the French law allows video recording in public places, but forbids any publication of such data without the written authorization of the people who appear. In the subway setting the latter is not possible, and this means we cannot show the contextual data in a way in which someone would be recognizable, which constrains our use of such data. Two of them used their cars and the rest used public transportation. This provided us with twenty five hours of recordings. After retrieving the apparatus and the recordings, we synchronized the camera glasses recordings with the screen capture data to produce the kind of split-screen images shown in figure 2 and throughout this paper. This was our raw material for the analysis. We then scanned it to constitute collections of gaze switching events. In five instances we showed one set of data to the user to elicit their own interpretation of it.

At first glance, such data appear overwhelming in the sense that a few minutes will provide many instances of gaze switching. Moreover so many contingencies may be subtly conducive to gaze switching that it is not possible at this stage to treat them systematically (for instance with coding procedures). It is more fruitful to look at sequences which display recognizable sequential patterns (such as around projectable pauses) and try to build a solid analytical frame. Since our aim here is to show exactly what kind of research perspective on multi-tasking our approach provides, we will focus here on a single case to illustrate the possible contributions of such work (a car-driving case, which also circumvents some of the ethical constraints the subway material brings to the analysis), and use the data related to other users and settings (those taken in the subway, and in some instances some data taken at home) as background resources to buttress our arguments, or to show how different settings may be contrasted with respect to a particular point.

In order to highlight the potential of our method in a limited space, we will limit ourselves to the analysis of car-driving cases and use the data obtained in public transportation as a way to enrich our findings.

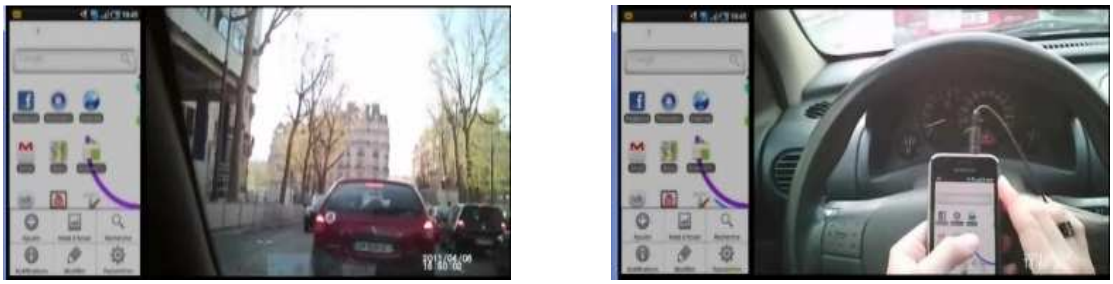
FINDINGS

OBSERVING THE MANAGEMENT OF JOINT ACTIVITIES: MOBILE COMMUNICATION AND TRANSPORT

Using smartphones on the move involves the moment-by-moment joint management of mobile communication and transportation. By using our empirical data on the occurrence and placement of gaze switches, we can understand some important aspects of the temporal organization of multiple involvements and shed some new light on the analytically elusive concept of multi-activity.

Gaze switches and the temporal organization of multiple involvements

In the example we will develop here, the smartphone user is a woman who connects to Facebook while driving. In the first set of data, she gets to a red light, stops (Figure 3a), looks down towards her lap and away from the road, places her smartphone on the driving wheel and launches the connection to Facebook, so as to check her list of recent posts on her wall (Figure 3b). Such conduct is illegal in France, but this is not the point here. What will interest us is rather the way she practically manages a dual orientation towards driving and using the Facebook application.



(a) (b)

Figure 3: a) Arriving at a red light behind the stopping traffic; b) Taking such an occurrence as an opportunity to gaze down, put the phone on the driving wheel and launch the Facebook application.

A red light is a constitutive feature of traffic management that is deceptively simple and familiar. It acts as an instruction for drivers to stop for a time that may be roughly anticipated by drivers with enough experience. It does not act by itself. As the speed bump studied by Science and Technology Studies [18], the red light is the surface of emergence of a large and heterogeneous network, an assemblage of people and artifacts 'delegating' their agency to the traffic light while remaining at a distance. It is this networked infrastructure that is 'agentive' in the way a red light recognizably and forcefully instructs us to stop. As a visual and semiotic cue which 'blackboxes' the networked infrastructure to which it is co-extensive, the red light projects a predictable pause for the surrounding traffic and therefore offers a relevant transition point in the driving activity. It affords a recognizable and convenient slot for doing something else such as picking up one's smartphone.

