GATHERING INDIVIDUAL TRAVEL DATA WITH GPS-ENABLED SMARTPHONES: A PROOF-OF-CONCEPT STUDY

J van Dijk and S Krygsman*

Stellenbosch University, Private Bag X1, Matieland, 7602
Tel. +27 (21) 8082624; E-mail: jtvandijk@sun.ac.za
*Stellenbosch University, Private Bag X1, Matieland, 7602
Tel. +27 (21) 8082624; E-mail: skrygsman@sun.ac.za

ABSTRACT

Policies aimed at Traffic Demand Management (TDM) rely heavily on the gathering of accurate individual level activity and travel data to understand and unpack the demand for transport. Moreover, self-report based data collection methods face problems such as a high-respondent burden and inaccuracies in the number and duration of the reported trips. On that account, this paper presents ongoing research to assess the reliability and feasibility of passively collecting high resolution spatiotemporal data on activity and travel behaviour using GPS-enabled smartphones. A small-scale pilot study was conducted in which respondents from the University of Stellenbosch, South Africa, were passively tracked for the course of two days by means of a purposefully designed smartphone application, termed TrackLog. The results of the small experiment indicate that while GPS technology in smartphones potentially holds a number of benefits for collecting activity and travel data, the technology is not without problems. This project has highlighted that these problems can be classified as: (1) user, (2) technology, and (3) methodology related problems. Notwithstanding these problems, the results indicate that gathering high resolution space-time data by means of GPS-enabled smartphones is feasible and that it opens doors to a range of possible applications that are unattainable by traditional survey methods.

1. INTRODUCTION

Urban transport systems and road networks are under pressure as a result of a rapid increase in private vehicle ownership and increasingly complex and fragmented travel patterns (Axhausen et al. 2002; Mokhtarian et al. 2006; Järv et al. 2014). The highly visible externalities of these modern societal trends include daily congestion and traffic gridlock, but also environmental degradation. In 2003, road transport accounted for 18% of the worldwide CO2 emissions. This dependency on road transport poses major challenges to the environment (Chapman 2007; Graham-Rowe et al. 2011; Finn 2012). While developed countries are still responsible for 70% of the worldwide transport related greenhouse emissions, the relative contribution of upcoming economies such as China, India and South Africa is expected to grow significantly in the next decades (Hickman & Banister 2010).

To address these challenges, policy makers and researchers have started to shift their attention to innovative measures to curb the demand for car use with travel demand management (TDM) policies (Kitamura et al. 1997; Loukopoulos et al. 2004; Stradling &
Anable 2008). Also in South Africa the focus has shifted from supply-side to demand-side passenger transport planning (Behrens & Mistro 2010). This is exemplified by the National Land Transport Act (Act 5 of 2009), in which it is stated that municipalities have to formulate and implement TDM techniques in their transport planning. Travel demand management, however, is not an undemanding task, as it “relies on the agencies’ ability to accurately predict future demand, suggest viable improvements and shift the demand away from the automobile to more sustainable transportation modes like walking, biking and transit (Jariyasunant et al. 2011, p.3).”

In order to understand travel behaviour and accurately predict future demand, precise and reliable activity and travel data are essential (Shafique & Hato 2015). However, obtaining disaggregate travel data through traditional methods such as paper-based survey instruments is a complex endeavour as a result of, but not limited to, a high respondent burden, the underreporting of trips, costly administrative processes and the variability of human travel behaviour (Behrens 2004; Stopher et al. 2009; Jariyasunant et al. 2011; Nitsche et al. 2014). Although more cost-effective, also computer-assisted methods do not solve all limitations associated with traditional methods because these methods are also contingent on the respondent’s ability to accurately remember his or her movements and activities (Shafique & Hato 2015). This makes administering and collecting travel surveys a challenging exercise, even more so for longitudinal studies (Behrens & Mistro 2010).

Over the past decade, technological developments have advanced the state of research with regard to travel surveys (Cottrill et al. 2013). In particular, Location Aware Technologies (LAT) such as Global Positioning Systems (GPS) have greatly enhanced the opportunity to collect more accurate data on human spatiotemporal behaviour for longer periods of time (Shen & Stopher 2014). These technologies are nowadays also available on most smartphones. The advancements in the technological realm in combination with the need for high-quality data on spatiotemporal human behaviour for transport management and planning, offers challenging new avenues for researchers and practitioners. It follows that this paper describes ongoing efforts to assess the reliability and feasibility of smartphone tracking, in the context of South Africa, to suggest ways in which location aware technology can be employed in collecting data on individual activity and travel behaviour.

