ITLS

**WORKING PAPER**

**ITLS-WP-10-19**

# Development of a GPS/GPRS prompted-recall solution for longitudinal driving behaviour studies

By

**Stephen Greaves, Simon Fifer, Richard Ellison, Yun Zhang and George Germanos**[1]

[1]**Smart Car Technologies**
**Sydney Australia**

**October 2010**

**ISSN 1832-570X**

**INSTITUTE of TRANSPORT and LOGISTICS STUDIES**
The Australian Key Centre in
Transport and Logistics Management

The University of Sydney
*Established under the Australian Research Council's Key Centre Program.*

| | |
|---|---|
| **NUMBER:** | Working Paper ITLS-WP-10-19 |
| **TITLE:** | **Development of a GPS/GPRS prompted-recall solution for longitudinal driving behaviour studies** |
| **ABSTRACT:** | This paper details the development of a GPS/GPRS data collection solution for a longitudinal (twelve week) study of driving behaviour in Sydney, investigating behavioural responses to variable rate charging. The study calls for data to be regularly downloaded to check the quality of data as it is being collected and provide the basis for a web-based prompted recall (PR) survey in which participants can view their trips, confirm details and provide information on who was driving, number of passengers and trip purpose. Following details of the technological setup, we detail the data processing issues involved and the development of the PR survey. Pilot testing of the approach on thirty motorists demonstrates that contrary to popular belief, data of this nature can be collected for several weeks with little respondent burden at high levels of accuracy. |
| **KEY WORDS:** | ***Travel surveys, transport, GPS/GPRS applications, prompted recalled surveys, longitudinal driving behaviour study*** |
| **AUTHORS:** | Stephen Greaves, Simon Fifer, Richard Ellison, Yun Zhang, and George Germanos |
| **CONTACT:** | Institute of Transport and Logistics Studies (C37)<br>The Australian Key Centre in Transport Management<br>The University of Sydney NSW 2006 Australia<br><br>Telephone: +61 9351 0071<br>Facsimile: +61 9351 0088<br>E-mail: itls@sydney.edu.au<br>Internet: http://www.sydney.edu.au/business/itls |
| **DATE:** | October 2010 |

# 1. Introduction

Since the mid-1990s, the potential for using GPS technology to automate the collection of travel survey data that previously relied on manual recall/recording methods has been increasingly realised. No fewer than sixteen regional household surveys in the United States have used GPS technology (NuStats 2008) in some context and at time of writing the first survey to be done entirely using GPS is in the field in Ohio. Driving behaviour studies have also proven a popular application for GPS technology, particularly in pay-as-you-drive applications (e.g., Elango et al., 2007; Nielson 2004). Here, in addition to the traditional elements of origins, destinations, times, distances and routes, detailed information and speeds discernible from GPS have added to the appeal and potential for using the technology (Mazureck and van Hattern, 2006; Greaves and Somers, 2003).

Building on the use of GPS as a means to improve existing data collection efforts, is a growing recognition of the potential of GPS to open up new possibilities in our understanding of travel behaviour. First, GPS provides the potential to extend the period of data collection with (arguably) little additional respondent burden. For instance, applications in which a GPS is installed within a vehicle have successfully collected data for several weeks (e.g., Nielsen, 2004) and in the case of the Atlanta Commute project, multiple years (Elango et al., 2007). Other researchers have demonstrated that well-designed personal GPS systems can be used to collect data for around one month (Stopher et al., 2008a; Li and Shalaby, 2008). Second, through integration with GIS, GPS has facilitated the potential to go back to participants to both confirm trip details are correct and prompt them for information such as trip purpose, who was driving, number of passengers etc (e.g., Stopher and Collins, 2005; Doherty et al. 2006; Li and Shalaby 2008; Auld et al. 2009).

Given this context, the current paper details the development of a GPS prompted-recall (PR) data collection effort to support a longitudinal study of driving behaviour in Sydney. Briefly, the aim of the study is to facilitate, predict and detect changes in driving and encourage safer driving practices through kilometre-based charges. These charges will vary based on the drivers themselves and how much, when and how they drive (specifically speeding). The study calls for a six-week ‘before’ period of monitoring to establish how motorists drive normally, followed by a six-week ‘after’ period of monitoring in which charges are levied and changes assessed. Incentives are paid to motorists for the difference in the charges between the two six-week periods. The data requirements to support this project call for 12 weeks of detailed driving usage data (Vehicle Kilometres of Travel, speeds, routes, times), daily downloads and checks to verify the quality of the data as it is being collected, a mechanism for contacting participants to verify data and device issues, and additional information from participants on trip purpose, who was driving and numbers of passengers. Following a review of recent GPS PR surveys, the paper details the technological configuration for the study, data processing issues involved and the development of the PR survey. The approach is tested on 30 Sydney motorists for a period of eight weeks, to gauge their acceptance of the technology and use of the PR over this extended a period of time before drawing conclusions on the merits of the methodology.

# 2. Literature review

The rationale and arguments for using GPS to support transport data collection as well as the challenges involved are well acknowledged among the research community (e.g., Stopher and Collins, 2005; Wolf, 2006). Rather than repeat this information, the focus of this review is on specific issues related to GPS-based data collection that are of relevance to this paper, namely 1) the ability to collect data over an extended period of time, meaning several weeks if not longer, 2) the potential to use the data to go back to participants to ‘prompt’ them to correct GPS-based information and/or provide information that cannot be directly derived from the GPS data, and 3) the capability to collect elements about speeds and speeding.

## *2.1 Longitudinal GPS studies*

What might best be described as the 'conventional' use of GPS in surveys of travel behaviour is to check/validate information coming from a travel diary, typically collected for one or two days (NuStats, 2008). Extensions of the GPS component to capture more days was originally driven to try to counter the so-called 'halo' effect – here the device is given to people before their survey day so that any unusual behaviours inadvertently generated by the novelty of having a GPS device and (hopefully) not repeated on the survey day itself (Stopher and Greaves, 2009). Taking this further, was the realisation that the marginal cost of collecting more days of data using GPS was actually very low, particularly when considering the expanded benefits in terms of reduced sample size requirements and greater richness of data on multi-day variability in travel behaviour (Stopher et al., 2008b).