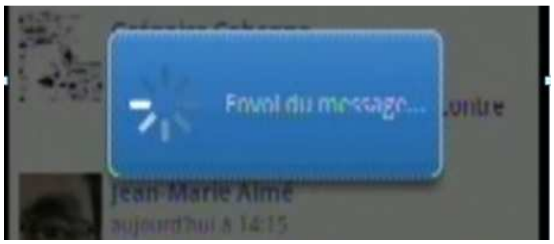
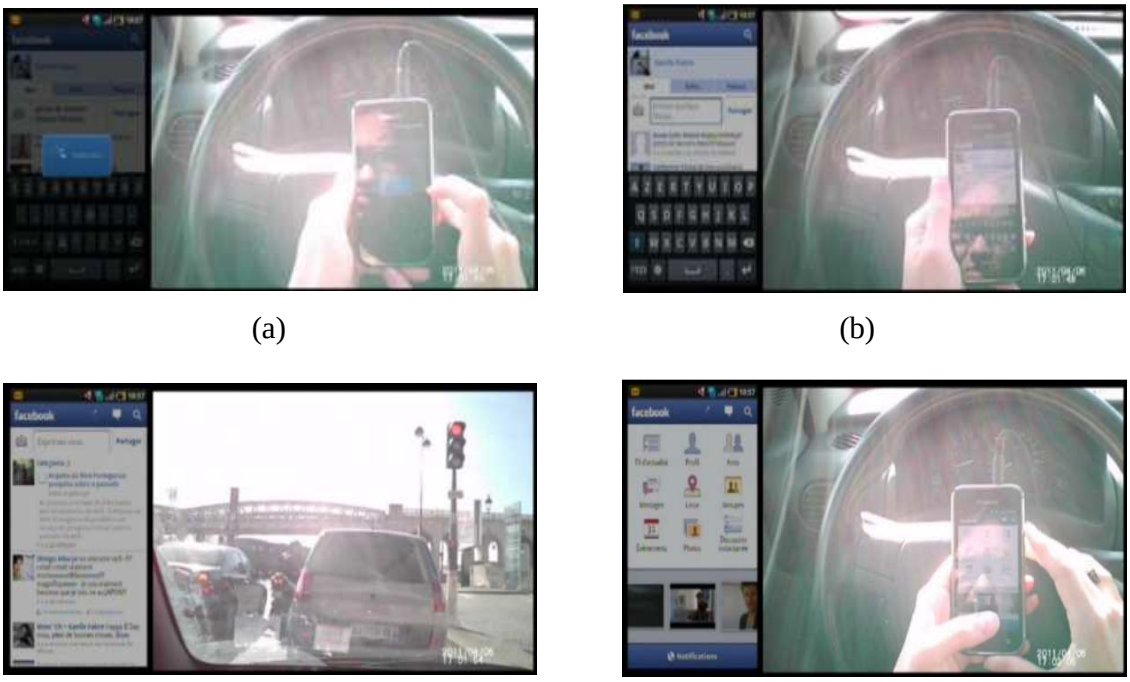


Figure 4: A typical visual display for the circular progress bar.

Such an artifact-mediated temporal patterning of periods of activity and inactivity is not restricted to driving. It is, for example, very common with mobile devices to intervene on the interface, and get a 'circular progress bar icon' (Figure 4), indicating to the user that the system is doing something and that until it is done, it is not responsive to the user's actions. The progress bar does not 'instruct' the mobile

phone user to stop as the red light does because it is grounded in another assemblage of human and material resources. However, like the red light, it projects a pause in the smartphone activity. It works as a ‘prospective indexical’ [8], signaling that some expectable delay is to be experienced before the expected action is accomplished and the interface becomes actionable again. It somehow ‘points’ towards the future moment in time in which this might happen.

For instance, three minutes and two traffic lights after the first one, our mobile Facebook user drives to another red light at which she looks down at her smartphone again. That she has left it on the wheel all along displays her orientation towards multi-activity: it is a way to construct an environment particularly conducive to this, making it easy to gaze swiftly at the mobile terminal and away while driving. At this new red light, she composes and sends a ‘happy birthday’ message to a Facebook friend. When she sends it, she gets the ‘progress circular bar’ icon (Figure 5a and 4).



(a) (b)

(c) (d)

Figure 5: a) Sending a message and getting the circular progress bar; b) the right hand goes to the right to engage a gear; c) the gaze moves up to look at the road and ‘discover’ that the light is still red; d) she immediately gazes down at the smartphone.

Her right hand moves towards the right to engage a gear. She takes that smartphone-related pause as an opportunity to engage a gear, an action the preparatory character of which shows evidence of her re-orientation towards the traffic about to resume. What is significant here is the way expectable pauses in the smartphone activity are also treated as slots to re-direct attention towards driving (even though driving is not yet relevant). Then she looks ahead to the road (Figure 5c) and sees that the light is still red. This shows that her engaging the gear was not related to any traffic event but indeed to the pause projected in the mobile communication activity. Characteristically, finding the traffic still at the stop is taken as an opportunity to gaze down at the smartphone again (Figure 5d), rather than continue on monitoring the road.