2. LITERATURE REVIEW

In order to capture, understand, predict and possibly manage travel behaviour, high quality activity and travel data on an individual level is essential. Given the complexities associated with gathering disaggregate data, researchers have started to experiment with acquiring these data by means of new technologies such as mobile phones (cf. Asakura & Hato 2004; Krygsman & Schmitz 2005; Krygsman et al. 2008), mobile phone call detail records (cf. Järv et al. 2014), and Bluetooth technology (see for a recent overview of mobile technologies used in activity-travel data collection Rasouli & Timmermans 2014). Increasingly, starting in the late 1990s, Global Positioning System (GPS) technology and, more recently, GPS-equipped smartphone technology have been put forward as a solution in supplementing or (partly) substituting for traditional travel data collection methods (see for instance Wolf 2000; Bohte & Maat 2009; Chen & Kwan 2012; Feng & Timmermans 2013; Nitsche et al. 2012; Nitsche et al. 2014; Shen & Stopher 2013; Shoval et al. 2014; Shen & Stopher 2014).
GPS technology provides opportunities to collect data that is otherwise difficult to acquire such as the exact start and end time of a trip, the chosen route and the distance travelled (Blazquez & Miranda 2014). In addition, whereas paper-based travel surveys can be costly and put a significant cognitive burden on respondents (see for instance Behrens 2004), GPS technology has the potential to reduce respondent burden whilst increasing the spatiotemporal accuracy of the collected data (Du & Aultman-Hall 2007). As Cottril et al. (2013, p.59) note: "whereas the decades-long experience with household travel surveys typically depended precariously on the vagaries of human memory, agent’s (...) movements and activities can now be tracked in great temporal and spatial detail." The increasing availability and ubiquity of smartphones makes it even easier for researchers and policy makers to take advantage of GPS technology.

One of the major advantages of collecting locational data by means of smartphones when compared to dedicated GPS-devices and other mobile technologies, emanates from the fact that: “people habitually carry their mobile phones with them much of the time as this pervasive technology offers its users to a means of constant and available communication as well as personal entertainment (Horanont et al. 2013, p.1).” Several studies have thus far begun to examine the use of smartphones in collecting data on activity and travel behaviour (e.g. Bierlaire et al. 2010; Nitsche et al. 2012; Nitsche et al. 2014). Some researchers (for example Froehlich et al. 2009) provided their participants with a smartphone capable of tracking their movements, whilst others have experimented with delivering a smartphone application to participants who possess a smartphone (for instance Jianchuan et al. 2014).

Although there seems to be agreement that using mobile devices for the collection of activity-travel data is viable, a number of challenges are outlined for future studies. Not only the privacy of the respondents is often mentioned in this context, but also technological and methodological issues have been raised with regard to the level of battery consumption and the management of large data sets (Rasouli & Timmermans 2014; Shen & Stopher 2014). In addition, for GPS-based technologies to fully replace traditional travel surveys, analytical method should be able to deal with imperfect locational information – especially because the signal of GPS-receivers in smartphones can be blocked or weakened when the phone is carried in a purse or a pocket (Bierlaire et al. 2013).

3. COLLECTING AND ANALYSING GPS-TRACKS

In August 2014, an experiment was conducted at Stellenbosch University which aimed to assess the reliability and feasibility of GPS-enabled smartphones in collecting data on transport and activity behaviour. An application, termed TrackLog, was developed for the Android mobile operating system. The choice for the Android operating system was based on the outcomes of a survey conducted at Stellenbosch University in 2013, which indicated that the market share of Android was estimated to be 31% in 2013 and projected to increase to 47% in 2014 (Stellenbosch University 2013). Upon activation, TrackLog allowed the recording of its user’s positional information using the GPS-location registered by the smartphone. This information included the geographical coordinates \((lat, long)\) and the time \((t)\) of each measurement \((x)\).
The recorded information was uploaded via the GSM network to a server from where the information could be downloaded by the researchers. A location measurement frequency of 30 seconds was set for the data collection, leading to the potential recording of 2880 records per 24 hours. These 2880 records resulted in a database file of approximately 220 KB. Uploading these data over the cellular network did therefore not lead to high data costs as on average 1 MB of data does not exceed R1. The uploaded information could also be accessed by the participants: a feedback website was created that allowed participants to login to a web interface where they could view their personal tracks projected on a Google Maps background.

