From *in the wild* to *in vivo:* Video Analysis of Mobile Device Use

Donald McMillan, Moira McGregor, Barry Brown
Mobile Life Centre
Stockholm University, Sweden
{Don, Moira, Barry}@mobilelifecentre.org

ABSTRACT
The explosion of mobile applications and services presents challenges for evaluation and user study. One successful approach has been to deploy instrumented applications, logging their use over long periods of time. We present an expansion of this by remotely recording video and audio of use, while also capturing device and app context. *In vivo* combines five data collection techniques – *screen recording, ambient audio recording, wearable cameras, data logging* and *distributed remote uploads*. This data provides a range of insights and we discuss examples from previous work which reveal interaction design issues where interface confusions or task mismatches occur. We see how apps are integrated into ongoing activity and environment (such as how maps are used *in situ*), and how recorded conversations around and about apps may be used for evaluation purposes. We conclude by arguing that this combinative method helps us to move from considering app use in isolation, to studying app use in interaction.

Author Keywords
Screen Recording; Evaluation; Logging; Video Analysis

ACM Classification Keywords
H.5.2. User Interfaces: Evaluation/methodology

INTRODUCTION

A recent paper by Reyal, Zhai and Kristensson [37] presents an interesting challenge to lab studies of mobile device use. It presents results from two studies of mobile text entry – one conducted in a traditional lab study setup, and the second based on an 'in the wild' experience sampling test. Their first test clearly shows that a standard 'tap' keyboard is faster and more accurate for text entry. Yet their second, the 'in the wild' experiment, shows the opposite result – that a gesture style swipe keyboard is faster and more accurate. With the 'in the wild' method being closer to real world text entry, this paper casts doubt on the validity of laboratory results for this task.

Moving beyond the user interface, evaluating complex, social applications on mobile phones presents even more serious challenges. Since much device and app use is triggered by the situation and environment of use – conducting meaningful lab-based evaluation has become difficult. For example, while using a navigation application a user may make use of their location, but also the broader situation, including road conditions, public transport information, or street configuration. Similarly, a mobile retail payment relies upon interaction with the customer *in* a store–and the store environment can impact how a mobile payment solution is used. As Robinson, Marsden and Jones [38] remind us, what goes on *around* devices is as relevant to its use as what goes on *with* devices. A variety of methods have been used to record and study the context of use, usually relying upon recollection – e.g. interviews [35], diary studies [10] and surveys [44]. Researchers are also increasingly using instrumented applications to collect log data on device and app use [32].

Following the call for new field methods in MobileHCI [24] this paper explores combining methods that record video and audio from participants' devices, with optionally having participants wear lightweight cameras to record activity. We discuss combining the longstanding tradition of work that uses video recordings of naturalistic activity, with the opportunities presented by distributed remote devices uploading data, without researcher involvement. We deploy five complementary recording techniques: *screen recording* allows for detailed understanding of interface, device and multiple application use; *audio recordings* taken from mobile devices during use, provides access to the talk and conversation in which the device use is embedded; *data logging* automatically captures location, time, sensors and device attributes such as applications launches; *wearable cameras* facilitate filming environmental details; and lastly, *remote uploads* lets participants record and upload data without direct intervention from researchers.

Packaged under the title *in vivo*, we propose that this combinative method supports evaluation of new applications and prototypes in use, in real world settings. It makes user behaviour easier to re-construct and understand because it provides contextual data, such as conversation

MobileHCI '15, August 25 - 28, 2015, Copenhagen, Denmark

ACM 978-1-4503-3652-9/15/08…$15.00
DOI: http://dx.doi.org/10.1145/2785830.2785883

and location. This leads to more grounded re-design lessons. Moreover, this method also helps change, as Rogers puts it [39], "the unit of analysis" – it presents mobile device use as something embedded within sequences of everyday activity that span across digital media, physical media, interaction, and activity [7]. As we show below, one crucial part of this is capturing talk and interaction between co-present users. The time consuming nature of video analysis, however, does presents significant challenges to using this *in vivo* method.

The paper begins by contrasting two different approaches to evaluation; using automatic device logging, and interactional analysis. We then briefly present two studies in which we deployed different configurations of recording technology–after which we discuss issues in data collection, preparation, analysis, and presentation involved. In particular, we focus on the trade-off involved in data collection between collecting richer data in more constrained situations with use of additional hardware, versus the discrete deployment of software on phones to record less of the surrounding situation of use but for longer periods, and less obtrusively. From this we move on to discuss some of the issues involved in analysing video, data, triangulation and summary techniques, and cases where smaller scale recording is more appropriate. Lastly, we explore the privacy implications of recording interactions between study participants.

BACKGROUND

Log Based Studies

Interest has grown in finding ways of studying mobile device use through log and sensor data, deploying instrumented apps to app stores. McMillan et al [32], Böhmer et al [3], and Henze et al [20] led the way in the large scale collection of log data on device use, an approach that allows for a broad understanding of user behaviour, particularly discrete user behaviours such as app launches, switches and typing. Online services, such as mobile search, also provide an avenue for collecting data on mobile device use, for example Kamvar's work [23]. These approaches have generated powerful insights into device use and user behaviour arising from the intersection of time, local environment and sequence of application launches combined together. Deployment of instrumented applications through global app stores allow for a wide range of data to be collected from users recruited from a diverse population. Moreover, these apps make use of remote upload of data to a server from users' own device meaning the data may be collected relatively unobtrusively.

Although automated logging can produce powerful data sets, log data has its limitations. Usage data is stripped of much contextual data, and this can make analysis of situated use difficult to achieve [32]. While location and on-board sensors can collect revealing data, understanding complex or social situations is challenging. For example, being a car passenger or car driver when using a device might appear similar in a system log but with obvious consequences for each situation of use.

