Studying long-term, fragmented data sets

Stuart Reeves, Alistair Morrison and Matthew Chalmers
Computing Science,
University of Glasgow.
Glasgow G12 8QQ, UK
{stuartr, morrisaj, matthew}@dcs.gla.ac.uk.

CHI 2009, April 4 – 9, 2009, Boston, MA, USA
ACM 978-1-60558-246-7/09/04.

Abstract
In this paper we suggest two challenges for the study of fragmented data sets generated from long-term studies. The first of these is the wide range of temporal perspectives from which a single data set may be inspected (from seconds to weeks). The second challenge involves the importance of considering user experience of time as a useful resource in analysis. Finally we briefly conclude with a call to consider new analytic tools that move beyond solely timeline-bound representations.

Keywords
Temporal data sets, qualitative and quantitative analysis tools, ubicomp systems

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

Introduction
The evaluation of ubicomp systems often creates difficult demands both for the practicalities of data collection, and the evaluative practices of researchers during and after the period of use. The fragmented, socially embedded and typically mobile nature of ubicomp systems is coupled with the frequent desire of researchers to understand the ways in which

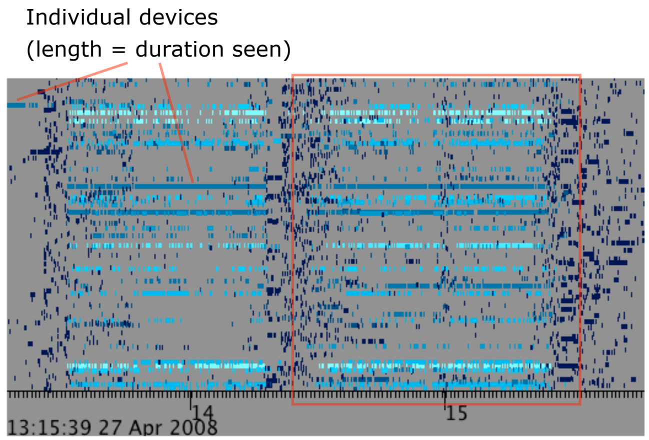

Figure 1: The density of several hours of Bluetooth devices around an individual, before, during and after a football match. The individuals that have been seen for the longest accumulative time are shown in a lighter colour. Within this data, game patterns, such as the two halves of the match (second half marked) and subsequent journey home are visible.

technology comes to be woven into everyday life and 'mundane' through adaptation, appropriation and contextualisation of system use. More often than not this means that periods of use under evaluation are as long-term as possible (although there are notable exceptions, particularly within ubicomp research that examines performance settings, e.g., [7, 1]), and, increasingly often, involves trials and deployments that are conducted 'in the wild' of everyday life rather than in controlled environments [6].

Temporal range

This concern with evaluation periods lasting days, weeks and sometimes months in natural settings, in combination with the disparate, transient and mobile forms of interaction that occur in many ubicomp systems, results in data sets fragmented both spatially and temporally [2]. Assuming the considerable problems of data collection are overcome, analysts must then reconstruct and piece together fragmented sets of diverse data types (e.g., audiovisual recordings, field notes, system transaction log files, sensor logs such as GPS or wifi scanning data, etc.). A subsequent issue that faces analysts is the potential range of temporal scales under which this data set can then be examined.

To illustrate this we can consider examples from our own and others' research projects. Recently, for example, we have been investigating the conduct of crowds at and around stadium-based sporting events. As part of this we have collected a body of video recordings documenting the forms of interaction that spectators engage in. Additionally, we have recorded Bluetooth device identity data using scanners carried by those attending these events, in an attempt to document the crowd's conduct at large. In studying this data we are interested both in the large-scale trends of crowd activity detected by the simple scanning

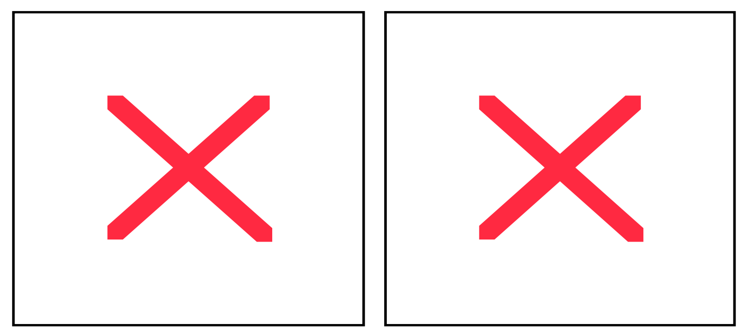

Figure 2: A small moment of interaction from a match-day video recording in which two sets of football supporters located in the same pub – fans of the Scottish and Norwegian national teams – 'break the ice', bridging between opposing crowd groupings via intimate gestures.

