WORKING PAPER
ITLS-WP-08-06

Deducing mode and purpose from GPS data

By

Peter Stopher, Eoin Clifford, Jun Zhang & Camden FitzGerald*

*Parsons Brinkerhoff

April 2008

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.
Deducing mode and purpose from GPS data

Over the past several years, there has been increasing interest in the use of Global Position System (GPS) devices as a means to measure people’s travel. These devices have been used quite widely to check the validity of self-reports of travel through travel diaries and similar conventional travel survey procedures. With the development of lightweight, high sensitivity GPS devices, there has been increasing interest in their use. GPS devices can collect and record data on points during a person’s travel, providing the latitude, longitude, (and altitude, if desired), the time, the speed of motion, and the heading at each such point, along with data quality measures. Data can be collected as frequently as every second, thereby providing a clear documentation of the path travelled. GPS devices cannot, however, collect data on the mode or the purpose of travel, both of which are essential components of the data record required for transport planning purposes. In this paper, we describe work that we have recently undertaken in which we have developed a set of heuristic rules for determining each of the mode of travel and the purpose of the trip, albeit with the use of comprehensive GIS data bases for the areas where the GPS surveys have been conducted.

In the case of mode of travel, it is necessary to have a GIS that includes not only the street system, but also the bus routes and any rail lines or other modes of travel (such as ferry in the case of Sydney), where those modes do not travel along the street system. The rules we have developed consider the average speeds, the maximum and minimum speeds, and also use certain procedures to eliminate potentially spurious spot speed information. They also use information about the transport network and the availability of bicycles and cars to the survey participants. In the case of trip purpose, it is necessary to have a GIS of the land use of all parcels of land in the urban area for the survey, and we also ask people to tell us the addresses of all workplaces of members of the household, the school addresses for schools attended by household members, and the two most frequently used grocery stores or supermarkets. The rules then use the length of stay at the location and whether or not the location is one of the workplaces, schools, or stores identified. Frequency of visits during one or more weeks of data recording is also used to help identify the purpose of the trip. For trips that have one or both ends at a location other than home or one of the given addresses, the purpose is deduced from the land use information and the duration and frequency of the visit.

KEY WORDS: Travel surveys, GPS, GIS, trips, travel mode, trip purpose

AUTHORS: Peter Stopher, Eoin Clifford, Jun Zhang & Camden FitzGerald

CONTACT: Institute of Transport and Logistics Studies (C37)
The Australian Key Centre in Transport Management
The University of Sydney  NSW  2006  Australia
Telephone: +61 9351 0071
Facsimile: +61 9351 0088
E-mail: itlsinfo@itls.usyd.edu.au
Internet: http://www.itls.usyd.edu.au

DATE: April 2008
1. Introduction

Since the mid-1990s, Global Positioning System (GPS) technology has been applied to various aspects of the measurement of transport-related issues. Initially, GPS devices required an external power source, which limited them to being used in vehicles, and sometimes only in specially-equipped vehicles (1), (2), (3), (4), (5). Subsequently, versions were developed with an attached or built-in power source (6), (7), (8). These devices have been used for a number of alternative applications in transport, ranging from objective measurement of infrastructure (9), to various measurements of traffic flow, system performance, and related phenomena (10), (11), (12), (13), (14), to validating the measurement of personal travel behaviour (5), (15), (16), (17), (18), (19), (23), to evaluating behaviour-change policies (20), (21). Although it has been suggested that GPS devices could take the place of more conventional travel surveys (22), this notion seems to have become lost subsequently in the drive to use GPS as a complement to more traditional methods of conducting household travel surveys (HTSs), and to improve methods of tracking behaviour change.

Recent developments in the technology have produced small and lightweight personal devices that are very sensitive and able to pick up GPS signals in locations that were hitherto not possible for such devices. This includes much better and more accurate recording in urban canyons and areas of dense tree cover, and also even recording in some buildings, shopping centres, and in public transport vehicles. As a result, the time appears to be ripe to reconsider the potential for GPS devices to be used as an alternative to conventional travel surveys.