This has led some to argue for a mixed methods approach – in particular for combining log data with participant interviews to gain access to the context of use and also user attitude and experience [34]. Experience sampling has attempted to overcome this by asking users for responses at random intervals. Two pioneering systems include MyExperience [15] and Momento [9]. These systems use instrumentation and user alerts on mobile devices to collect data of use and broader user interaction. However, these approaches have not been widely used for the study of modern mobile devices. An exception is the experience sampling work on mobile search by Church and collaborators[11], and continued more recently in work on micro-app usage by Ferreira et al [14]. An alternative way of dealing with these issues forwarded by Morrison et al, is to make use of a local cohort of users alongside a large scale log based deployment [34]. In this work a local cohort was interviewed and studied more intensively. This allowed researchers to use individual cases to investigate and understand the broader patterned themes observed in the large scale data – using the "small to understand the large." In turn, Morrison et al also discuss using the "large to understand the small", where particular behaviour may be identified in the local cohort can be classified as typical or unusual by reference to the larger data.

Interactional Video Analysis

In contrast, interactional video analysis is a distinctive approach to studying technology use that has proven particularly popular within the field of CSCW (computer supported collaborative work). In this research tradition [19, 21, 29] situated cameras, in settings such as control rooms, surgeries, homes, offices, and museums, are used to capture technology use *in-situ*. These videos support analysis of the details of interaction – looking at how users bring together physical and technical resources in artful and complex ways. Interactional video analysis draws on conversation analysis and ethnomethodology with its focus on fine-grained analysis of interaction and activity.

One example of the application of interaction analysis has been in workplace studies [19], looking at the how the sequence of activity and interaction, such as gaze and talk, are consequential. Mobility has been one concern of this work, as in Luff and Heath [29] on mobile interaction, developing the concept of 'micro mobility': the ways in which information – particularly paper documents – come to be deployed in face to face interactional settings. Critically, video forces attention on the moment-by-moment production of technologically mediated action. Analytically, interactional video analysis pays close attention to a small number of incidents, rather than attempting to present general findings about use. Following conversation analysis, the focus is on trying to understand

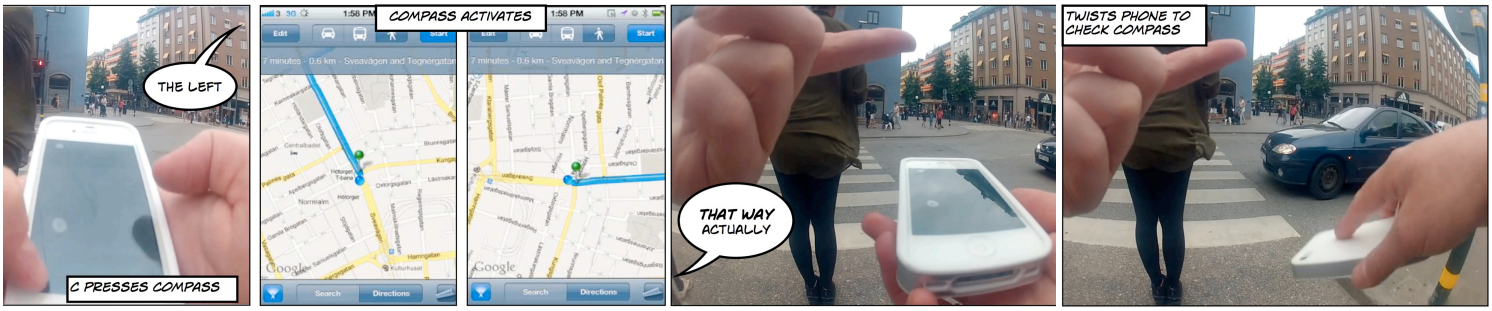


Figure 1: Navigating with the compass

'why that here': how sense is made of a situation through the use of talk, gesture and interaction. Within HCI, interaction analysis has been applied to study non-work situations such as research on driving with GPS [4] using cameras mounted in cars to record drivers' navigation.

Remote Recording of Device Use

How might we combine these two diverse methods? Interaction analysis provides a way of understanding the rich details of technology *in situ*, whereas log analysis provides the ability to collect device data *in situ*. While they are different, both approaches analyse data collected from actual system use. A recent technique, growing in popularity, is to combine these methods and remotely record videos of user behaviour and screen interaction. In [5] we discuss the use of screen recording on mobile devices combined with wearable cameras. We later extended this technique, using the collection of video data of mobile device use recorded via the device itself [6]. Licoppe studied mobile interaction using camera phone glasses [27] combined with video equipment connected to mobile phones. Kjeldskov [24] also experimented with camera on stalks pointed at mobile phone screens. Lastly, while not used specifically for evaluation the Sensecam research made use of wearable cameras [41] to record experiences. Commercial applications of video recording in mobile usability work are also growing in popularity, with recording applications such as Lookback, Appsee and Usertesting, and such systems being used in the design of popular apps such as Airbnb and Spotify.

IN VIVO

We have conducted two separate studies in which we made use of these video techniques to examine mobile device use [5, 6]. Both these studies involved recording individual and group iPhone use. The resulting video provided access to the social life around mobile devices – aspects of interaction with the device, with the local environment, and with those around. The first study used screen recording, wearable cameras and audio recording, with twelve groups of participants on a city daytrip. For the second study, wearable cameras were omitted and audio recording on the phone was optimised, and the software was deployed for a much longer period of time – on average seven days for each participant. These studies highlight the trade-offs in the approaches used. Let us present an example from each of the two studies. This will allow us to show the different analysis opportunities afforded by the different recording setups. We will also present the video in two different ways to give the reader a side-by-side comparison of two possible styles. Both of these examples use interaction analysis.