The main draw-back of course to multi-day surveys (GPS or not) is the additional participant expectations of taking a device with them for several days if not weeks. Here, it is crucial to distinguish between devices installed in a vehicle versus devices that are carried by a participant. In the former case, devices are typically hard-wired into the vehicle or powered through the cigarette lighter and designed to either switch on/off automatically with the engine or infer trips based on movement (Elango et al., 2007; Nielson 2004). The issue of participant burden in this case is little/none other than checking the device has not been inadvertently knocked or removed. In the case of personal devices, the burden is higher in the sense participants need to remember to take the device with them, but again devices have been designed that require very little of the participant (Stopher et al. 2008a). A final point to make here is that little/nothing appears to be documented about willingness to participate in a long-duration study including whether certain demographic/personality types are more likely to want to participate and what biases this might create.

## *2.2 GPS and prompted-recall surveys*

Prompted-recall (PR) surveys are designed to replay trip information to respondents and ask them questions which validate and/or supplement the collected data (e.g., trip purpose, who was driving, number of passengers, costs etc). They can be administered via telephone, face-to-face or (increasingly) the Internet. Since the concept was first explored (Bachu et al. 2001), PR surveys have become increasingly integral to passive GPS surveys[1] for several reasons: First, is the ability to easily create maps, which are in themselves believed to be useful memory joggers although there is some debate in the profession about exactly how crucial maps are for this process (e.g., Doherty et al. 2006). Second, the global trend in Internet penetration (particularly high-speed[2]) has opened up new possibilities for prompted-recall surveys both in terms of how information is presented and participant interaction is facilitated. Third, recognising the need to present the information as close to the actual time and date of travel as possible, advances in telecommunications have allowed for wireless transfer of GPS data to centralised servers, which can be processed and presented to respondents in the form of a PR survey within 24 hours of completing travel (Marca, 2002; Doherty et al. 2006).

Doherty et al. (2006) describe PR surveys as being either or a combination of sequential, spatial, or temporal/tabular formats. The most popular approaches generally present a temporal calendar style format with a spatial (map) section included for memory assistance. Table 1 (shown at the end of the paper) outlines the key published studies to date in this field. Most of these PR studies are limited in sample size and duration due to the exploratory nature of the objectives. The information collected using GPS PR surveys is very similar across the current reported studies. The key components that are often included are questions surrounding activity purpose and other persons involved in the activity (e.g., passengers if driving). In some of the more detailed PR studies, questions surrounding route choice and activity planning have also been included (Auld et al. 2009).

[1] Passive GPS data collection refers to GPS data that is collected and processed with little respondent interaction at the time of collection. Active GPS surveys require respondent interaction most commonly through a PDA or other interface.

[2] As of 2007-08, 52% of Australian households have high-speed Internet connection (ABS 8146.0 - Household Use of Information Technology, Australia, 2007-08).

A largely unexplored issue of great relevance to the current study is respondent burden in long-duration PR studies and how it could be managed/reduced. To the authors' knowledge, the longest published study is that of Li and Shalaby in 2008, which was completed by 15 respondents and lasted 35 days in duration. The authors use learning algorithms to infer activity purpose cutting down on participant time to complete the PR, an approach which has been used by other researchers (Doherty et al., 2006; Auld et al., 2009). It is not clear from any of these studies what impact these changes have on the participant's willingness to complete the study and what the impacts are on completion time, particularly given the inevitable false-positives from such a system.

### *2.3 GPS and (speeding) behaviour*

Three fundamental issues surround the use of GPS for the measurement of speeds and speeding in particular. First, is the issue of accurate derivations of speed from the GPS device itself, something which has been the subject of considerable investigation (e.g., Greaves and Somers, 2003). GPS Doppler speeds in themselves are generally highly accurate (±0.1 m/s is what most manufacturers will claim), but the main problems come from the lag (typically one to two seconds) and missing data points meaning that some sort of inference or smoothing algorithm is typically used (Jun et al. 2006). Second is the accuracy, comprehensiveness and availability of a digital representation of the speed limit street network not just spatially but temporally (e.g., school zones, variable speed limits). Third, is the specific problem of accurately matching GPS data points to the correct speed limits, which is typically not dealt with explicitly in map-matching algorithms or if it is, does not appear to be well documented.

# 3. System configuration

The system configuration for the survey involved various elements working together to provide the desired outputs (Figure 1). Motorists were given a GPS data logger, which transmitted standard NMEA sentences in near real-time by wireless communication to our project partner, Smart Car Technologies (SCT) central server for map-matching (including the tagging of speed limit) and processing into trip records. Processed trip summary files and second-by-second GPS files were then transmitted nightly to the University of Sydney Webserver, where they were downloaded to form the basis for an Internet-based PR survey. Participants could then log in to the survey and provide the required trip information, which would be automatically written to the database. Each of the system elements are now described in more detail.