Participant recruitment was done by means of an e-mail invitation that was sent out to 100 individuals at Stellenbosch University, including post-graduate students, academics and support staff. The invitations yielded a non-representative sample of 15 individuals (6 females and 9 males). After accepting the invitation, the participants received a follow up e-mail which included an ethical consent form, the procedures of the experiment, and detailed instructions on how to install TrackLog. Because the project was reviewed by the Departmental Research Ethics Committee, participants were required to accept the terms and conditions of the ethical consent form. Only when their consent was received, he or she was provided with a link to download TrackLog and requested to switch the application on for the period of two days. In order to remind the participants to switch on the application, every participant was sent an automated text message on both days of the experiment.

Data from the phone was automatically uploaded every hour to the servers from where the data could be downloaded in a database format (.csv). Using the Python scripting language, the spatiotemporal measurements of each individual were imported and projected in ArcGIS 10.2 to visualise the locational information. IBM SPSS Statistics version 22 was used to visualise the spatiotemporal trajectory.

4. RELIABILITY AND FEASIBILITY OF SMARTPHONE TRACKING

After the two days of tracking, the locational information was downloaded from the server and merged in chronological order. This way an automatic time relationship was created between the records that represent the switching off and switching on of the device, be it either on purpose or as a result of other issues. The measurements were subsequently imported and projected in a GIS-environment (Hartebeesthoek LO19) – an example of which is given in Figure 1. It can be seen from the figure that there are both clusters of points, indicative of an activity, and individual points, indicative of travel. Taking the time of the occurring clusters in account, the clusters are indicative of a place of work in Stellenbosch (during office hours) and a place of residence in Kuilsrivier (outside office hours), respectively. In turn, the distance between two subsequent points indicative of travel, gives an indication of the speed with which was travelled.
Although elements of travel and activity patterns can be discerned from a simple visual inspection, a number of issues occurred during the tracking period. First, the automatic upload did not function as per design. Some phones experienced connectivity problems when switching from the Wi-Fi network to the cellular network and vice versa. Second, a number of participants reported that they could initially not log on to the TrackLog application which withheld them from starting to record their locations. Third, it was found that the location (GPS) services of the smartphones of some users were not activated or that the smartphone’s settings restricted background data transfer. This resulted in no locational information being recorded or uploaded at all. As a consequence of these technical issues, only 20 tracks (from 11 individuals) were suitable for further analysis.

For the 11 individuals qualifying for further analysis, Table 1 presents a basic overview of the collected data. The last two columns of Table 1 refer to the number of records that were captured. In general the participants did not record their location for the full 48 hour period, but switched the application off during the evening. In order to give an indication of the quality of the data, the percentage of records that could have been recorded was calculated for each participant. The percentages are relative to the time the participant initiated the tracking and terminated the tracking, respectively – whether this was on purpose or as a consequence of a technical failure. A value of 100% indicates that the phone continuously recorded locational information with a measurement frequency of 30 seconds between the starting and ending of TrackLog.
Table 1 | Overview collected data of the TrackLog study

<table>
<thead>
<tr>
<th>#</th>
<th>Phone Type</th>
<th>Android</th>
<th>Records Day 1</th>
<th>Records Day 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Samsung Galaxy S4 Mini</td>
<td>4.2.2</td>
<td>1259</td>
<td>95%</td>
</tr>
<tr>
<td>2</td>
<td>Samsung Galaxy S4</td>
<td>4.2.2</td>
<td>No Data</td>
<td>460</td>
</tr>
<tr>
<td>3</td>
<td>Motorola Moto E</td>
<td>4.4.4</td>
<td>1793</td>
<td>97%</td>
</tr>
<tr>
<td>4</td>
<td>Samsung Galaxy Pocket</td>
<td>2.3.6</td>
<td>653</td>
<td>37%</td>
</tr>
<tr>
<td>5</td>
<td>Samsung Galaxy S4</td>
<td>4.4.2</td>
<td>1320</td>
<td>125%</td>
</tr>
<tr>
<td>6</td>
<td>Samsung Galaxy S3 Mini</td>
<td>4.1.2</td>
<td>749</td>
<td>45%</td>
</tr>
<tr>
<td>7</td>
<td>Sony Xperia Z1</td>
<td>4.4.4</td>
<td>463</td>
<td>69%</td>
</tr>
<tr>
<td>8</td>
<td>Samsung Galaxy S4</td>
<td>4.2.2</td>
<td>No Data</td>
<td>827</td>
</tr>
<tr>
<td>9</td>
<td>Samsung Galaxy S5</td>
<td>4.2.2</td>
<td>323</td>
<td>26%</td>
</tr>
<tr>
<td>10</td>
<td>Samsung Galaxy S3</td>
<td>4.3.0</td>
<td>190</td>
<td>14%</td>
</tr>
<tr>
<td>11</td>
<td>Samsung Galaxy S4</td>
<td>4.2.2</td>
<td>1268</td>
<td>65%</td>
</tr>
</tbody>
</table>