Study 1: Using Maps – Location Based Applications

The most obvious location-based application available on smartphones are GPS enabled Map applications used for route-finding. We recorded twelve 2-3 hour sessions in which we instrumented the phone with screen recording software, and also furnished each participant with small, wearable cameras. The research team interacted directly with participants during deployment at the beginning of each session, and again during collection of the wearable camera and mobile software at the end. The participants were then given the task of choosing, navigating to, and visiting the tourist attractions of their choice for the rest of the afternoon, so for much of the data collection participants were on a city visit (using public transport, visiting local attractions, dining and so on) alone or in pairs. For this study we used *screen recording, ambient audio recording,* and *wearable cameras.*

Figure 1 is a simple clip in which a participant and friend talk around the iPhone's map app trying to work out which direction they should walk to follow the route on the map, to a restaurant. In the graphical transcript we have included the view from the phone user's camera, and the view from the iPhone screen. In this clip, while trying to find the road that corresponds to their route – they activate the compass on the phone and use this to work out that they need to walk to their right rather than left as they originally thought.

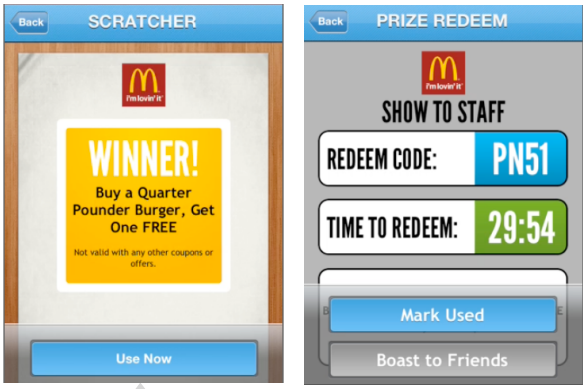
Understanding maps on a phone can be a real challenge to users, and this recording affords us a new perspective on this common device use. Importantly, we can see how phones are not just devices interacted with, but interacted around. Map use is 'environmentally coupled' – the participants have to align the map on the phone to the streets they see in front of them to get to their destination. During the use of the app the compass feature is activated causing the map to spin round, leading to an unexpected realignment of the route from left to right. This suggests that some indication of the maps orientation prior to, or during the activation of the compass feature, may help the user. We can judge by the twisting of the phone to align it with the street ahead that the compass on the phone might be unreliable in its use.

The wearable cameras provide access to gestures around the phone. In this case, the extra video angle shows how, when and where the user interacts with the phone in hand to align the visual interface of maps to the physical landscape. The participant's talk contains much in the way of indexical phrases (references to the environment) "left", "this way", terms which would be much harder to identify without the video of the hand to see, for example, that the participant is pointing at their phone to refer to different roads and streets. In general, the interactions that happen with and because of objects in the environment and those that are determined on, or determine, gestures and physical expressions around the device are those which are uniquely captured and recorded for analysis using this method.

Study 2: Using a coupon app at the point of sale

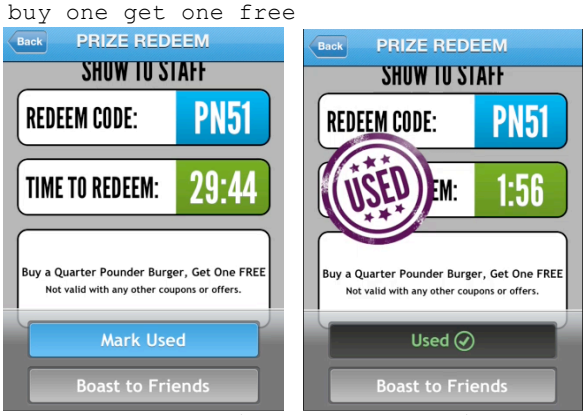
For our second example we will look at a clip which shows how this method can be used to better understand the flow of social interactions that the application, in this case a coupon app, necessitates. The second clip is taken from a study in which we deployed recording software on the phones of 15 participants to record videos of all their interactions with the screen, all the ambient audio (such as conversations around the phone) while they were using the phone, alongside contextual information such as where, when and with what applications the phone was used. Eight of these participants were recruited remotely using social media and Mechanical Turk, and the researchers never physically met them. The participants were asked to use their phones as normal throughout the week and in return they were paid for each full day of use they uploaded. This gave us a geographically and demographically diverse user group. Videos were automatically uploaded to a website, which allowed users to hide videos that they preferred not to share, and also for us to collect diary entries from users describing what they were doing when each clip was recorded. With this approach there were no video recordings of the participants themselves nor their physical surrounds. As such we collected data using *screen recording, ambient audio recording, data logging* and *distributed remote uploads*.