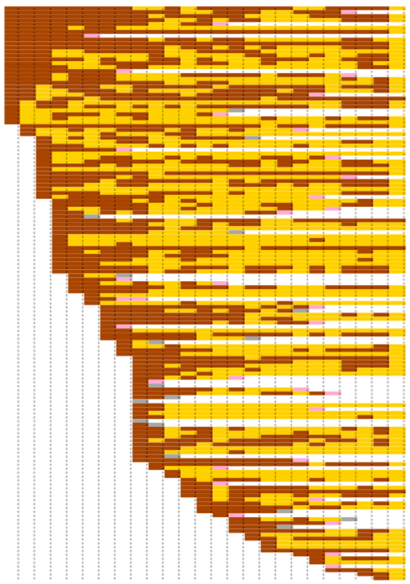

Figure 3: Engagement of players in Day of the Figurines over the course of 24 days (taken from [3]). Red bars indicate days players were 'engaged' in the game, yellow bars indicate that they were 'disengaged' and white indicating that they had exited the game.

technique (see Figure 1), and the smallest of interactional moments of crowd activity (e.g., see Figure 2). In spite of the differences, analysis of extreme ends of the temporal spectrum can often be mutually informing, providing context and richer understanding.

Analysis of other projects, such as Day of the Figurines (DoF), a narrative-driven interactive SMS-based game for mobile phones, reveals an even wider temporal scope, and further issues when examining temporal data. The game itself was intended to be played episodically, unfolding "in the background of players' daily lives" [5]. Once again, log data and video recordings collected from weeks of play by hundreds of players were examined both for long-term patterns of engagement (see Figure 3, covering 24 days) [3], as well as for fine-grained ethnographic analysis of orchestration practices [4].

Thus, we can see how an analysis of data collected from a ubicomp system trial may have stages that focus on very different levels or scales of temporal detail.

User experience

Interestingly, the DoF analysis also highlighted how players may experience the flow of time differently due to their varying levels of engagement in the game (e.g., playing every day versus playing once a week). Thus, when addressing temporal data sets, understandings of different layers of experienced "temporal trajectories" [5] becomes key. In DoF, the transactions between players and game orchestrators in collaboratively producing a narrative, and player engagement in the game (Figure 3) can be understood in terms of different temporal narrative trajectories. For analysts, therefore, not only is the sheer range of temporal scales potentially very wide for any relatively diverse data set, but also we can begin to see how the temporal experience of those being evaluated may vary greatly. The relevance of this may well need to be folded into the analytic methodology.

We might also consider this subjective temporal experience for shorter periods. For the football supporters in the first example, for instance, 90 minutes of the match experience will differ greatly in pace when compared to discussing the game with others subsequently. Pushing this notion further, we can consider more extreme analytic examples, such as the study of interactive systems that operate on fairground rides [1]. In this instance, as with many 'stressful' experiences, perception of time radically changes [12], and so interpreting sensor and video data of participants may well need to take this into account when evaluators try to make sense of these experiences.

Challenges

From these observations on the evaluation challenges posed by varied data sets and projects, we can make some more general comments on the broader challenges for the evaluation of fragmented, long-term data sets.

Firstly we looked at the potential for very wide temporal range within sets of data. The issue here then is what might be the best tools and methods for analysts concerned with differing ends of the temporal detail spectrum. We saw both aggregations of data over the course of hours, days and weeks, and how segments of (typically video) data from this same data set that may last a mere few seconds can also be of interest. For example, long term appropriation of a technology may hinge on many brief moments of interpersonal interaction.

In dealing with data at varying temporal scales, an analyst is likely to need a variety of different types of tools, each with varying strengths and weaknesses. For example, whereas it is possible to look at trial-wide overviews of log data to examine long-term trends of system use, it is harder to summarise the contents of video data in a useful way without sitting through the full corpus of collected footage. Driven by the recognition that varying types of recorded data form part of the same data set, one might ask whether it is of greater benefit to use a collection of tools, each specialised for a single type of data (often a single end of the temporal spectrum) or a system that attempts to merge such tools into one application, but the most likely way forward may be the linking of specialised tools to form one co-ordinated ensemble. A key design characteristic of information visualisation is to show fine-grained detail in the wider context of, ideally, the entire data set [8] Systems such as DRS [11] and our Replayer system [9] have taken this approach, aiming to show how fine-grained segments may elaborate upon the whole (and vice versa). We note, however, that a perennial design challenge with this approach is how to let users move smoothly between overviews of time and the most useful small periods within it. Drilling down, as such, is not a problem but knowing where to drill down is. For example, it would be technically straightforward to use the Bluetooth scanning data visualisation (Figure 1) as an index for video clips recorded by the individual conducting the scanning, but in practice the quantitative features that the visualisation shows may not be the most useful clues as to the qualitatively most significant short sequences of video.