In considering such a use, however, we have to be conscious of what GPS can and cannot do. GPS is able to collect accurate data on:

- Location at each instant
- Time at each instant
- Speed of movement at each instant
- Heading
- Data quality measures

Data can be collected as frequently as every second, or as far apart in time as the user desires. Ideally, if applied to transport planning, logging devices are all that is required, in which all the data collected at whatever time interval are stored for later retrieval and analysis. Transmission of the data in real time is not generally advantageous or useful to the transport planner and may be expensive to achieve and raise major issues about privacy and intrusion for the user. In our experience, we have found that a device with 4 or 8 Mb of memory is sufficient to store second-by-second travel for several months for the average urban dweller.

However, to substitute for a conventional travel survey, additional data are needed that GPS devices do not collect. Principal among the additional data items are the mode of travel used on each trip segment, and the purpose of each trip. GPS devices alone clearly cannot collect these. However, as this paper shows, it is possible to achieve a
high degree of accuracy in determining these from the GPS record, provided that certain supplemental data are also available.

2. Mode and purpose identification

GPS devices do not provide any information directly on purpose or mode. However, both of these can be inferred, with high accuracy, from the information provided by the GPS device, provided there are adequate GIS databases. To identify mode, the primary items required, in addition to the GPS data, are:

- GIS of the street network
- GIS of all public transport routes (including rail and subway lines)
- GIS of all bus stops and station locations (if possible)

All of these are needed in a common geographic referencing system which also matches well to the GPS data on latitude and longitude. Second, in our GPS surveys, we still collect demographic data, which will be needed in many analysis and modelling situations. In addition, to assist in purpose identification, we collect the following additional information:

- Address of each workplace for each working household member
- Address of each educational establishment for household members engaged in education
- Addresses of the two most frequently used grocery stores

In addition, the home address is already known, for delivery and collection of the GPS devices and other survey materials. All of these locations are then geocoded. Finally, it is of immense help to purpose identification if there is also available a parcel-based GIS of land use, even if it is not fully up to date. However, this will provide good information for determining the purposes of some trips that might not otherwise be readily identified to a trip purpose.

2.1 Preliminary steps

For the purposes of this paper, it is assumed that the GPS data have already been divided into trips (24). Briefly, we define a trip end as occurring any time that there is no movement for a period of 120 seconds or more. Some checks are made to look for exceptions to this, and, after all processing is done, we generate maps of all trips and do a visual check to look for possible missed trip ends and spurious trip ends. Generally, we find that the program misses about 5 percent of actual trip ends, because the stop lasts less than 120 seconds, and also identifies stops longer than 120 seconds as false stops (they are really traffic stops) about 5 percent of the time. Most of these are corrected through the visual checking. It is important to note that the rules we apply to
identifying trips actually identify trip segments. This is also ideal for being able to identify mode, as discussed in the next section of this paper.

3. Mode identification

The identification of travel mode proceeds in a hierarchical process. At the outset, the easiest mode to identify is walk, because of the consistently low speeds for the entire trip segment. This mode is identified principally by considering the 85th percentile highest speed on the trip segment, and then checking the 85th percentile accelerations and decelerations. Off-network public transport modes, such as rail and ferry, are identified next. These are readily identified because of the location of the paths used, which coincide with the rail lines or ferry routes, and not with the street network. Underground rail trips are identified separately, in the trip repair process, when gaps are found in the record that correspond to a rail trip with no GPS data available during the rail trip.

In all cases, when speeds and acceleration or deceleration are used to assist in identifying a mode, we use the 85th percentile value. This helps by removing excessive values that are a function of the GPS device, rather than an actual trace of movement. For example, in second-by-second GPS data, the device may give an occasional point that is located off the route, as a result of momentary problems in solving the position equations. Such off-route points can also produce high figures for speed and acceleration/deceleration. Using the 85th percentile value, while also being consistent with much traffic engineering practice, also ignores these points that are in reality spurious values.