Such a pattern of systematic gaze switching at projected pauses in one stream of activity was a recurrent pattern throughout our corpus. When our smartphone users were in the subway (a kind of mobility which is not as attention-consuming as driving), gaze switches away from the mobile phone usually occurred around the moments at which the train was entering into the station or leaving it, whether they were using their smartphones or not. Such a placement probably displays an orientation towards monitoring the progress of the train and the conduct of other passengers coming and going. When they were using their smartphones as well, whenever the circular progress bar appeared, they treated the projected pause as an occasion to gaze away from the phone and to the train environment, whether near a station or not, even if nothing was happening in the wagon at the time.

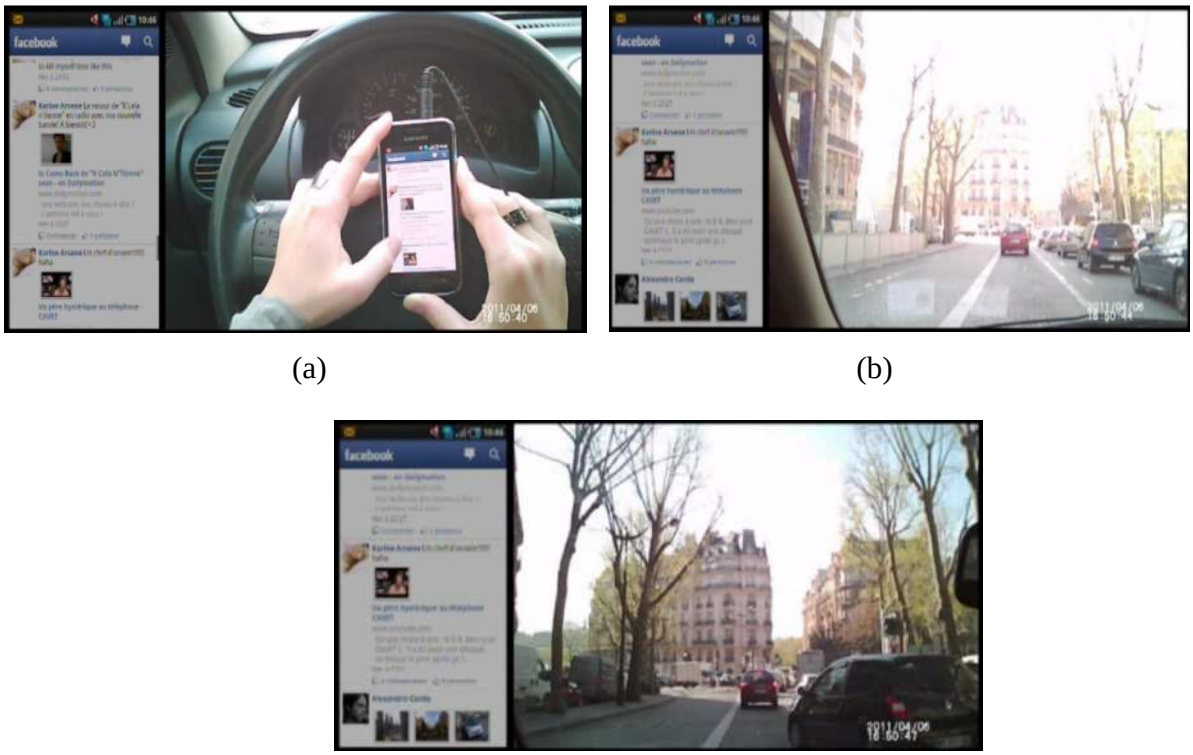
These observations show that the way we recognize unfolding activities as occasioning and projecting moments of activity or inactivity is a key resource in the management of joint activities. More generally it provides us with an original way to empirically grasp what the experience of multi-activity is about. We can thus pin down an elusive concept which is usually understood either in its commonsensical definition of doing two things at the same time (which at a fine-grained level is almost never the case) or as the limiting case of such an extreme fragmentation of tasks that switches between activities are so frequent as to become indistinguishable [14]. With our data, we can say that an orientation towards multi-activity becomes manifest when projectable pauses (and possibly other events) in a given stream of activity are systematically oriented to as opportunities to gaze away towards another domain of the environment, relevant to another activity. In that sense, mobile Facebooking and driving, or even mobile Facebooking and taking the subway, are experienced as multi-activity in our corpus. Conversely, being absorbed or engrossed in an activity can be glossed as a tendency to ignore projected pauses in the ongoing activity and remain focused upon it. Mobile Facebooking on a couch at home may be engrossing in that sense.

Using our recording apparatus to test a user in that situation, we could observe that while events in the environment occurred which might trigger his gazing away from the smartphone, many such potentially relevant transition points were ignored in that respect.