In one instance in Table 1, the captured records exceed 100%. This implies that in this case the measurement frequency was lower than was intended. This may be attributed to the fact that TrackLog was set to synchronise its measurement frequency with the settings on the server. However, given the connectivity problems alluded to earlier, it is likely that TrackLog used a different measurement frequency instead. In other cases it turned out that the measurement frequency was actually longer than 30 seconds. This most likely implies that when the smartphone did not have a GPS-fix at the moment TrackLog requested it, the application recorded the location the moment the signal was available again. As noted by Blazquez and Miranda (2014, p.154): "a validation of the GPS measurements is needed due to the data gaps caused by respondents forgetting their GPS device or satellite signal blockage."

The impact of a missing GPS-fix should not be underestimated. Following Hägerstrand’s conceptualisation of a space-time path (Hägerstrand 1970), the daily life of an individual is embedded in a set of complex spatial and temporal arrangements that determine an individual’s opportunities to engage in activities. This constitutes travel as a derived demand. As Pas (1980, p. 4) has aptly phrased: "If all activities in which an individual wished to participate were located at the same place, that individual would be expected to undertake little or no travel at all." Whilst a number of missing data points might not be directly visible on a Cartesian plane, particularly when they occur at the same geographical location, they become visible in a 3D-representation of the participants’ movements with time as a third dimension. The accuracy of the collected data therefore also depends on the temporal interval of the measurements. This is illustrated by Figure 2 in which the space-time paths, between 08h00 and 18h00, of four participants are visualised.
Figure 2 | Spatiotemporal representation of four GPS-tracks (08h00 – 18h00)

From the graphs in Figure 2 it becomes apparent that the temporal dimension of the data quality differs significantly between the four individuals. Whereas graph A and B show a rather uninterrupted path through time and space, graph C and D show a dispersed set of recorded data points. When we turn to graph C, there is only one measurement around 11h30, and a continuous stream of measurements from 13h30 until 18h00. In graph D on the other hand, most of the data points during the day are missing. This is problematic because the activity locations the participant might have visited during those hours are not captured. In fact, if one would purely look at the geographical location, it seems that the individual did not engage in any travel during the day. If it were not for the time element this measurement gap could go unnoticed. In turn, this could lead to incorrect conclusions with regard to his or her travel behaviour.

Although not per se problematic with regard to determining an individual’s activity pattern, also graph A and B manifest some issues. In graph A, for instance, it is very clear that there is noise in the GPS-measurements. This is a problem inherent to GPS Technology: even when an individual is not moving, the GPS might record a change in location as a consequence of the configuration of the GPS satellites or a signal being reflected by an obstacle such as a building; a problem typically referred to as positional drift. Lastly, also graph B shows a number of gaps that can most likely be ascribed to signal blockage. This implies that the reliability of the captured locations should be critically reviewed and data heuristics should be employed to tease out the most likely activity patterns.
In the final part of the experiment a small exit survey was distributed amongst the participants in order to capture their experience with the tracking process and the functioning of the application. The participants who responded to this exit survey (n = 7) were generally rather positive with regard to the GPS-tracking exercise. One participant commented: “I actually enjoyed taking part in the tracking and would certainly take part in any future studies if required.” Privacy concerns did not seem to be an issue either. All of the participants (n = 5) who responded to this item, stated that they did not experience an intrusion of their privacy or a discomfort with regard to sharing their locational information.

All respondents (n = 7) did express a concern with regard to the battery level and stated they encountered a significant impact on the battery life; something often reported on in the academic literature. This concern is not unfounded: in a battery consumption test it was found that the GPS-services TrackLog requires indeed considerable amounts of battery power. Using a Motorola Moto E as testing device, it was found that if the phone was put on standby for 24-hours the battery drained to 71% of its capacity. If the phone was put on standby, but with TrackLog running, the battery drained to 33% of its capacity.\(^1\) It should be noted that these numbers are contingent on and vary with the type of phone, user settings and usage characteristics. Notwithstanding this inconvenience, the majority (n = 6) stated that the TrackLog application did not interfere with the normal usage of their smartphone devices.