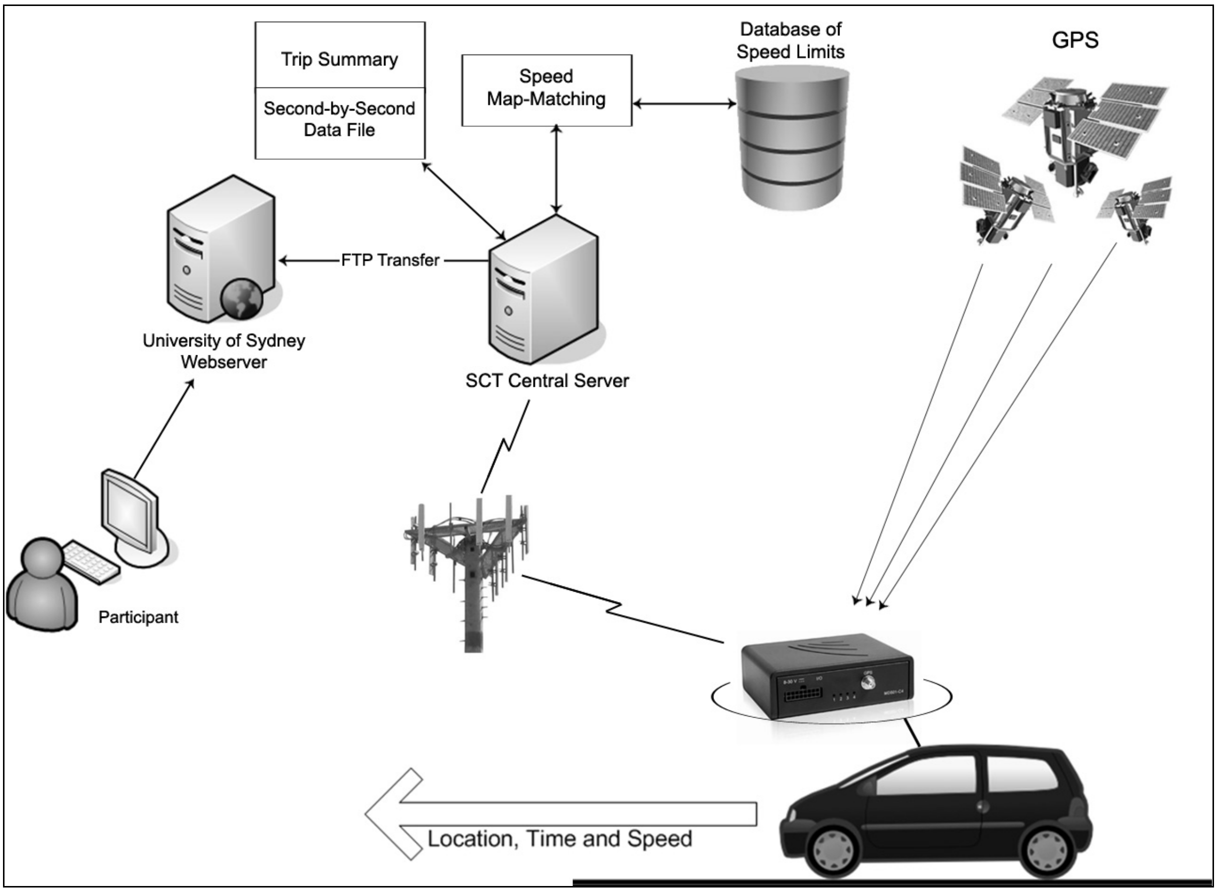


***Figure 1: Overview of the system setup***

## *3.1 The GPS device*

The specifications for a suitable GPS device were as follows. First, it needed to be simple to install in a motor vehicle, as unobtrusive as possible, and provide the required information with no involvement from a participant (i.e., passive). Second, it needed to be able to provide second-by-second GPS data for accurate speed information used in the assessment of speeding. Third, data needed to be regularly downloaded to check both the quality of the data as it was being collected and provide the basis for the daily update of the prompted-recall survey. Finally, it needed to be priced appropriately given the available project resources.

After an extensive search and testing of several devices, the C4 Mobile Device manufactured by Mobile Devices Ingenierie was selected (Figure 2)3. The C4 is a small (8.2cm x 10.5cm x 3.5cm) Sirf III GPS device that can (among many other features) be configured to regularly communicate second-by-second GPS data by GPRS (as little as ten second intervals if required) to a server for processing4. For this application, data were transferred every 20 seconds mainly to provide sufficient time to transmit the last package of data at the end of a trip before the GPS went into sleep mode. The device was set up to be powered from the cigarette lighter such that the GPS would 'wake up' when the engine ignition was switched on and 'sleep' when the engine was switched off to enable the automatic detection of trip-ends. The device could also have been configured to use an in-built motion-sensor to detect trip-ends, with the advantage here being it is more likely to detect trip-ends where the engine is still running (e.g., drive-throughs, dropping someone off). However, testing resulted in several false positives (e.g., sitting at traffic signals as a trip-end) and missed several genuine trip-ends and this was rejected as an approach.

[3] http://www.mobile-devices.fr/index.htm

[4] The device could also have been configured to communicate by SMS. However, this worked out at 25 cents per message and was prohibitively expensive for this application. GPRS requires a mobile phone plan, which was $10/device/month for this study.


Data Logger In-vehicle charger Antenna

***Figure 2: The C4 device***

## *3.2 Data processing and quality checks*

The data processing component of the project involved various phases. First, the regular transmission of the data enabled real-time monitoring through Google Earth to ensure the system was always working and to quickly address any general problems that arose. It must be stressed here participant privacy was strictly maintained in that details of participants were only accessible to the University of Sydney research team.

Second, the GPS data were map-matched to a GIS-based representation of the Sydney street network, which includes accurate speed limits, collected by SCT. The speed limit database has been developed from the ground up by driving all the streets in Sydney and includes temporal variations in speed limits such as school zones[5]. The map-matching algorithm includes several features designed to overcome problems that are particularly pertinent to the accurate detection of speeding including the GPS lag problem and the incorrect assignment of links (and hence speed limits). The GPS lag problem refers to the fact that GPS speeds typically have a lag of 1-2 seconds. This is addressed by matching GPS points to the speed segment 1-2 seconds in advance based on the speed of the vehicle and assumed distance that would be covered. The incorrect assignment of GPS points to links is most associated with roads in close proximity (e.g., intersections, off-ramps, service roads) and is generally caused by a combination of the way that GPS points are ‘snapped’ to the nearest line and the wander associated with these points. This is overcome by building the network representation up in more detail and looking at previous GPS points, to maximise the probability that the point is snapped to the correct segment.