Let us note finally that the orientation towards multi-activity in that sense is usually displayed in the organization of the body and the environment. Leaving the connected smartphone on the wheel when driving, or on one's lap when riding the subway, are ways to create an embodied spatial-material arrangement particularly conducive to swift and frequent gaze switches between the smartphone and the mobility-relevant visual domain.

Temporal mismatches in multiple streams of activity

An orientation towards multi-activity (in our case, mobile communication and transportation) makes the normal 'messiness' of urban environment a constraint and a resource. Urban environments have been described as 'messy' to account for the way they are crisscrossed with heterogeneous socio-technical infrastructures and occupied by urban denizens who have been thrown together there and must behave accountably with respect to one another [1]. Such messiness involves bringing heterogeneity and contingent eventfulness to the everyday urban experience. One consequence of this is that for a user engaged in different activity systems, there is no reason for a relevant transition point in one given activity to occur at the exact moment another stream of activity demands action. Different activities will generally project different temporal expectations and mismatched sequential opportunities. We have seen an example of such routine temporal mismatches in the previous section. The driver at the red light who takes advantage of the pause in mobile screen activity after sending a message (Figure 5b) finds the traffic light still red when she looks up (Figure 5c) so she looks down at her mobile phone again (Figure 5d). Being competent at jointly managing the demands of mobile communication and mobility (and more generally any kind of multi-activity) shows one's capacity to handle and minimize the potential consequences of temporal mismatches, so as to perform all relevant activities reasonably well with respect to the demands of the situation.



(a) (b)

(c)

Figure 6: a) The driver is scanning down her list of Facebook posts. The sudden motion of surrounding cars is detectable in data through the side window though not visible in the picture; b) she eventually looks down, and a large opening is now visible before her car, materializing the delayed character of her response; c) the black car on the right 'jumps' into the gap, before she eventually starts to move her car forward again.

Such phenomena occur at a micro level in the details of the way the ongoing situation unfolds. They often elicit responsive conduct, which is not reflexive and thought about. This is the reason why video recordings of naturally occurring situations such as the ones we are providing here are required to provide better understanding based on observation. For instance, let us return to our mobile Facebook user who had just come to a traffic light and used this as an opportunity to launch Facebook on her smartphone (Figure 3). When the traffic resumes at the green light, her gaze can be seen to remain on her smartphone for a lapse of time and not immediately raised towards the street ahead, even though cars visible in the camera glasses (and even more so in her peripheral vision) have started to move (Figure 6a). That her response is noticeably 'delayed' is not just the analyst's opinion since independent evidence of this can

be found in the data itself: a driver in the next lane takes advantage of the gap in front of a car created by her delayed response in order to change lanes and take her 'slot' in the line (Figures 6b and c).

The materialization of such a delay shows how using a smartphone while driving may have consequences on the driving performance, even if in a minor fashion here. Can we get one step further and provide a more precise interpretation of the reasons for such a delay? This is where the screen capture data comes in useful. At the moment the light turned green, the driver was in the midst of a scan of her list of Facebook posts: she was scrolling down with her finger, from the most recent (at the top) to older ones further down on the touchscreen. It is impossible to say, on the basis of the empirical evidence alone, whether or not she was looking for a specific post (goal-oriented activity) or just waiting for some post on her list to catch her attention (environment-driven activity). Whatever the case, what is significant is that the visual and pragmatic structure of the list does not offer obviously recognizable relevant transition points, the occurrence of which could be taken as an opportunity to gaze away and reorganize one's involvements. One may just scan down on and on until an attention-catching post has been attained, without the list in itself affording any generic and eventful asperity. In the case of a significant event happening outside of the screen, which is at that moment the focus of the visual attention (such as a light turning green), there is a fair chance that the user might remain absorbed in the scanning activity for an extra moment through the sheer inertia of focused attention, even though these 'outside' events are demanding some kind of response. Hence the kind of delay we observed when the traffic actually resumed.

Our research methodology not only allows us to observe the way users manage multiple and temporally heterogeneous involvements (here with respect to mobile communication and mobility), but it also provides us here with another grip on the concept of multi-activity. Situations of multi-activity are situations in which users are attuned to the potential sequential implicativeness of events (as discussed above). Yet they are also situations in which temporal mismatches in the occurrence (or projected occurrences) of meaningful events are expected to occur, to be recognizable as such, and to be consequential with respect to the joint accomplishment of the relevant activities. In our example, the delay in the resumption of driving is materialized by the 'gap' in the line of traffic in front of the observed driver. It is made consequential by the conduct of the driver on her right who immediately 'fills' the space. The expectability of such a sequence is perceptible in the fact it is not noticed or topicalized. It is treated as unremarkable. Even such a routine occurrence provides evidence for the claim that using a smartphone while driving is a form of multi-activity that is potentially consequential in line with social and legal concerns.