The clip in figure two shows a more conventional transcript of the conversation that unfolds, alongside 4 screenshots taken from the video, as the user employs a mobile payment method *in situ*. We have access to location (a McDonalds restaurant), a video of the screen and the audio. This clip is an example of an app-based mobile payment solution, including the multi-party interaction that is necessary for this method of payment to be successful. The audio is enough to start analysing some of the interaction around making payment with the device and to raise issues with interface design and larger issues around sharing devices. The first hesitation in this clip, line 3, points to a problem in the copy, design and layout of the coupon interface. Here we can see the participant struggle to find the appropriate shared language required to talk about this app-based



Music and noise of people.
Snippets of conversations can be heard.
A activates the coupon

B: Hi can I help you there sir
A: Yeah uhm I have a coupon for ahh [0.6] ah buy one get one free



A: quarter pounder
B: ↑o↓k [1.5]
A: is everything ok
B: yes hit marked use for me
A: ok
A: ok↑
B: Awsooome and whaicht eequarter pounders did you want ((those))|
A: |em
A: [1.1] ok ahm [1.4]
A: just quarter pounders I assume right
B: Yup
A: ok em and ill have a large ah coke
B: anything else↑
A: that will be all

[1.0] indicates a 1 second pause

Figure 2: Mobile Payment with a Coupon App

discount code. The nouns 'coupon, code, scratcher, prize' are all used in various places in this app to describe what the user has received. Settling on a common language on the interface could simplify this interaction

As the participant makes use of his mobile phone coupon, he displays a little hesitation while he works out how to inform the casher of this: "I have a coupon for ah: [0.6] ah buy one get one free quarter pounder". Presumably the issue here is that as he selects, 'Use Now', the details of what he is entitled to when using the coupon disappears

from the screen – as he pauses to examine the screen the white box behind the 'Mark Used / Boast to Friends' panel is scrolled into view to display the details of the coupon. Only when the details of the coupon are available to verify his request, does he complete his sentence and show the screen to the cashier. We can understand how the important information is the validity and the code the cashier should enter which, we can assume, would then inform him on the point of sale system of the details of the offer. However the obfuscation of information for the consumer – what the coupon entitles them to – has slowed the transaction down.

Without a situational video recording we do not have access to some important aspects here of the physical interaction involved in showing the code and then verification that the 'Mark Used' button has been tapped. Even without video access to the setting we can see the ways in which the phone is not just used by the owner, but becomes part of his interaction with the server. Presumably the server sees the phone (and the coupon app) before line 8, as he instructs the participant how to void the voucher and is either shown the device again or continues watching the screen until the coupon is marked as used in the app, to which he exclaims, 'Awsooome'. One limitation of this study is that without another camera angle or sensor readings we do not know definitively, recording the tilt sensor here would have give us some insight into this. This said, the clip shows a number of possible issues with the app: inconsistent terminology, the role of different information screens at different times, and the hiding of the voucher when "use now" animation is touched. We might also note the hesitations in the talk – the repeat of the question about quarter pounders, and the "is everything ok", which suggest some lack of alignment in conversation between the coupon user and the server

DEFINING *IN VIVO*

These examples illustrate some of the opportunities in using remote video and application logging to capture mobile device use data. As is often the case with empirical work, it was clear that the two studies feature different key trade-offs. In the first study the distinctive aspect was the deployment of wearable cameras and the participants being introduced to the technology by a researcher, in person. This set up potentially constrained the data recorded, resulting in less variety in the iPhone use collected. Alternately, in the second study we have a much greater diversity of data – with situations recorded that were never anticipated. Yet here we had no contact with those being recorded, and the recordings themselves were only of the device screen. Thus we had little sense of the broader context, or the micro-mobility of the device at key points. So while study 1 yielded higher quality data in some respects, the scope of the study was reduced – on the other hand, study 2 collected a greater diversity of use, yet much of it could not be reliably analysed. Clearly, there is not a sense that one size fits all for those using these approaches, but rather that choices in method design will result in different trade-offs which will impact results in different ways. By reflecting on the experiences of **data collection**, **analysis**, and **presentation** of the results for each of these studies we can start to consider this more broadly as a combinative method for Mobile HCI which may be adapted to satisfy research goals.

Data Collection

We have deployed five different data collection techniques, which can be combined in different ways: screen recording, remote uploads, recording environmental audio, data logging and, lastly, wearable and mobile cameras:

1. Screen Recording

One of the most distinctive parts of the *in vivo* method is the recording of the mobile device screen. Screen recording records videos of all onscreen interaction, turning phone use into video streams of activity. Recording the screen has been employed in both mobile [31] and stationary contexts [40]. It can be done by attaching small, external cameras to mobile devices to record the screen and the interaction [24]. It can also be done using screen capture software [45] within a single application for purposes of user evaluation, bug reporting, and feedback, or system wide. A number of commercially available applications for screen recording are now available, notably Lookback for iOS, and SCR screen recorder for Android.

Recording the screen solves the problem of interaction being occluded by the user or other objects when recording interaction using fixed or handheld cameras. In mobile devices where the form factor lends itself, and is expressly designed for, a single user it can be difficult for cameras to capture both interaction and what is being displayed. Yet this also misses important aspects of the interaction external to the screen – such as gestures over the screen, user gaze and vision, the position of devices in space, or the start of on screen interaction through gestures and other physical movement. With the increase in local storage, data transmission rates and processing power the opportunity to record more widely and with less disruption to the use and the normal practices of the participants presents itself. This data support the analysis of micro-level interactions within and between applications on a device, and it also supports a focus more on usability problems [12] and more generally for UX testing [36].

2. Remote deployment and uploads

Data collection involves not only recording users in some way, but also deploying equipment and retrieving the data. This can seem a rather prosaic part of data collection, but its subtleties can impact on the data that is collected and also the sorts of device use recorded. Often in research, where there is an experimental prototype, investigators need to be involved to 'manage' the devices in some way. Tolme et al [47] relate the story of a party thrown by participants for the

designers of a particular device being deployed in a study. The management of relationships with users is not a case of 'trying not to interfere', since any data that is collected is shaped by its collection process, but recognising that different relationships can result in different data [47].

We can exploit the widespread use of global app stores to deploy recording software to users, either in its own right or embedded within an app with another purpose (with proper permissions received from users before recording). This provides for an almost completely remote data collection and retrieval setup, with users downloading software under their own motivation, and agreeing to be recorded without any special recruitment. Alternatively, smaller groups of user can be recruited, using in-app advertisements.