Once the GPS data were correctly map-matched, these were then aggregated into trip files, which provided summary information (origin, destination, distance, time, speeding behaviour etc) for each trip undertaken by a participant for a particular day. The summary files were downloaded nightly (automatically through a batch process job) with the second-by-second GPS data to the University of Sydney Webserver. This download procedure included the first phase of a data checking component via a short report that covered issues relating to:

*Participants* (number of trips for the previous day, how long since they had logged into the PR survey, days since last GPS activity),

*Very short trips* such as moving a car in a drive-way (defined as less than 100 metres), and

*Potential missing trips* (defined as one trip-end starting more than one kilometre from the previous trip-end).

---

[5] School zones typically operate from 8:00 a.m. – 9:30 and from 2:00 p.m. – 4 p.m. in Sydney during which time the speed limit is reduced to 40 km/h.

These flags assisted the manual checking component conducted as part of the survey management process detailed in Section 3.4.

### *3.3 The internet-based prompted recall survey*

The specifications for the PR survey were that it needed to be simple for participants to use, avoid long refresh times and appealing and attractive to keep them motivated for several weeks. Following extensive testing and feedback from academic colleagues[6] and members of the public, the interface shown in Figure 3 was developed. Key elements of the interface are a familiar tabular format that is quick and intuitive to use, the use of open-source mapping software (Google Maps) that refreshes comparatively quickly, and optional viewing of trips. This last point emerged as important because of the slow refresh speed of Internet Explorer (still the most widely used web browser) even using a high-speed connection combined with the fact that the repetitive nature of most travel implied participants did not need to see a visual depiction for most of the trips to jog their memory on the other attributes.

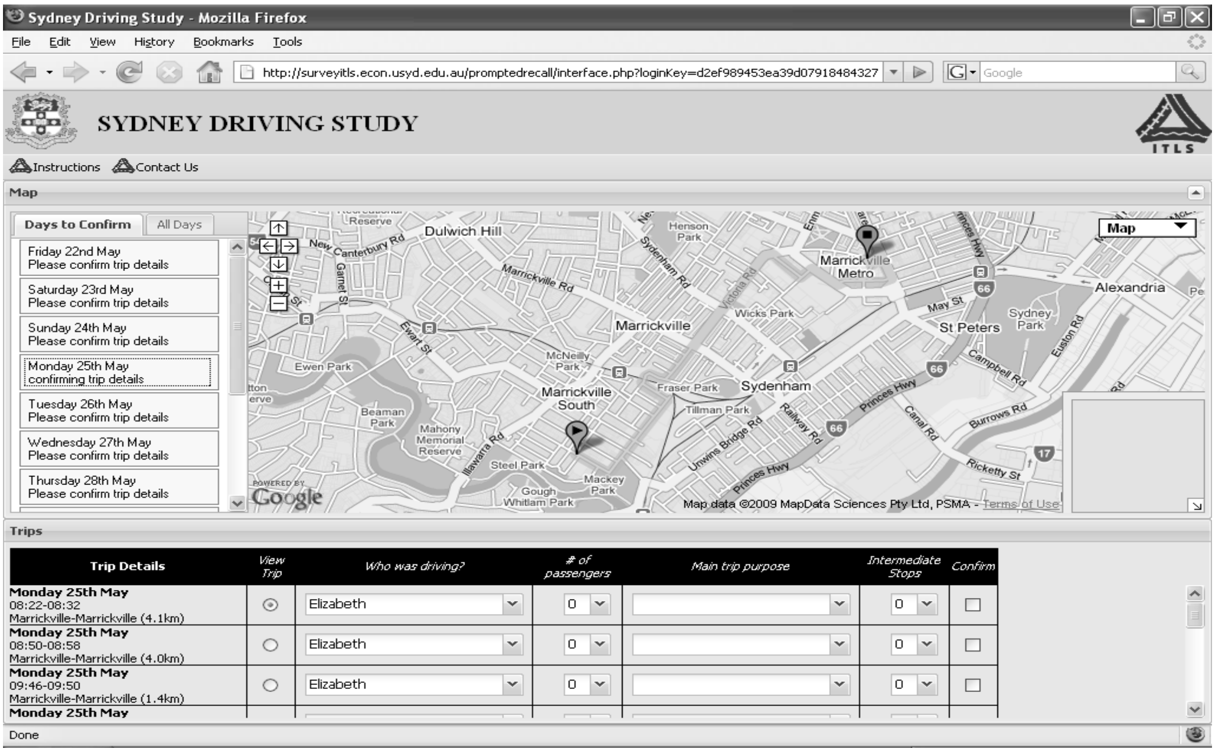


***Figure 3: Participant prompted recall interface***

To access the interface, participants were sent a URL via the survey management interface (see next Section), which was unique to them (they were advised to Bookmark) and took them straight to their trips without having to remember logins and passwords[7]. Once in the interface, clicking on a day under ‘Days to Confirm’ brings up the trip list for that day underneath the map. Participants can view the trip to ‘prompt’ their memory and then fill in some simple information on the trip; namely who was driving, the number of passengers, the main trip purpose and whether any intermediate stops were made (e.g., dropping off children at school on the way to work). At the bottom of the trip list is a dialogue box enabling participants to record any missing trips or other data issues they notice. Once they have finished the trip is confirmed, meaning that data is written to the database (they can un-confirm and change information) until all days are completed.

[6] The authors’ would like to acknowledge in particular the suggestions of Prof. Sean Doherty during the development of the interface.

[7] The login key is generated by running a cryptographic hash (MD5) on a random sequence of 32 characters which is then checked to make sure it is unique to that specific participant.

### *3.4 Survey management*

Among the unique features of this survey were the survey management system employed by the University of Sydney research team. A similar interface to the one shown in Figure 3 was developed, which was only accessible to survey managers (Figure 4) and enabled them to do the following:

Check the general status of each participant including last GPS update, time since last PR login and where the GPS devices were currently located. Participants would be sent a reminder e-mail if there had been no GPS activity for more than 72 hours and/or they had not logged onto the PR for more than seven days. This would be followed up by a phone call if necessary.