We do not claim here that all mismatches and discrepancies in the occurrence of relevance transition points are necessarily detrimental to the accomplishment of the relevant activities. We have mentioned above how at a further traffic light the occurrence of a pause in the mobile communication (the delay related to the sending of a message) provided the occasion for the same driver to engage a gear before the light turned green, thus displaying her orientation towards multi-activity (i.e. systematically treating pauses in one activity as opportunities to switch one's gaze and/or involvements. The accomplishment of such preparatory action is potentially useful and certainly not disturbing with respect to the impending resumption of traffic. The actual meaning and significance of temporal mismatches in the occurrence of relevant transition points cannot therefore be defined *a priori*. Such meaning is constructed locally; it is situated and contingent to the details of the unfolding situation. Being a competent user of Facebook on the move relies on being able to manage the consequences of such potential mismatches as smoothly as possible.

THE 'TEXTURE' OF MOBILE INTERFACES

We have repeatedly seen here how some events in a given stream of activity were treated as sequentially implicative in multi-activity situations, i.e. recognized as potentially relevant transition points, and occasions to gaze away and perhaps reshape one's involvements. That they actually trigger such a response is related to the situation as a whole and is particularly sensitive to the design of the environment. The treatment of the red light-green light change in the above example is sensitive to the finer details of the way the driver is involved in mobile Facebooking, and particularly the way the navigation of the mobile Facebook application might provide (or not) recognizable occasions for disengagement from the mobile screen. We have for instance argued that scanning an (almost) unlimited list of posts afforded few opportunities for gazing away, similar posts following one another in the navigation until one catches the user's attention and is treated as salient. On the other hand, and particularly in multi-activity settings, projected pauses such as the circular progress bar that appears when one sends a message were treated almost systematically as occasions to gaze away, that is, as sequentially implicative.

Navigation-mediated events occurring during the use of mobile applications may be quite subtle. The empirical methodology we propose here proves quite powerful in unveiling some fine-grained interface-mediated embodied conduct in real life situations. For instance, our argument that scanning down a list of Facebook posts with one finger is 'smooth' has to be nuanced when we analyze the data more closely.

The driver at the red light starts with the most recent posts and scans down rather slowly with her left index (Figure 7a). After a few posts, she rearranges her hand so as to scan now with her left thumb, with a larger span scan-down movement (Figure 7b). When we correlate this observation to the screen capture data, we see that this change in the hand arrangement occurs at the moment her finger gets to the transition between today's posts and yesterday's posts. That difference is marked linguistically in the interface design, the categories of which distinguish the posts of the day, labeled as having "arrived X hours (or minutes) ago" and those of the day before, labeled as "arrived yesterday at …" This lexical shift, from hours to days, introduces a minute spatial-semantic discontinuity as one scans down the list of posts. In the course of scanning down the list, the passing down of this spatial boundary occasions a minute transition, a potential temporal event that might be ignored. What our data shows is that in this situation it is not. It is marked as meaningful and noticeable by the change in the scanning finger on the touchscreen.

(a) (b)

Figure 7: a) The user scans down the list of posts with her left index finger (top image) until she reaches the last post of the day and the first one of the day before; b) At this juncture she removes her index and starts to scan with her left thumb.

Such embodied conduct is non-reflexive and finely tuned to the details of the interface through the way design features may occasion meaningful events in situations of use, which may acquire sequential relevance. Making the change in temporal descriptors for the incoming Facebook posts is treated as an "affordance" [6, 7, 13], i.e. as a relational feature which is as much in the interface as in the habituated user, and through the mediation of which some non-reflexive embodied conduct is "afforded". In multi-activity settings, such occurrences may or may not have sequential implicativeness. For instance, while projected pauses are almost always sequentially implicative, the scanning event indexed by the change in

finger is not, at least in the situation described above. It was not treated as an occasion to gaze away. It might have been in other circumstances (for instance if it had occurred at the moment the traffic resumed), for the meaning of such 'events' and their sequential implications depend very much on the moment they occur and their placement within temporal organizations of each relevant activity.

We could describe such affordance-like features as composing the 'pragmatic texture' of interfaces. Interfaces may be described as more or less pragmatically 'rugged' according to the way in which they might occasion such occurrences and perceptible embodied responses to them. In multi-activity settings, we can speak more specifically of the 'sequential texture' of interfaces to account for the way they provide opportunities for sequentially implicative events, that is events that are treated as occasions to change one's focus of attention and involvement. We have shown how the navigation in an interface projects many moments of pause and how it can be described as having a 'rugged sequential' structure. This is not limited to mobile applications. The traffic environment can be considered 'rugged' in the urban environment according to the number of red lights, acting as instructions to drivers to stop for a short lapse of time, one encounters.