After identifying the walk and off-network public transport, the next mode to be identified is bus. This is based again on maximum (85th percentile) speed, acceleration and deceleration, and on the trip segment beginning and ending on a bus route. If bus stops are available on the GIS, then the bus trip segments can be identified even better by requiring that the trip segment begin and end at a bus stop. An additional check can be made by looking at the trip segment immediately before and immediately after the trip segment that is identified as being a bus. At least one of these must normally be a walk, with a relatively short period of time being spent between the trip segments.

Bicycle trips are identified next. The first check that is done is to determine if the person comes from a household where bicycles are available. If not, then no trip segments are assumed to be by bicycle. If, however, at least one bicycle is owned, then the bicycle trips are identified by examining the maximum speed and acceleration, and that the trip origin is either home or is a location to which this person has already travelled by bicycle. If all of these tests are passed favourably, then the trip segment is allocated to bicycle.

At this point, all remaining trips should be trips by car. However, a further check is made of maximum speed and acceleration, and also that the trip segment remains on the road network. If these are correct, then the trip segment is identified as being by car. We have not pursued the identification of car passenger at this point. However, some car passenger trips could be identified in theory through the following two checks. First, if the person does not hold a current driving license, then any trip segments identified as car would be presumed to be car passenger. Second, if the origin is not home, and car was not used to reach the origin of the trip, but the next trip segment is by car, then car...
passenger would normally be assumed. Also, although we have not yet implemented this, car occupancy of household members could be determined by matching car trip segments among two or more household members by time of day and mode.

4. Trip purpose identification

To identify origins and destinations more easily as noted earlier, we ask respondents to the GPS survey to list workplace addresses for each job held by members of the household. They are also asked to provide the school addresses for each student in the household. Finally, they are asked to provide details of the two most frequently used grocery stores or supermarkets. The home address of each household is known, of course. All of these addresses are geocoded. The first issue to determine is the level of detail desired in the trip purpose identification. If trip purposes are to be defined only to a small set, such as home-based work, home-based education, home-based other, non-home-based (work), and non-home-based (nonwork), then the address information we collect, together with an examination of the frequency over each week that the places are visited and the length of time spent there will largely identify all the trip purposes. It is only if the detail to be obtained on trip purpose is greater than this that significant effort is required to identify purposes.

On the average, about 30 percent of household trips are made to or from either school or work (25), while another 13 percent involve regular shopping, and about 40 percent of the remaining trips will have one end at home. As a result, asking these address questions leads to the identification of about 70 percent of the trip purposes, and may result in identifying almost all trip purposes in a simple three to five purpose schema. If more detail is required, then our procedure follows the steps shown in Figure 1.
For greater detail or for the remaining trips whose purpose has not been identified, three pieces of information can be used that are available in the GPS and which should be available for the region where the GPS survey is undertaken. First is the duration of the activity or stop, which is the difference between the time the trip to the location ends and the following trip from the same location starts, recorded by the GPS device. Second is the frequency that the location is visited within the multiple days of the GPS survey. (Any GPS survey is likely to be a multiday survey.) Third, if available, a GIS land use map, that permits identification of the land use at each end of each trip recorded by the GPS device, will provide additional help in determining the trip purpose. With this information and the address data discussed previously, the purpose of most trips can be inferred, at least to the level of trip purposes used in most transport planning analysis. Typically, the purpose categorisation would be home-based work, home-based education, home-based shopping, home-based other, non-home-based work, and non-home-based other.

The main problems that will be encountered in this process, and that are largely only an issue if quite disaggregate purposes are desired are:

- Multiple use parcels and
- Shopping centres.
People may use these locations for multiple trip purposes, or may use them for one specific one that is not necessarily the one that would be picked most readily. For example, a person spending 90 minutes at a shopping centre might be there to eat a meal or to see a medical person. Normally, one would assume that this duration of time would be for shopping, but it may not be, in fact. Such activities at such locations will continue to present some difficulty for the identification of trip purpose.