Check short trips and missing trips flagged as part of the daily report.

Check missing trips flagged by participants.

Export the GPS second-by-second data and trip data by date range and/or participants as required into a csv format for further analysis.

E-mail participants with reminders and their unique URL (generated through encryption programming) for logging into the PR survey.

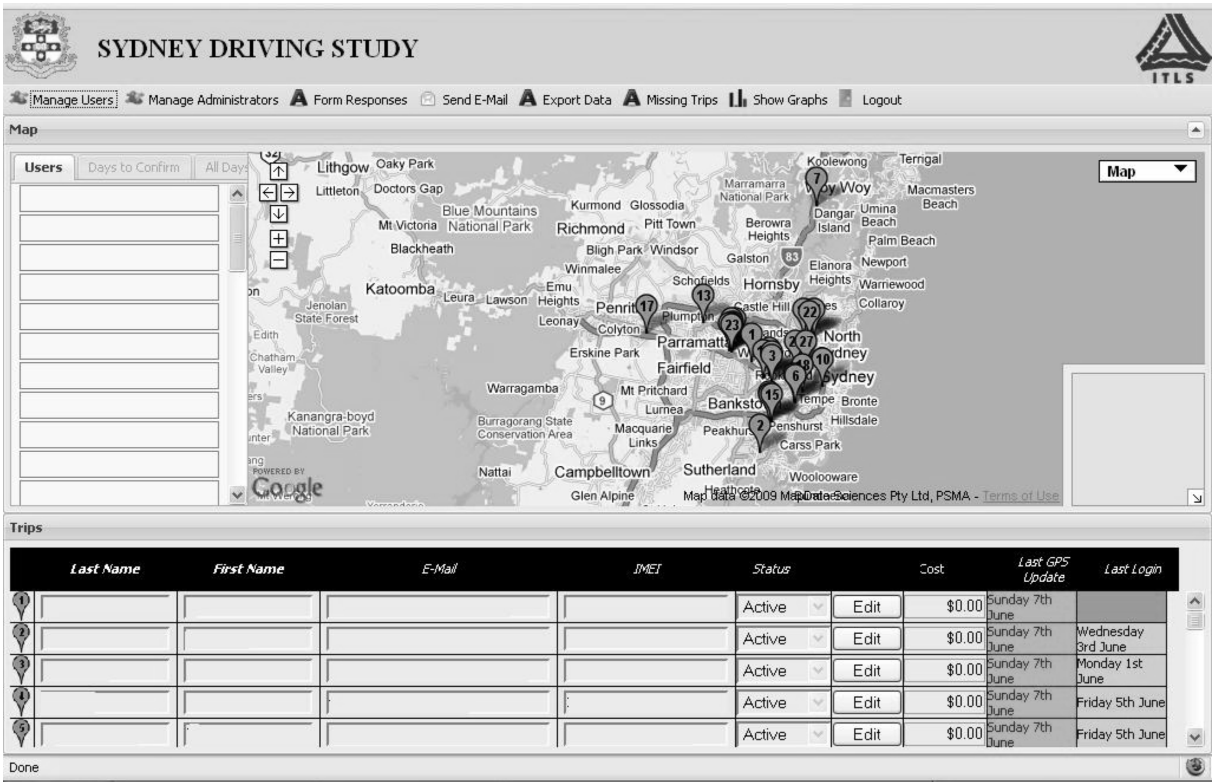


*Figure 4: The survey administration interface*

# 4. Testing of the configuration

Thirty participants were recruited from four suburbs (Chatswood, Hurstville, Parramatta, Strathfield) for the pilot testing phase of this project. The sample comprised 12 males and 18 females ranging in age from 19 to 61. Participants were recruited according to strict criteria that reflected the main aims of the study as well as practicalities about using the equipment. In terms of the main aims of the study, only participants with a valid licence from one-car households were recruited8 and they needed to be the primary driver and drive more than two days per week on average. In terms of practicalities, cars

[8] The proportions of one-car households in the selected suburbs were Chatswood (48%), Hurstville (46%), Parramatta (50%), and Strathfield (35%).

needed a working cigarette lighter, which did not stay on when the engine turned off (a problem for a very small proportion of high-end vehicles in Australia) and needed to be parked off-street at night. Unfortunately, this last criterion was imposed from an insurance/liability perspective as ‘smash and grabs’ of electronic equipment from motor vehicles is common in Sydney.

Both courier and face-to-face methods were used for testing delivery of the device. Despite the added expense for face-to-face (about $50/participant versus $10/participant for courier), this method emerged as more favourable for several reasons. First, the device could be installed by the recruiter – even though the installation was simple, never-the-less, it was critical to ensure this was done correctly. Second, odometer information (used to cross-check the distance information coming off the GPS device) could be collected and consent forms signed rather than having to wait for people to return this information. Third, because of the way couriers schedule deliveries (they are paid based on attempted deliveries if a signature is required) and specify wide time-windows in Sydney, it proved very difficult to ensure the devices were delivered safely and/or cost-effectively. Finally, and most importantly, the survey required considerable ‘buy-in’ from participants, which was more effectively achieved using face-to-face trained interviewers.

## *4.1 Data quality*

The data were generally of an extremely high quality with 97 percent of trips captured by the GPS device. The three percent of trips classified as definitely missing (either inferred as missing and/or confirmed by participants via the dialogue box in the trip interface) were attributed to connection issues associated with the in-vehicle charger coming loose from the socket (deliberately or accidentally removed) but generally an e-mail to the participant resolved this issue. Invariably, these issues occurred early on and because of the data management system in place, issues were quickly identified and dealt with, a key to the success of the data collection effort.