The sequential 'ruggedness' of the activity environment plays a considerable role in multi-activity settings, where users are particularly alert and attuned to the potential sequential implications of events occasioned by their actions in all streams of relevant activity. The more 'rugged' the environment, the more occasions it provides for such switches in gaze and attention. The more frequent such opportunities, the smaller and less frequent the possible time discrepancies that will appear contingently, when the projected pauses in one stream of activity are not temporally fitted to the demands in others. For instance, if navigating the mobile interface projects one recognizable pause every second, delays in the resumption of driving at a traffic light change might not exceed one second. What we suggest here is that the frequency of projected pauses defines a rough upper limit on possible response delays with respect to the pressing demands of other engagements.

Our research methodology is therefore particularly important for the analysis of smartphones on the move since it provides empirical evidence at a fine-grained level of the occasions in which, in a naturally occurring user situation, smartphone users will 'respond' to features in the interface design. This retrospectively reveals the 'pragmatic texture' of the interface, i.e. the features in the design that may occasion meaningful events in the course of a given activity, with possible sequential implications. This has significant consequences for design, since design can significantly shape the sequential texture of the interfaces.

IMPLICATIONS FOR DESIGN

First, our research methodology is in itself a powerful resource for design. It allows us to observe very subtle interface-mediated embodied conduct, of which users are mostly unaware and therefore unable to talk about in surveys. Most crucially, such data can be obtained in real life situations with smartphone users on the move. It can thus inform designers on the way such interface-mediated embodied conduct may be sensitive to the actual situation of use.

Secondly, our observations have highlighted the importance of the temporal organization of activities and their sequential implicativeness (that is, their offering potentially relevant transition points to deal with something else) for jointly managing mobile communication and mobility, and more generally in situations of multi-activity. We have proposed to call this the sequential texture of activities, that can be deeply altered by design. In current versions, the Facebook wall of posts has a relatively smooth pragmatic structure with respect to finger-mediated scanning on a touch screen (with the exception of the subtle 'bump' made by the transition from the day's posts to those of the day before). One may scan from the top to the bottom almost continuously with the same kind of gesture. One could imagine designing the same function with a more sequentially discontinuous feel. For instance, messages could arrive by chunks of two or three on the touch screen and an icon would be pressed to get the next batch to appear. In this second version, a more 'rugged' environment would be provided for scanning down the list of posts, which would offer many more potential transition points to gaze away or direct one's attention towards another domain of activity while scanning the list.

This opens up the possibility of designing specifically for mobile settings and the kind of multi-tasking which occurs then, not only by adapting the look and feel of the interface to the rhythm of the activity [12], but also with respect to the sequential properties of the interface, i.e. to try to increase the number of potentially relevant transition points in the accomplishment of a given activity. For instance, it might be imagined that the 'rugged' version of the list interface is safer at a red light, since it would minimize the potential temporal mismatch between the green light and the resumption of traffic. However, mobile design better adapted to the mobility context would probably also come at the expense of user friendliness (more likely to be associated with the design of pragmatically 'smooth' interfaces). So 'ruggedness' should be reserved for specific situations. Pervasive smart environments might offer a way to get the best of both worlds. For instance, in a smart traffic environment, the smartphone application might be enabled to detect the demands of the surrounding traffic situation (such as the current state of the closest traffic

light) so that some form of transition point could be ‘created’ in the smartphone activity when such demands become pressing. Or there could be ‘smooth’ and ‘rugged’ versions of the mobile interface simultaneously available, the system switching to the second when conditions require it. In this way one might combine a concern for safety and adequacy with respect to the mobility context with a concern for user-friendly design, which is bound to dominate in more ‘ordinary’ and less attention-demanding situations.

We have also shown how projected pauses in a stream of activity tended to be systematically treated as opportunities to gaze away in multi-activity settings and therefore constituted important resources in the management of joint activities. Yet environmental cues that project pauses may vary with respect to the degree of information they project about the pauses they index. Red lights index a pause in traffic, but it is only habituation that tells drivers how long this pause might be. Circular progress bars index a pause but do not let the viewer anticipate their length. This differs from linear progress bars where a cursor moves at a certain speed towards its extremity. More subtlety can be introduced in the design of progress bars to refine the temporal anticipations of their audience [10]. Based on our findings, cues designed to let the viewer better anticipate the end of a pause projected in a stream of activity are better suited to multi-activity settings.

CONCLUSIONS

To solve the complex problem of observing the uses of mobile communication on the move, we have developed a portable recording apparatus, based on the combination of light camera glasses (worn by the user) and the capture and recording of his mobile terminal activity as it appears on the mobile device’s screen. We have shown how such a recording apparatus allowed the observation of a key phenomenon, i.e gaze switches to and away from the mobile phone (and how these are timed with respect to other perceptible events in the situation).