While this does entail some careful design, it opens up the possibility of studies where data can be collected with little involvement by the investigator. This has been used to collect log data on application use and can also be used to support remote recording and upload of audio and video data from mobile devices. This potentially allows for collecting data from users who have only the lightest of contact with researchers. Moreover, it is straightforward to collect data from a more geographically diverse collection of users, as setup and collection is done remotely.

3. Recording Environmental Audio

As technology is integrated into our lives it becomes something that we talk around. This is clearly the case with mobile devices, where devices might be brought out to search and settle disputes in conversation, or to share a photograph. Talk surrounding device interaction has been studied in great detail, drawing on conversation analysis to better understand not only the mechanics of interaction but how talk fits more broadly into activity [7, 27]. Recordings of talk around a device can be used to see how a device is included in on-going activity, such as in our first example above, or where the action is done by talk and the phone in combination, such as example two.

Phones themselves can be used as audio data collection devices. This can be limited to when a device is in use, but recordings can also be made in the background when a device is not in use. This provides the possibility for mobile devices to collect audio broadly from a much wider range of different contexts than might be practical with a fixed microphone set-up. Alongside speech, audio recordings can also capture ambient audio, this can also provide insight into the actions of the users as any number of changes in situation can be accompanied by a change in the nature of the audio that would be recorded – the cessation of engine noise can indicate that the user has arrived at their destination, and the 'ding' of an arriving elevator can precede a loss of connectivity and the start of actions to mitigate its impact on use.

4. Data logging

With the increasing number of sensors available on mobile devices, logging those sensors can provide useful additional information for understanding use. Along with sensing the environment, system functions can also be recorded. As we discussed this has been an established part of 'in the large' style studies, where system events such as web requests or applications launched are logged. In combination with other approaches here, however, such information can be used to index and contextualise other recording streams. For example, GPS position can, with a little work, be mapped onto semantic categories – such as work and home [1], but also giving a more specific context to an interaction – such as above with our example of phone use in a McDonalds restaurant. Recent work has also suggested how sensors could be used to detect how a device is being held [49] and infer the hand position of a device given the nature of the touches registered on the screen [33]. Yet currently this is not able to provide the same window on the bodily actions of the user and their interactions with others that a camera can provide.

5. Wearable Cameras

Wearable cameras have been used as integral parts of support systems to help the visually impaired to read on the go [16], to help people track what they eat [46], and to help people be more active [25]. They have also been used to understand how users can remember events [8]. In using wearable cameras, as opposed to following cameras or researchers observing *in situ*, the level of scrutiny felt by the third parties in any situation can be reduced. Some studies have been made of researchers following participants with cameras, such as in [2]. This obviously causes some straightforward influence on the participants, in that having a 'following' researcher can hinder the activity being undertaken. This would be even more apparent if the system or phenomena under examination was one in which the participant was expected to interact with people. While the participant may have an understanding and a degree of comfort with the researcher following them, others may not. One advantage of using a 'following' camera is that researchers have more control over what is recorded. With wearable cameras a researcher only has limited influence over the angle and orientation of the camera. This said, with wearable cameras considerable information about the context of use can still be collected.

This reduction of influence is of course dependent on the make, model and mounting of the device on the participant. In the first study discussed above we made use of 'map bags' worn around participants' necks. One key issue with different recording technologies we tried was the quality of recording when the camera is moving quickly. Cameras designed specifically for action recordings produced much more legible recordings. Licoppe experimented with recordings made from cameras embedded in glasses, and while this provided an unobtrusive perspective, there were

issues with the fast moving scenes recorded [27]. If more than one participant in a study uses wearable cameras then a combined picture of the situation of use is possible – with access to interactions and gestures that present themselves as awareness, control and participation.

Analysis

The biggest challenge in using video methods comes from the lengthy nature of the analysis phase. While there are some techniques that can be used to speed analysis, video analysis is a topic of enquiry in its own right. Here we will give pointers on narrowing the collection of video data to be analysed, and contrast coding and interaction analysis as two approaches to analysing video data – be that video of interaction with devices or screen recordings of use.

Data reduction – initial classification

It is a common criticism of the use of video in user studies that it is time consuming – just watching once through the 100 or so hours of recordings made from our two studies would be a considerable undertaking. One approach is to narrow down the data before analysis by triangulation with the other data collected. For example, location can be used to narrow down clips to those taking place in public or application logging can be used to collate and analyse examples of the use of certain apps or features – such as social media apps, or navigation apps. One technique we have found useful is to focus on recordings where conversations are heard taking place alongside device use, providing materials for the analysis of interactions with co-present others.

A contrasting technique is to use video summarisation, such as [28], however these systems can struggle with the visual similarity of screen recording. Alternatively, audio can be used to identify and summarise recordings, such as with tools like Panopticon [22]. For those who can build their own tools open source libraries such as Sphinx [26] are available to help the development of audio analysis solutions, and libraries such as VLFeat [48] can be useful for the analysis and automatic annotation of video. One simple technique we found useful was to use ffmpeg to extract frames from videos – producing "flick books" of screenshots that summarised particular videos of use and were helpful during analysis.

Coding

Perhaps the most popular video analysis techniques are approaches that code data in some way, annotating video clips to look for generalizable patterns and correlations [17]. Coding behaviour in video draws most directly on the behavioural sciences, and particular behavioural and animal research [17]. Such coding of recorded data is the first step in many analytical techniques. In deciding on the individual codes there are a multitude of methods which can be employed, from grounded analysis where only the subjects' language is used to form the codes, to those such as structural coding where the data is segmented and assigned codes based on the research questions that prompted the data collection.