The other main issue affecting data quality was the perennial ‘cold-start’ problem (Greaves and Somers, 2003; Stopher and Collins, 2005). The issue here is that when a GPS receiver loses connection with the satellites (such as when a trip ends) it has to re-acquire contact to accurately compute position. The longer the time between trip-ends, the longer it takes the receiver to re-acquire satellites to compute position. Up to a certain time interval, the GPS uses information on its last known position to compute position, known as a ‘warm start’, which means position is re-acquired quickly (within a few seconds). However, after a certain amount of time, the receiver has to literally start over again in determining position, known as a ‘cold-start’. The ‘cold-start’ acquisition time is also negatively affected by whether a vehicle is in motion and any blockages to satellites such as trees and heavily built-up areas.

The pre-testing and subsequent pilot testing of the device, showed the first trip of the day was particularly susceptible to this cold-start problem (reflecting long time gaps, in general, from the last trip of the day before), with trips affected by anywhere from 30 seconds to two minutes. This meant that in the computation of distance, the missing segment (i.e., previous trip-end to start of new trip) was inferred back to improve the accuracy of the VKT computations. Putting some numbers on this, using the ‘raw’ GPS data, VKT was under-estimated by an average of 12% across the 30 participants (compared to self-reported odometer readings, which are in themselves susceptible to error). Inferring back based on a simple straight-line distance reduced this under-estimation to 5% - intuition would suggest this accuracy could be improved further using the network information.

## *4.2 Participant reaction*

Arguably the most pertinent reflection of participant reaction to the survey is that at time of writing out of the thirty participants, only one has dropped out while the remaining 29 have all completed the PR survey for at least four weeks, with six now having completed eight weeks. Reaction to the survey was also gauged through various mechanisms, including e-mail enquiries during the survey, exit surveys conducted on participants at the end of the survey and usage statistics collected from the PR survey. In terms of e-mails, over the course of the eight weeks, 70 e-mails were received, an average of just over two per participant. Ignoring issues to do with device delivery/retrieval and other administrative concerns, 15 of these e-mails were to do with device connection issues and five about the PR. The main issues with the PR was to do with fire-walls (two enquiries), web browser (initial problems with Google Chrome, which were quickly resolved) and refresh rate (only one participant had dial-up). No issues were raised about the prompted-recall itself during the study, which was attested to the ease of use and the clear instructions given.

Face-to-face exit surveys conducted on participants confirmed that both the use of the device and the PR were simple and intuitive. Once installed, participants seemed to forget about the device and continue driving as per normal. It was clear also that participants had enjoyed the study, particularly the ability to visualise their travel. In fact, all participants interviewed indicated that they would happily continue with the study for at least another 4 weeks and some even longer.

Of particular interest in this study was to gauge participant usage of the PR survey, something which to our knowledge has not been formally documented. The use of a Web server combined with some 'clever' programming meant we were able to track this usage automatically, providing unique insights into how often they accessed the PR survey, how long it took to complete, and when people typically completed it. On average, people accessed the interface once every 4.5 days (median was 4 days) with 20% of participants accessing the interface at least four times per week, 80% accessing it at least once per week and 20% accessing it once a fortnight. While the average session length was seven minutes (median of 3.5 minutes), given the completion time depended on how often the interface was accessed and how many trips were made, arguably of more relevance is the time taken to complete details for each trip (shown in Figure 5). Using this metric, the median time for completion was 14 seconds/trip – note the average time of 25 seconds is heavily skewed because of some very high values due to people probably being interrupted when doing the PR and leaving the interface open.