We have tested this research methodology on a sample of mobile users commuting between home and work. The observation of gaze switches appears to be a powerful resource in understanding how users try to manage and juggle the contingent demands of mobile communication and mobility. Based on the way users treated projected pauses in the mobile communication or transport activity, we have been able to propose an original and user-centered definition of what it means for a person to be engaged in a situation

of multi-activity: an orientation towards systematically treating projected pauses in one stream of activity as an opportunity to re-direct one's gaze and attention towards another stream of activity and vice and versa. In that respect, mobile communication in a transport situation is definitely experienced as a situation of multi-activity.

Events triggering a gaze switch in this way can be described as being sequentially implicative: they are treated as possible slots for reshaping one's attention and involvements in a sequential way. However, their occurrence (or non-occurrence) is conditioned by what happens in one stream of activity and not necessarily adjusted to the timing and demands of the other relevant activities. Temporal mismatches occur all the time. We have discussed one example in depth: a driver scanning down her list of incoming Facebook posts (which did not afford recognizable transition points) recognizably 'lagged' when the light turned green. The management of such temporal mismatches is a central feature in the competent handling of multi-activity situations.

The way in which navigating activity environments projects possible transition-relevant points, and more generally any kind of event, therefore plays an important part in the management of multi-activity situations. Examining the way our driver was navigating her list of Facebook posts in more detail, we have been able to show more subtle phenomena, such as a change in the scanning finger from the thumb to the index, which did not appear to be oriented-to as sequentially implicative, in the sense that it was not seized by the user as a possible transition point for gazing away from the mobile phone to monitor the environment.

We have eventually introduced two different notions of texture to capture the propensity of interfaces to generate or project response-eliciting occurrences in the course of their use. The 'pragmatic texture' of an interface describes the way the interface design will more or less afford interface-mediated events and responses as it is navigated. The 'sequential texture' accounts for the propensity of the interface to generate events with sequential implicativeness. An interface can be described as 'rugged' (vs. smooth) when it frequently (vs. rarely) generates events of a given type. Textures are relational in the sense that, like affordances, they are as much a feature of the technology design as something relative to a user or a community of users. For instance, in multi-activity situations, users display a heightened awareness of the sequential 'ruggedness' of their environments. Designing specifically for multi-activity environments where safety concerns are critical might therefore involve increasing the 'ruggedness' of the mobile interfaces to augment the frequency of possible transition points and minimize temporal mismatches in

the demands of the varied activities the user is engaged in, even if such a design rationale might run against more conventional design strategies centered on user-friendliness.

REFERENCES

1. Bell, G., and Dourish, P. Yesterday's tomorrows: notes on ubiquitous computing's dominant vision. *Personal & Ubiquitous Computing 11, 2* (2007), 133-143.
2. Brown, B., and Laurier, E. The normal natural troubles of driving with GPS. In *Proc. CHI 2012*, ACM Press (2012), 1621-1630.
3. Drewes, H., De Luca, A., and Schmidt, A. Eye-gaze interaction for mobile phones. In *Proc. Mobility 2007*, ACM Press (2007), 364-371.
4. Esbjornsson, M., Juhlin, O., and Weilenmann, A. Drivers Using Mobile Phones in traffic: An Ethnographic Study of Interactional Adaptation. *International Journal of Human Computer Interaction 22, 1-2* (2007), 37-58.
5. Figeac, J. "The Appropriation of Mobile TV through Television Preferences and Communications Networks". In Martin, C. and Von Pape, T., *Images in Mobile Communication. New content, new uses, new perspectives?* Springer VS Verlag (2012), 121-141.
6. Gaver, W.W. Technology affordances. In *Proc. CHI 1991*, ACM Press (1991), 79-84.
7. Gibson, J. J. *The ecological approach to visual perception*. Boston: Houghton-Mifflin (1979).
8. Goodwin, C. Transparent vision. In Ochs, E., Schegloff, E. A., and Thompson S. A. (Eds.), *Interaction and grammar* (1996), 370–404, Cambridge University Press.
9. Haddington, P., and Rauniomaa, M. Technologies, Multitasking and Driving: Attending to and Preparing for a Mobile Phone Conversation in a Car. *Human Communication Research 37* (2011), 223-254.
10. Harrison, C., Amento, B., Kuznetsov, S., and Bell, R. Rethinking the progress bar. In *Proc. of the ACM Symposium on User interface Software and Technology*. ACM, Press (2007), 115-118.

11. Isomorsu, M., Marika, T. Soili, V., and Kuutti, K. Experimental evaluation of five methods for collecting emotions in field settings with mobile applications. *International Journal of Human-Computer Studies 65, 4* (2007), 404-418.