The act of coding itself is a primary stage of analysis, where the data is examined in detail and sought to be understood – at least on some level – by those involved in the coding. In many cases it is advantageous to perform a multi-party coding phase, and to work towards coder reliability. Such coding need not be limited to the video data. Log data, and other channels that have been collected, need not be used only to reduce the corpus of video data to be watched. By combining information from multiple sources it is possible to build more nuanced coding schemes. One field that has made extensive use of video is education, where it has been key in, for example, understanding infant development [13].

Interactional Analysis

In the two studies outlined above, however, rather than coding the data we made use of interactional analysis to take a more detailed look at smaller incidents – looking to understand specific activities. It is beyond the scope of this paper to give a lengthy explanation of interactional analysis – Silverman presents two useful introductions [42, 43], in particular of the foundation of interactional analysis in conversation analysis and video analysis more broadly [18].

In practice, interaction analysis means looking closely at examples of interaction and understanding that each of these examples forms one or more of these patterns, then examining how – through the talk, device and bodily interaction taking place as well as through the other channels of information available to the researcher such as log data and diary entries – this pattern is negotiated. In the examples above the interactions are simultaneously taken as individual, unique incidents of use – but are also exemplifying patterns that we can extrapolate to be present in many other situations, with many other participants. In bringing this method to bear on the recorded interactions from mobile device use in natural everyday settings it is possible to build an understanding of the underlying structures seen in the interactions which we examine. In doing so we are able to gain a greater understanding of where technology fits into these structures and patterns of human interaction, where it breaks these patterns – for good or ill – and where the opportunities for design and development are situated.

With respect to our interest in mobile interaction this has guided us to be as much as possible interested in asking what a user is orienting towards in a particular situation, what are they trying to do, why do they do this at this point in time and place, what resources do they draw upon, how does the interaction display that, and so on. To answer these sort of questions involves a detailed analysis of particular video clips – not the repetition of a specific particular method as such, but a much more crafted set of sessions of interrogation of video, and theoretically informed

inspection of clips. In our analysis of mobile interaction two foundational concepts have proven useful: *Sequentiality,* and *Recipient Design*. Looking through the lens of *Sequentiality* we identify not only the order of interactions, but the reasons behind that order and the some of the other possible paths that the interaction could have taken. An answer, for example, should follow a question. In our second example above we start to see how a customer interaction works within a particular sequential form. In contrast, using the concept of *recipient design* lets us try to understand where the user is crafting their interactions with the intention to make their actions understandable in a certain way to an observer, be that the system or another person. Actions are 'designed' for their recipients. In this way we are able to craft a holistic view not only on what is happening in our clips, but also on why it is happening and how the system design and local context are exerting influence on the pattern of interaction.

Presenting Data

A final challenge in our method is presenting data in presentations and papers. It can be tempting in talks to just 'play the video', or in papers to include lengthy transcriptions. In our second example above, for example, we made use of a transcript annotated with screenshots of activity recorded on the screen. In many cases this can be sufficient – here having a transcript reveals some of the timing issues around the talk between customer and server and interaction with the phone. A transcript allows for close attention to be given to what is said, and the pauses in interactions that can be consequential.

Yet for some extracts where bodily orientation is important, – such as in our first example – 'comic strip' style views can illustrate well how interaction is going physically alongside talk. In this case we used the Comic Life application to produce the transcript, and while this is often not practical for longer extracts, it can illustrate small extracts well if care is taken to extract suitable images.

The privacy and control issues related to deploying such recording devices raise another aspect where presentation of this data is important. These issues are nuanced and vary from situation to situation, however the pervasive and invasive nature of screen and audio recording has been already noted [45]. In recording aspects of their lives that may not have been open to replay, scrutiny and distribution before there is a risk. Participants should have the opportunity and tools to both control their exposure in the moment, and after the data has been recorded. Control mechanisms must be tuned to the data collected and the expectations of the participants. In using remotely automatically uploaded video recordings we provided a website that required participants to approve specific video recordings that we could see and use. This is harder with video recordings from cameras, where it can be more difficult to force participants to edit their video recordings, but it is important to make participants aware of, and approve, what they are sharing as part of their participation.

DISCUSSION

Analysing logs of usage and interactional video analysis may seem very different approaches. However they are both methods of recording events and analysing them later – repeatedly if required. The power of log based studies of device use tends to lie in the number of examples that can be collected, and then used to gain greater certainty of the extent to which a phenomena in the data can be reasonably expected to be present in the everyday interactions of the majority who are not being logged [20]. The power of video analysis, on the other hand, is in the deeper understanding and identification of interactional phenomena that can then be recognised time and again in other contexts of use. This approach may seem to draw only on very short singular clips of mobile device interaction, yet its power comes from the focus on understanding the 'methods' of device use – the means by which individuals put their devices to useful purposes. So while the identification of methods of device use might originate in the close analysis of single clips, this is not to say that it only exists in a single case, but rather than once identified it can be compared, contrasted, and found in other examples of use – other video clips in the first instance, but also more broadly in other observations of device use.

A second point to make is that while observations are embedded in the close analysis of a single clip, the analyst, in turn, draws more broadly on the whole collection and the other streams of information collected from the participants. The automated collection of log data of app use, location, interviews, diary entries, and direct observation are all drawn upon to form an analysis. In the decision of what is suitable for closer analysis, an understanding of what is novel, what is exemplary and what constitutes a complete episode of interaction is necessary and this can only come from a broader understanding of the examples of use that make up any corpus of data. We recommend that logging and interaction analysis be seen as complimentary tools in the quest for both deeper and broader understanding of mobile device use, interaction, and integration into the lives of those that use them.