A similar issue with wide-ranging fragmented temporal data sets is the often large differences in the rates of collection or sampling of quantitative data. For example, accelerometer and related sensors may need sampling rates of over 50Hz in order to get data accurate enough for some modelling approaches, but it is not immediately apparent how to visualise—or visualise well—that data alongside GPS readings being taken once every 10 seconds, or field notes taken every few minutes, especially if the patterns under investigation are happening at a very fine scale. Like the bridging between fine-grained video clips and coarse-grained Bluetooth patterns, such multilevel design seems to call for forms of abstraction that bridge between very different quantitative models. For example, what features or patterns in the accelerometer data for a short period of time lead to interesting GPS data around that same time—and vice versa? Such bridging or ‘metamodelling’ is not yet well understood or established.

Secondly, we explored how the temporal experience of those we observe may play an important role that might not be represented in our current analysis tools. It might be practical to include self-reporting mechanisms that explicitly track user experience of time, to be used to change or adapt representations within temporal data analysis tools. For example, individual players in DoF did not necessarily share the same temporal patterns of engagement in the game, and having a representation of such engagement alongside raw log and video data might helpfully inform how we choose to interpret participants' understandings of the passing of time and the user experience as a whole.

Similarly, such treatment might be used to adapt quantitative algorithmic analysis. One example of such analysis from our own work concerns Bluetooth scan data of the type visualised in Figure 1. By dividing the data set into short windows of time, we can find sets of Bluetooth devices that were within range of each other in a significant proportion of such windows and use this to suggest membership of informal and temporary groups of fans (although further analysis is usually needed to confirm this). Similarly, in other work, we have used an information theoretic measure, mutual information, to find periods of time in which pairs of participants appear to show coordination of movement and system use. For example, in the Treasure game [10], where players competed in teams of two, we used mutual information to examine the extent to which team members coordinated movement as opposed to acting completely independently of their team mate; we could thus automatically find pairs of players who had developed tactics involving separately covering different parts of the game arena. Such techniques generally rely on analysing windows of time of uniform length, but the length of time is chosen rather arbitrarily. Analysts may guess that, for example, windows of 5 minutes' duration would be useful, but these strict divisions of time might not fit well with the way players organise their own reports of the experience. We suggest that algorithms for quantitative data analysis might adapt according to participants' reports of engagement, dynamically adjusting window sizes to fit with, for example, users' impressions of their own activity, the pace of interaction, and the intensity of the user experience.

Conclusion

These two challenges, of wide temporal ranges of data sets and the representation of subjective temporal experiences of participants, thus lead us to question the way that analysis tools might be designed in order to collect together such long-term and fragmented temporal data. As different forms of analysis may require separate tools or components and introduce multiple subjective timelines, we might also consider representing these lengthy subjective time ranges of data in ways that are not restricted by predominantly timeline-based forms. For instance, a single timeline, however augmented, may compress subjective temporal experiences into an objective representation that distorts how a series of events was actually experienced. Whilst not wishing to abandon the timeline representation, we would ask what alternative spatial and representational forms might help us arrange data sets in new ways that are more sensitive to the nature of the user experience. Or, in other words, when is a timeline a useful way of arranging such diverse data, and when does it result in an over-simplifying distortion?

References

[1] Schnädelbach, H. et al. Performing thrill: Designing telemetry systems and spectator interfaces for amusement rides. *Proc. ACM CHI 2008*, pp. 1167-1176.

[2] Sherwood, S. et al. Adapting evaluation to study behaviour in context. *International Journal of Mobile Human Computer Interaction*. In press.

[3] Capra, M. et al. Day of the Figurines Evaluation Report. http://www.pervasive-gaming.org/Deliverables/D12.6-DOF-Evaluation.pdf

[4] Crabtree, A. et al. The cooperative work of gaming: orchestrating a mobile SMS game, 2007, J.CSCW.

[5] Benford, S. and Giannachi, G. Temporal trajectories in shared interactive narratives. *Proc. ACM CHI*. 2008.

[6] Crabtree, A. et al. Supporting Ethnographic Studies of Ubiquitous Computing in the Wild. In Proc. Designing Interactive Systems (DIS), 2006, ACM Press.

[7] Benford, S. et al. Can you see me now? ACM *Transactions on Computer-Human Interaction (ToCHI)*, 2006, ACM Press.

[8] Card, S. K, Mackinlay, J. D., Shneiderman, B., *Readings in information visualization: using vision to think*. Morgan Kaufmann Publishers, 1999.

[9] Morrison, A. et al. Coordinated Visualisation of Video and System Log Data. In *Proc. International Conference on Coordinated & Multiple Views in Exploratory Visualisation*, p. 91-102, 2006.

[10] Barkhuus, L. et al. Picking Pockets on the Lawn: The Development of Tactics and Strategies in a Mobile Game. Proc. *UbiComp*. 2005.

[11] Brundell, P. et al. Digital Replay System (DRS) - a tool for interaction analysis. In Proc. of International Conference on Learning Sciences, Utrecht: ICLS, 2008.

[12] Hancock, P. A. and Weaver, J. L. On time distortion under stress. *Theoretical Issues in Ergonomics Science*:6(2), pp. 193-211, Taylor & Francis, 2005.