12. Kane, S.K.**,** Wobbrock, J.O. and Smith, I.E. Getting off the treadmill: Evaluating walking user interfaces for mobile devices in public spaces. *Proceedings of MobileHCI '08.* ACM Press (2008), .

13. Kaptelinin, V., and Nardi, B. Affordances in HCI: toward a mediated action perspective. In *Proc. CHI 2012,* ACM Press (2012), 967-976.

14. Kenyon, S. What do we mean by multitasking? Exploring the need for methodological clarification in time use research. *International Journal of Time Use Research* ***7,*** *1* (2010), 42-60.

15. Kjeldskov, J. Skov, M.B., Als, B.S., and Hoegh, R.T. Is it worth the hassle? Exploring the added value of evaluating the usability of context-aware mobile systems in the field. In *Proc. Mobile HCI 2004*, ACM Press (2004), 61-73.

16. Lahlou, S. L'activité du point de vue de l'acteur et la question de l'inter-subjectivité : huit années d'expériences avec des caméras miniaturisées fixées au front des acteurs (subcam). *Communications 80* (2006), 209–234.

17. Lahlou, S. How can we capture the subject's perspective? An evidence-based approach for the social scientist. *Social Science Information 50, 3-4* (2011), 607-655.

18. Latour, B. *Pandora's hope. Essays on the reality of science studies*. Cambridge: Harvard University Press (1999).

19. Laurier, E, Lorimer, H., Brown, B., Jones, O., Juhlin, O., Noble, A., Perry, M., Pica, D., Sormani, P., Strebel, I., Swan, L, Taylor, A., Watts, L. and Weilenmann, A. Driving and passengering. Notes on the ordinary organization of car travel. *Mobilities, 3, 1* (2008), 3-23.

20. Lee, I., Kim, J., and Kim, J. Use contexts for the mobile internet: A longitudinal study monitoring actual use of mobile internet services. *International Journal of human-Computer Interaction 18, 3* (2005), 269-292.

21. Mark, G., Christensen, U., and Shafae, M. A methodology using a microcamera for studying mobile IT usage and person mobility, *CHI 2001,* Workshop on Mobile Communication: Understanding Users, Adoption and Design, (2001).

22. Mondada, L. 'Talking and driving'. Multiactivity in the car. *Semiotica 191,* (2012), 223-256.

23. Morel, J. and Licoppe, C. “Studying mobile video telephony”. In Büscher, M., Urry, J. and Witchger K. “*Mobile Methods*”, Routledge (2011), 164-182.

24. Oh, Y.S., and Lee, K.P. The development of user Observation method and analysis tool in Mobile environment - With emphasis on the development of observation method at user's point of view by wearable micro-camera and analysis tool. http://dpl.kaist.ac.kr/web_wiki/images/1/18/Ohys2005.pdf

25. Omodei, M. M., McLennan, J., and Whitford, P. Improving decision making in complex natural settings: Using a head-mounted video camera to improve performance of competitive orienteers. *International Journal of Sport Psychology 29, 2* (1998), 115-131.

26. Oulasvirta, A., Tamminen, S., Roto, V. and Kuolahti, J. Interaction in 4-Second Bursts: The Fragmented Nature of Attentional Resources in Mobile HCI. In *Proc. Of CHI 2005*, ACM Press (2005), 919-928.

27. Pink, S. *Visual Anthropology*. London: Sage (2007).

28. Relieu, M. Mobile phone “work”: disengaging and engaging mobile phone activities with concurrent activities. In R., Ling, and S., Campbell (Eds.). *The reconstruction of space and time: Mobile communication practices.* New Brunswick (NJ): Transaction Publishers (2009).

29. Roto, V., Oulasvirta, A., Haikarainen, T., Kuorelahti, J., Lehmuskallio, H., and Nyyssönen, T. Examining Mobile Phone Use in the Wild with Quasi-Experimentation. In *HIIT Technical reports* (2004).

30. Tamminen, S., Oulasvirta, A., Toiskallo, K., and Kankainen, A. Understanding mobile contexts. *Personal and Ubiquitous Computing* (2004) 8: 135-143.

31. Tossell, C., Kortum, P., Rahmati, A., Shepard, C., and Lin, Z. Characterizing web use on smartphones. In *Proc. CHI 2012,* ACM Press (2012), 2769-2778.

32. Yilmaz, A., Javed, O., and Shah, M. Object tracking: A survey. *ACM Computing Surveys* 38, 4 (2006), 1-45.

33. Zouinar, M., Relieu, M., Salembier, P., and Calvet, G. Observation and capture of multimodal interaction in mobile situations. In *Proc.UbiMob 2004*, ACM Press (2004), 5-8.