However, we note that traditional methods may still be more successful in gaining access to specific understanding required in specific circumstances. Using more traditional interview or survey techniques is necessary when trying to understand reflections, or the feelings of participants is the goal. For example, video data may help, but the resulting analysis would never replace an expert users' reflection on the introduction of technology to a practice.

A privacy problem often cited with such recording is that of incidental capture of those interacting with the participant or those who just happen to be within the range of the recording device. While the value and efficacy of informed

consent in user studies research has been recently under debate [30] at its core it is based on the principle of ensuring the autonomy of those you are scrutinising – for those incidentally captured there is no option to give them access to the data in which they feature, or to explain to them the study procedures, as there are no communication channels between them and the researchers. With respect to this challenge recording video and audio with large, visible and recognisable equipment gives much more awareness, and therefore control, to those third parties. There are a number of technological solutions that can go towards mitigating some of this risk.

CONCLUSION

In this paper we have described *in vivo,* a combinative method which encapsulates five data collection techniques – *screen recording, ambient audio recording, data logging, wearable cameras,* and *distributed remote uploads* – collecting real-time sequential data of mobile device use that brings together the opportunities of AppStore mobile deployment and qualitative video analysis. By presenting a real world example of two scales of data collection used in this approach we have shown how video, of both user and screen, combined with audio recordings give us insights into situations of use that would have been impossible with other standard methods. Current mobile technology, including the storage space necessary for this data and the bandwidth to transfer it back to researchers, supports this practice with little detriment to the users. While the analysis of video and audio data can take considerable time, and the collection of a large amount of video data from a number of participants could provide a daunting corpus, we have discussed how using current logging techniques to triangulate analytic efforts allows the refined detail for analysis to be extracted from a user study with an, in our opinion, proportionally acceptable cost in terms of researcher time and effort. We have shown how the application of *in vivo* helps to change the "the unit of analysis" [39] by allowing us to see mobile device use as something embedded within sequences of everyday activity that span across digital media, physical media, interaction, and activity [7]. This greater understanding of the context of use allows for greater understanding of the use itself.

In closing we would argue that the approach provided here has potential to enable better understandings of how devices are used in context, bring greater depth to mobile evaluation, and lead to a more substantive understanding of what it is to use a mobile device. Through understanding how our devices are woven into the complex tapestry of everyday interaction, we can provide the foundations for both understanding and designing systems that fit with and enhance our interactions with technology and each other.

ACKNOWLEDGEMENTS

We would like to thank Eric Laurier for drawing figure 1, as well as help with the analysis discussed in this paper.

REFERENCES

1. Ankolekar, A., Szabo, G., et al. Friendlee: a mobile application for your social life. *In Proceedings of MobileHCI '09* (2009), 27, ACM.
2. Benford, S., Seagar, W., et al. The Error of our Ways: The experience of Self-Reported Position in a Location-Based Game. In Ubicomp *2004,* Springer (2004).
3. Böhmer, M., Hecht, B., et al. Falling asleep with angry birds, facebook and kindle: a large scale study on mobile application usage. *In Proceedings of MobileHCI'11* (2011),47-56 ACM.
4. Brown, B. and Laurier, E. The normal natural troubles of driving with GPS. *In Proceedings of CHI'12* (2012),1621-1630 ACM.
5. Brown, B., McGregor, M., et al. iPhone in vivo: video analysis of mobile device use. In *Proceedings of CHI'13* (2013),1031-1040 ACM.
6. Brown, B., McGregor, M., et al. 100 days of iPhone use: understanding the details of mobile device use. In *Proceedings of MobileHCI'14* (2014), 223-232 ACM.
7. Brown, B., McGregor, M., et al. Searchable Objects: Search in Everyday Conversation. In *Proceedings of CSCW'15* (2015), 508-517 ACM.
8. Brown, B., Taylor, A. S., et al. Locating family values: A field trial of the Whereabouts Clock. In *Proceedings of Ubicomp'07* (2007) Springer.
9. Carter, S., Mankoff, J., et al. Momento: support for situated ubicomp experimentation. In *Proceedings of CHI 2007* (2007),125-134 ACM.
10. Church, K., Cousin, A., et al. I wanted to settle a bet!: understanding why and how people use mobile search in social settings. *In Proceedings of MobileHCI'12* (2012),393-402 ACM.
11. Church, K., Neumann, J., et al. SocialSearchBrowser: a novel mobile search and information discovery tool. In *Proceedings of Intelligent user interfaces* (2010) ACM Request Permissions.
12. Cooley, J. and Smith, S. Privacy-preserving screen capture: Towards closing the loop for health IT usability. *Journal of biomedical informatics*, 46, 4 (2013), 721-733.
13. Derry, S. J. *Guidelines for video research in education: Recommendations from an expert panel*. NORC at the University of Chicago, Data Research and Development Center, 2007.
14. Ferreira, D., Goncalves, J., et al. Contextual experience sampling of mobile application micro-usage. In *Proceedings of MobileHCI (2014)*,91-100 ACM.
15. Froehlich, J., Chen, M. Y., et al. MyExperience: a system for in situ tracing and capturing of user feedback on mobile phones. In *Proceedings of MobiSys '07* (2007),57-70 ACM.