5. An example of the application of the procedures

An example of the processing steps and their results is provided to show what is currently produced from our software. GPS devices are carried by respondents for a period of time, and the devices are then returned to us and the data stored are downloaded and processed. The initial data output, converted to a csv format, will appear generally like that shown in Figure 2.

```
V,07/03/2006,12:58:49,138.509622,-34.843843,500,78,3,8,3
V,07/03/2006,12:58:51,138.509650,-34.843809,500,65,2,3,8,3
V,07/03/2006,12:58:55,138.509714,-34.843786,500,66,3,8,8,3
V,07/03/2006,12:58:57,138.509732,-34.843779,500,68,2,8,3
A,07/03/2006,12:59:00,138.509741,-34.843580,500,26,8,4,11,6
A,07/03/2006,12:59:02,138.509778,-34.843454,500,34,3,4,11,6
A,07/03/2006,12:59:04,138.509805,-34.843373,500,53,3,4,11,6
A,07/03/2006,12:59:08,138.509833,-34.843252,500,28,4,11,5
A,07/03/2006,12:59:11,138.509879,-34.843183,500,30,2,4,11,5
A,07/03/2006,12:59:14,138.509943,-34.843149,500,38,2,4,11,4
A,07/03/2006,12:59:16,138.509970,-34.843126,500,56,4,4,11,4
A,07/03/2006,12:59:18,138.510016,-34.843098,499,63,4,4,11,4
A,07/03/2006,12:59:20,138.510007,-34.843115,499,72,3,4,11,4
A,07/03/2006,12:59:25,138.509970,-34.843149,498,82,2,4,11,3
A,07/03/2006,12:59:30,138.509961,-34.843172,498,84,2,4,11,3
A,07/03/2006,12:59:34,138.509970,-34.843188,497,89,2,4,11,2
A,07/03/2006,12:59:37,138.509961,-34.843199,496,93,2,4,11,2
A,07/03/2006,12:59:39,138.509970,-34.843202,496,75,3,4,11,1
A,07/03/2006,12:59:41,138.509961,-34.843206,495,73,3,4,11,1
A,07/03/2006,12:59:45,138.509925,-34.843222,494,76,2,4,11,1
A,07/03/2006,12:59:54,138.509933,-34.843337,492,95,2,4,11,0
A,07/03/2006,12:59:57,138.509924,-34.843373,491,107,2,4,11,0
A,07/03/2006,12:59:59,138.509933,-34.843390,491,110,2,4,10,9
A,07/03/2006,13:00:01,138.509870,-34.843390,490,109,2,4,10,9
```

*Figure 2: Data downloaded from a GPS device*

Information is stored with a file name that includes the deployment information. The data shown here provide the position, time, heading, speed, and so forth for each data point, where data are collected second by second. The entire processing procedure is shown in Figure 3.
Deducing mode and purpose from GPS data
Stopher, Clifford, Zhang & FitzGerald

Figure 3: GPS data processing procedure

Figure 4 shows an example of what the data might typically look like before editing, and on a background map that shows layers of the streets, bus stops, bus routes, etc.

Figure 4: Unprocessed GPS data on a base map

The actual traces of the person’s travel are shown by point layers for each of the four trips that were identified by the initial processing of the data. After editing, the point layers are converted to line layers, directional arrows are included in the lines and the trips are referenced correctly to the underlying network. This is shown in Figure 5.
At the next processing step, the mode of travel is identified for each of the trips. This is indicated in the key to the map, in this case, where trips 1a and 2a are identified as walk, trips 1 and 2 as bus, and trips 3 and 4 as car. Finally, the purpose of the trips is deduced, and a final map and summary table are produced, as shown in Figure 6.
6. Conclusions

Mode and purpose can be deduced with a quite high degree of accuracy from GPS traces, provided that certain supplementary data are available and also some are collected from respondents to the GPS survey. Appropriate GIS layers are essential for this process, and the supplementary data required from participants is bicycle ownership/availability and addresses of certain frequently-used locations. All of the remaining data required for mode and purpose identification are readily available from the GPS record.

It would be useful to assess the proportion of trips reported in conventional surveys where the mode and purpose are well defined. It is possible that the deduced mode and purpose, as described in this paper, achieves a similar level of accuracy to that achieved in self-report diaries. Further refinement of the procedures outlined here is underway currently, and improvements may be possible by introducing some form of Artificial Intelligence procedures to further refine the results obtained.

7. References