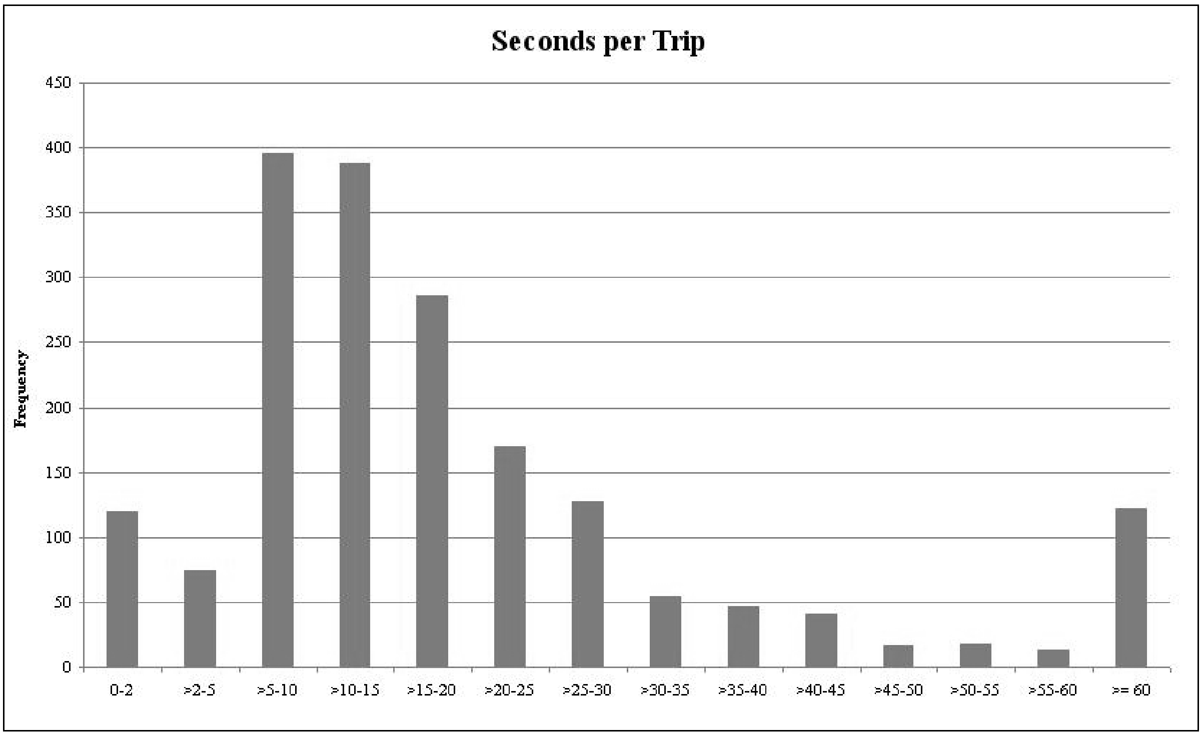


***Figure 5: Time taken to complete the prompted recall per trip***

It was also of interest to determine when people typically completed the PR and how this related to more general Internet usage patterns. Figure 6 shows the distribution of session start times. Evidently, the survey is primarily being completed during work-time hours, particularly between 9 a.m. – 11 a.m., with a lull between 6 p.m. and 8 p.m. (commuting home, eating dinner) followed by another small peak in activity between 9 p.m. and 10 p.m. It is also notable that a small, yet significant number of sessions occur during the late night/early morning. In terms of how this compares to more general Internet usage patterns, a recent study of Internet usage in Australia, shows that usage of the PR follows similar patterns except for a larger drop-off in the early afternoon and the fact that general Internet usage peaks between 7 p.m. – 8 p.m. (Spennemann, 2006). This suggests that in general, completion of the PR is part of people's normal web browsing routines, an assertion backed up by evidence from the exit interviews conducted thus far.

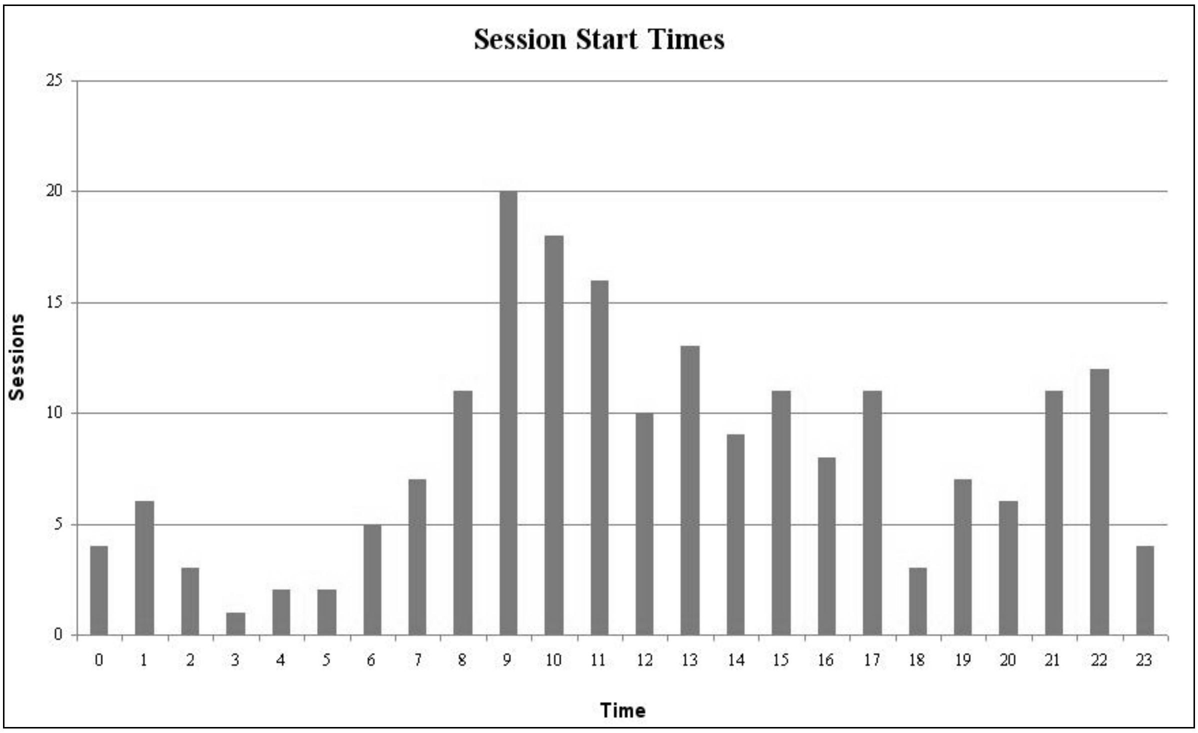


***Figure 6: Session start times for completing the prompted recall survey***

# 5. Conclusions

This paper details the development of a technological solution designed to capture detailed driving information (including speeding) over several weeks as a precursor to investigating behavioural responses to variable rate charges. Among the unique features of the survey are 1) the combination of a longitudinal GPS data collection effort with an interactive web-based PR survey to capture additional trip information, 2) highly accurate assessments of speeding facilitated by a precise spatial and temporal speed limit database, and 3) real-time wireless transfer of data and a survey management system that enables daily checks on the quality of data and the capability to respond quickly to any problems.

Overall the GPS data were of a high quality with only three percent of trips classified as genuinely missing. Cases of inadvertent knocking/removal of the device were a problem for some participants, but the survey management system enabled these problems to be detected and quickly dealt with. It is an important if obvious observation that despite all the best efforts to effectively nullify the opportunity for participants to affect GPS data collection it is (in the authors' opinion) still not possible to do this with 100% guarantee, even if the device is physically installed in the vehicle. The cold-start issue remains an issue (in Australia at least) meaning VKT computed from 'raw' GPS data, will be an under-estimate compared to odometer-based VKT (itself prone to inaccuracy). In the case presented here, VKT estimates from 'raw' GPS data were under-estimated by 12 percent, but this discrepancy was reduced to 5 percent using simple inference of missing trip segments.

Participant reaction to both the GPS and PR components was generally very positive, reflected in exit interviews and the fact that (at time of writing) 29 of the 30 recruits have completed the PR survey for at eight weeks. The success of the PR is attributed to the design of an interface that is quick and appealing to use such that completion has become a regular component of Internet activities.

In terms of wider implications for the transport research community, this study shows that (perhaps) contrary to popular belief, data of this nature can be collected for several weeks with little respondent burden at high levels of accuracy. Key to this is using technologies not only to make data collection easier but also to inform and engage participants in their daily travel-decision making process in a straight-forward and intuitively appealing manner.