16. Goto, H. and Tanaka, M. Text-Tracking Wearable Camera System for the Blind. In *ICDAR '09* (2009),141-145 IEEE.
17. Hailpern, J., Karahalios, K., et al. A3: HCI Coding Guideline for Research Using Video Annotation to Assess Behavior of Nonverbal Subjects with Computer-Based Intervention. *TACCESS*, 2, 2 (2009).
18. Heath, C., Hindmarsh, J., et al. *Video in qualitative research: analysing social interaction in everyday life*. Sage, London, 2010.
19. Heath, C. and Luff, P. *Technology in action*. Cambridge university press, Cambridge, 2000.
20. Henze, N., Rukzio, E., et al. 100,000,000 taps: analysis and improvement of touch performance in the large. In *Proceedings of MobileHCI '11* (2011), 133-142 ACM.
21. Hindmarsh, J., Heath, C., et al. Creating assemblies: aboard the Ghost Ship. In *Proceedings of CSCW'02,* (2002), 156-165 ACM.
22. Jackson, D., Nicholson, J., et al. Panopticon: a parallel video overview system. In *Proceedings of UIST '13* (2013),123-130 ACM.
23. Kamvar, M. and Baluja, S. A large scale study of wireless search behavior: Google mobile search. In *Proceedings of CHI'06* (2006),701-709 ACM.
24. Kjeldskov, J., Skov, M., et al. Is it worth the hassle? Exploring the added value of evaluating the usability of context-aware mobile systems in the field. In *Proceedings of MobileHCI 2004* (2004),61-73 ACM.
25. Lam, M. S., Godbole, S., et al. Measuring time spent outdoors using a wearable camera and GPS. In *Proceedings of SenseCam'13* (2013),1-7 ACM.
26. Lee, K.-F., Hon, H.-W., et al. An overview of the SPHINX speech recognition system. *Acoustics, Speech and Signal Processing,* 38, 1 (1990), 35-45 IEEE.
27. Licoppe, C. and Figeac, J. Direct video observation of the uses of smartphone on the move. Reconceptualizing mobile multi-activity. In *Mobility and Locative Media. Mobile Communication in Hybrid Spaces,* Routledge (2015), 48-64.
28. Lienhart, R., Pfeiffer, S., et al. Video abstracting. *Communications of the ACM*, 40, 12 (1997), 54-62.
29. Luff, P. and Heath, C. Mobility in Collaboration. In *Proceedings of CSCW'98,* (1998), 305-314 ACM.
30. Luger, E. and Rodden, T. An informed view on consent for UbiComp. In *Proceedings of Ubicomp'13* (2013),529-538 ACM.
31. Lyons, K. and Starner, T. Mobile capture for wearable computer usability testing. In *International Symposium on Wearable Computers* (2001), 69-76 IEEE.
32. McMillan, D., Morrison, A., et al. Further into the wild: Running worldwide trials of mobile systems. In *Proceedings of Pervasive '10* (2010),210-227 ACM.
33. Mohd Noor, M. F., Ramsay, A., et al. 28 frames later: predicting screen touches from back-of-device grip changes. *In Proceedings of CHI 2014*, 2005-2008 ACM.
34. Morrison, A., McMillan, D., et al. A Hybrid Mass Participation Approach to Mobile Software Trials. *In Proceedings of CHI '12* (2012),1311-1320 ACM.
35. Müller, H., Gove, J., et al. Understanding tablet use: a multi-method exploration. In *Proceedings of CHI'12 (2012)* ,1-10 ACM.
36. Pitkänen, J., Pitkäranta, M., et al. Usability testing in real context of use: the user-triggered usability testing. In *NordiCHI '12* (2012), 797-798 ACM.
37. Reyal, S., Zhai, S., et al. Performance and User Experience of Touchscreen and Gesture Keyboards in a Lab Setting and in the Wild. In *Proceedings of CHI'15 (2015)* ,679-688 ACM.
38. Robinson, S., Marsden, G., et al. *There's not an app for that: Mobile User Experience Design For Life*. Morgan Kaufmann, 2014.
39. Rogers, Y., Scaife, M., et al. *Interdisciplinarity: an Emergent or Engineered Process?*, University of Brighton, Cognitive Science Research paper 556, 2003.
40. Schusteritsch, R., Wei, C. Y., et al. Towards the perfect infrastructure for usability testing on mobile devices. In *Proceedings of CHI '07* (2007),1839-1844 ACM.
41. Sellen, A. J., Fogg, A., et al. Do life-logging technologies support memory for the past? In *Proceedings of CHI '07* (2007),81-90 ACM.
42. Silverman, D. *Harvey Sacks: social science and conversation analysis*. Polity Press, Cambridge, 1998.
43. Silverman, D. *A Very Short, Fairly Interesting and Reasonably Cheap Book about Qualitative Research*. Sage, London, 2007.
44. Streefkerk, J. W., Esch-Bussemakers, M. P. v., et al. Field evaluation of a mobile location-based notification system for police officers. In *Proceedings of CHI'08 (2008)*,101-108 ACM.
45. Tang, J. C., Liu, S. B., et al. Unobtrusive but invasive: using screen recording to collect field data on computer-mediated interaction. In *Proceedings of CSCW '06* (2006),479-482 ACM.
46. Thomaz, E., Parnami, A., et al. Technological approaches for addressing privacy concerns when recognizing eating behaviors with wearable cameras. In *Proceedings of UbiComp '13* (2013),739 ACM.
47. Tolmie, P. and Crabtree, A. Deploying research technology in the home. In *Proceedings of CSCW'08 (2008), 639-648 ACM*.
48. Vedaldi, A. and Fulkerson, B. VLFeat: An open and portable library of computer vision algorithms. In *Proceedings of Multimedia* (2010),1469-1472 ACM.
49. Wimmer, R. Grasp sensing for human-computer interaction. In *Proceedings of TEI'11* (2011), 221-228