5. CONCLUSION AND DISCUSSION: LESSONS LEARNT

This work described ongoing efforts to assess the reliability and feasibility of smartphone tracking, in the context of South Africa, to suggest ways in which location aware technology can be employed in collecting data on individual activity and travel behaviour. The results of the small-scale pilot study indicate that while GPS technology in smartphones potentially holds a number of benefits for collecting activity and travel data, the technology is not without problems. This project has highlighted that these problems can be classified as: (1) user, (2) technology, and (3) methodology related problems.

Despite the fact that the application seemed to be quite well-received by the participants, a number of users struggled with installing and activating the application on their smartphone. Some users are thus more smartphone literate than others, and care should be taken when designing future experiments. A related problem is that users tend to switch off TrackLog when they are home, which increases the chance that they will forget to switch it back on when they leave their homes again. The incorporation of, for instance, an ‘active tracking’ function could solve this problem by signalling the researchers that a user forgot to switch on the application. Another option currently being experimented with, is the usage of Near Field Communication (NFC) tags that automatically switch on the application. If one were to place a NFC tag on, for instance, a smartphone car mount, the application would not only switch on automatically, but it would also automatically record the exact starting time of a car trip.

\(^1\) Motorola XT1021 running Android 4.4.4 with a Lithium-ion battery with a capacity of 1980 mAh. Wi-Fi disabled. No other functions of the phone were being used during the test.
Whereas none of the participants in this study expressed concerns with regard to their privacy, this does not imply that other people neither have these concerns. In an attempt to mitigate these concerns it is therefore recommended that a research project involving GPS technology should come with a clear description on what happens to the locational information being recorded and on who has access to this information. A concern that did emerge amongst the participants was related to the relative high impact on the battery level. This should not just be regarded as a trivial complication; participants may actually drop out of the study to avoid their smartphone running out of battery. The battery consumption is, however, not solely a user related problem, but also a technology related problem. Whereas it is hard to control the battery consumption of location services in general, the academic literature suggests the usage of additional sensors in order to capture location data more energy efficiently – for instance, by only recording a location when the user is actually on the move.

Technological problems not only surface on the data collection side, but also on the data analysis side. Whilst in a small experiment such as this with only 15,000 measurements the number of measurements is manageable, a large-scale study on activity and travel behaviour of, for instance, 500 individuals for the course of five days with a measurement frequency of 30 seconds, could potentially lead to a dataset comprising of 7.2 million measurements. When analysing such large numbers of measurements, one easily moves into the realm of big data analysis and significant requirements with regard to computing power. In addition, the automatisation of data cleaning processes, data heuristics and data analysis becomes imperative.

Aside from methodological problems with data analysis, methodological problems emerge furthermore from the measurement instrument itself. As only persons with a smartphone are eligible to participate, this leads to questions regarding the representativeness of the collected data. This is also related to different smartphone markets. In the context of this study, it meant that only individuals with Android-based operating systems qualified to participate. This not only requires additional attention to the research design, but might also call for the segmentation of the sample population whereby different groups within a population are targeted with different data collection instruments. These issues should be addressed before smartphone technology can be used as a large scale data collection technique.

Notwithstanding these problems, the collection of high-accuracy data using smartphones to address a number of the requirements imposed by activity-based analysis and spatiotemporal travel behaviour seems promising; both given the possibility of collecting data that would have been otherwise extremely difficult to acquire and because of the spatiotemporal detail with which locational information is recorded. Maybe sophisticated tracking systems are not the perfect way of gathering truly reliable origin-destination data, if that is possible at all, but not only can they improve accuracy, they also open a whole new range of possible applications that were previously unattainable. One can think of accurately capturing both financial and environmental costs of individual travel, precise route reconstruction that can be used in the calibration of traffic models, longitudinal studies with a decreased burden on the respondent, the measurement of behavioural changes that require a high level of spatial detail, a reduction in the reliance on inadequacies of retrospective surveys, increased access to hard to reach groups and a cost reduction of large scale travel surveys. Although GPS-equipped smartphones may not be a perfect
instrument for measuring activity and travel behaviour yet, they are definitely part of the future of transport research.

ACKNOWLEDGEMENTS

The financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the authors and are not necessarily to be attributed to the NRF. The authors would also like to thank the two reviewers for their valuable comments.

6. REFERENCES