# References

Auld, J., C. Williams, A. Mohammadian, and P. Nelson. (2009). An Automated GPS-Based Prompted Recall Survey with Learning Algorithms. Transportation Letters: The International Journal of Transportation Research, 1(1), 59-79.

Australian Bureau of Statistics (ABS). 8146.0 - Household Use of Information Technology, Australia, 2007-08). Accessed on 7/6/09 from

http://www.abs.gov.au/ausstats/abs@.nsf/0e5fa1cc95cd093c4a2568110007852b/acc2d18cc958bc7bca2568a9001393ae!OpenDocument

Bachu, P. K., T. Dudala, and S.M. Kothuri (2001). .Prompted recall in global positioning system survey: Proof-of-concept study. Transportation Research Record: Journal of the Transportation Research Board, 1768, 106-113.

Doherty, S. T., D. Papinski and M. Lee-Gosselin (2006). An Internet-based Prompted Recall Diary with Automated GPS Activity-Trip Detection: System Design. CD-ROM proceedings of the 85th Annual Meeting of the Transportation Research Board, Washington DC, January 2006.

Elango, V., R. Guensler, and J. Ogle (2007). Day-to-Day Travel Variability in the Commute Atlanta, Georgia, Study. Transportation Research Record: Journal of the Transportation Research Board, 2014, 39-49.

Greaves, S.P. and A. Somers (2003). Insights on driver behaviour: What Can Global Positioning System (GPS) Data Tell Us? CD-Rom Proceedings of the 21st ARRB & 11th REAAA Conference, Cairns, 18-23 May 2003.

Jun. J., R. Guensler, and J. Ogle (2006). Smoothing Methods to Minimize Impact of Global Positioning System Random Error on Travel Distance, Speed and Acceleration Profile Estimates. Transportation Research Record: Journal of the Transportation Research Board, 1972, 141-150.

Li, Z.J. and A.S. Shalaby (2008). Web-based GIS System for Prompted Recall of GPS assisted Personal Travel Surveys: System Development and Experimental Study. CD-ROM proceedings of the 87th Annual Meeting of the Transportation Research Board, Washington, D.C., January 2008.

Marca, J.E. (2002). The Design and Implementation of an On-Line Travel and Activity Survey. Center for Activity Systems Analysis. Paper UCI-ITS-AS-WP-02. Accessed on 7/6/09 from: http://repositories.cdlib.org/itsirvine/casa/UCI-ITS-AS-WP-02-1.

Mazureck, U. and J. van Hattern (2006). Rewards for Safe Driving Behaviour: Influence on Following Distance and Speed. Transportation Research Record: Journal of the Transportation Research Board, 1980, 31-38.

Nielsen, O.A. (2004). Behavioural responses to pricing schemes: Description of the Danish AKTA experiment. Journal of Intelligent Transportation Systems, 8(4) 233-251.

NuStats (2008). Non-Response Challenges in GPS-Based Surveys - Resource Paper Prepared for the May 2008 International Steering Committee on Travel Survey Conferences, Annecy, France, August 2008.

Spennemann, D.H.R (2006). The Internet and Daily Life in Australia: An Exploration. The Information Society, 22, 101–110, 2006

Stopher, P.R. and S.P Greaves (2009) Missing and Inaccurate Information from Travel Surveys – Pilot Results. CD-ROM proceedings of the 32nd Australian Transport Research Forum, Auckland, New Zealand, September 2009.

Stopher, P., E. Clifford, and M. Montes (2008a). Variability of Travel over Multiple Days: Analysis of Three Panel Waves, Transportation Research Record 2054, pp. 58-63.

Stopher, P.R., K. Kockleman, S.P. Greaves and E. Clifford (2008b) Reducing Burden and Sample Sizes in Multiday Household Travel Surveys. Transportation Research Record 2064, 12-18.

Stopher, P. and A. Collins (2005). Conducting a GPS Prompted Recall Survey over the Internet. CD-ROM proceedings of the 84th Annual Meeting of the Transportation Research Board, Washington DC, January 2005.

Wolf, J. (2006). Application of New Technologies in Travel Surveys, in Travel Survey Methods -- Quality and Future Directions, P. Stopher and C. Stecher (editors), Elsevier, Oxford

.

***Table 1: Recent GPS/PR surveys***

| Reference | Date | Country | Households | Duration | Tracking | Data | Map | Processing | Recall Method | Questions |
|---|---|---|---|---|---|---|---|---|---|---|
| Bachu et al. (2001). | 2001 | US | 10 | 2-3 days | Vehicle | Manual upload | Static | Basic | Face to face interview | Trip purpose?<br>Vehicle occupancy?<br>Number of family members? |
| Marca (2002). | 2002 | US | Pilot | *Not specified* | Vehicle | Wireless | Static | Basic | Internet Self completion | Traffic conditions?<br>Important route decisions?<br>Location name?<br>Activity name?<br>People involved in activity? |
| Stopher and Collins (2005) | 2005 | Australia | 64 | 1 day | Vehicle | Manual upload | Interactive | Basic | Mixed methodology (including internet) | Trip purpose?<br>Who was the driver?<br>Number of passengers?<br>Was payment for parking required? |
| Doherty et al., (2006) | 2006 | Canada | Example cases | *Not specified* | Person | Wireless | Interactive | Predictive | Internet Self completion | Event type?<br>Location?<br>Involved persons/passengers? |
| Li and Shalaby (2008) | 2008 | Canada | 15 | 35 days | Person | Manual upload | Interactive | Predictive | Internet assisted / self completion | Location and trip editing?<br>Activity type and details?<br>Mode?<br>Trip purpose? |
| Auld et al. (2008) | 2008 | US | 5 | 15 days | Person | *Not specified* | Interactive | Predictive | Internet Self completion | Type of activity?<br>Planning time horizon?<br>Other persons involved?<br>Locations available for this activity?<br>Timing decisions for activity?<br>Reasons for mode and route choice? |