CycleTracks – a Bicycle Route Choice Data Collection Application for GPS-Enabled Smart Phones

Billy Charlton[1], Elizabeth Sall[1], Michael Schwartz[1], Jeff Hood[1,3], Matt Paul[2]

[1] San Francisco County Transportation Authority, 100 Van Ness Avenue, 26th Floor, San Francisco, CA 94102

[2] Mopimp Productions, http://www.mopimp.com

[3] Department of City and Regional Planning, University of California, Berkeley

Keywords: Bicycle Route Choice, GPS, Data Collection, iPhone

Word Count: 2,311

Submitted for Consideration at the 2010 Innovations in Travel Modeling, a conference of the Transportation Research Board.

Original Submission Date: November 5th, 2009

Introduction

Non motorized modes such as walking and biking have played an increasing role in daily travel in major metropolitan areas, accounting for 9% and 4% respectively of work tours by San Francisco residents in 2000 (Bay Area Travel Survey 2000). Because of their relatively low costs, emissions, and high potential public health and welfare benefits, most public agencies, including the San Francisco County Transportation Authority (SFCTA) are actively seeking ways to promote even more non-motorized mode choice. Accordingly, attention to non-motorized modes within the regional travel demand models has increased. It is now standard practice in many urban areas to include walking and biking within mode choice models. However, due to lack of data, little is known about the routes and actual infrastructure that are used. There are two main benefits to adding a non-motorized mode route choice model to a regional travel demand model: (1) decision makers can geographically target non-motorized capital investments and operational improvements, and (2) travel demand models can quantify the accessibility benefits from non-motorized infrastructure within a log-sum context.

CHAMP is San Francisco's tour-based travel demand model, used citywide to quantify benefits and understand the implications of various transportation projects, plans, and policies (Outwater & Charlton 2006). Like many advanced travel models, CHAMP is currently able to forecast the quantity of bicycle and pedestrian trips, but lacks the capability to make non-motorized route choices to assign these trips to the network. To remedy this deficiency, Caltrans awarded SFCTA a State Planning and Research Grant to develop a bicycle route choice component of CHAMP. The goals of the research are to understand where cyclists ride in San Francisco and the Bay Area, as well as how network, personal, and trip-based factors affect those route choices (including slope, cycling infrastructure, age, gender, and trip purpose). Potential bicycle route choice methodology already exists (a detailed review of already existing methodology can be found in the companion paper (Hood et al. n.d.), but SFCTA's larger hurdle is obtaining sufficient revealed preference bicycle route choice data within a limited budget. This paper details SFCTA's strategy for collecting a sufficient amount of revealed preference bicycle route data by utilizing the satellite-based geographic positioning system (GPS) capability of consumer-grade smartphones.

Review of Route Choice Data Collection Techniques

Several data collection techniques from existing bicycle route choice studies were reviewed: (1) web-based stated preference surveys; (2) route recall; (3) personal GPS devices; (4) bicycle-mounted GPS devices; and (5) smart phone GPS.

Web-base stated preference surveys have the advantage of being quick and cheap. In addition, the stated preference nature allows the surveyor to force the respondent to make certain tradeoffs to lower the needed sample size, and base those tradeoffs on respondents' answers to previous questions (Sener, Eluru, and Bhat 2008). However, estimating a model from such forced tradeoffs is not nearly as reliable as from a good revealed choice data set where many biases can be eliminated.

Non-GPS methods can be costly on a per-record basis and are prone to human error in description and translation. McDonald and Burns (2001) used route-recall for respondent's "most recent commuting route." An additional problem with this methodology is that it is unlikely capture much stochasticity among individuals that occurs in the route selection process over the course of several days.

GPS devices are small and lightweight enough to carry on-person for long time periods. Price has also come down significantly in the past ten years and individual units can be obtained for less than $100. However, if the user is not continually prompted for mode information the GPS points for bicycle modes can be confused with various other modes. Meghini et al (2009) was successful in gleaning 2,657 bike routes from a personal GPS dataset with 11,000 total trips and 2,435 people. Dill and Gliebe used bicycle-mounted GPS devices to record bicycle routes of 164 adults in Portland, Oregon. The bike-mounted GPS eliminated much of the data-cleaning required by the personal GPS devices, but researchers still had to clean out portions of routes where the cyclist was on transit. The participants used the bike-mounted GPS device for seven days, at the end of which the project team retrieved the unit and downloaded the GPS data. Doherty (2009) provided smartphones to participants to track second-by-second GPS locations, which were regularly uploaded wirelessly to a remote server and stored in a MySQL database. The smartphones operated continuously for 17 hours at a time with the help of extra battery packs. Wireless data uploads negated the need for researchers to spend much time in the field retrieving data.

The techniques discussed above were contrasted with SFCTA's project needs and constraints. The SFCTA team had the following desired features for a data collection instrument:

- Revealed preference as opposed to stated-preference
- GPS-native data format to reduce error and researcher hours
- Cheap to deploy in terms of time (researcher hours) and money (capital investment)

These attributes pointed towards a smartphone-based GPS data collection that could wirelessly transmit data to an SFCTA server. Instead of providing devices to participants, the team could also leverage the high smartphone penetration in San Francisco and allow participants to download the application to their own phone. The iPhone was chosen as the initial deployment platform because it is easy to develop applications for, and has a high market penetration. The next section describes the CycleTracks iPhone Application, which SFCTA developed to collect bicycle route choice data.

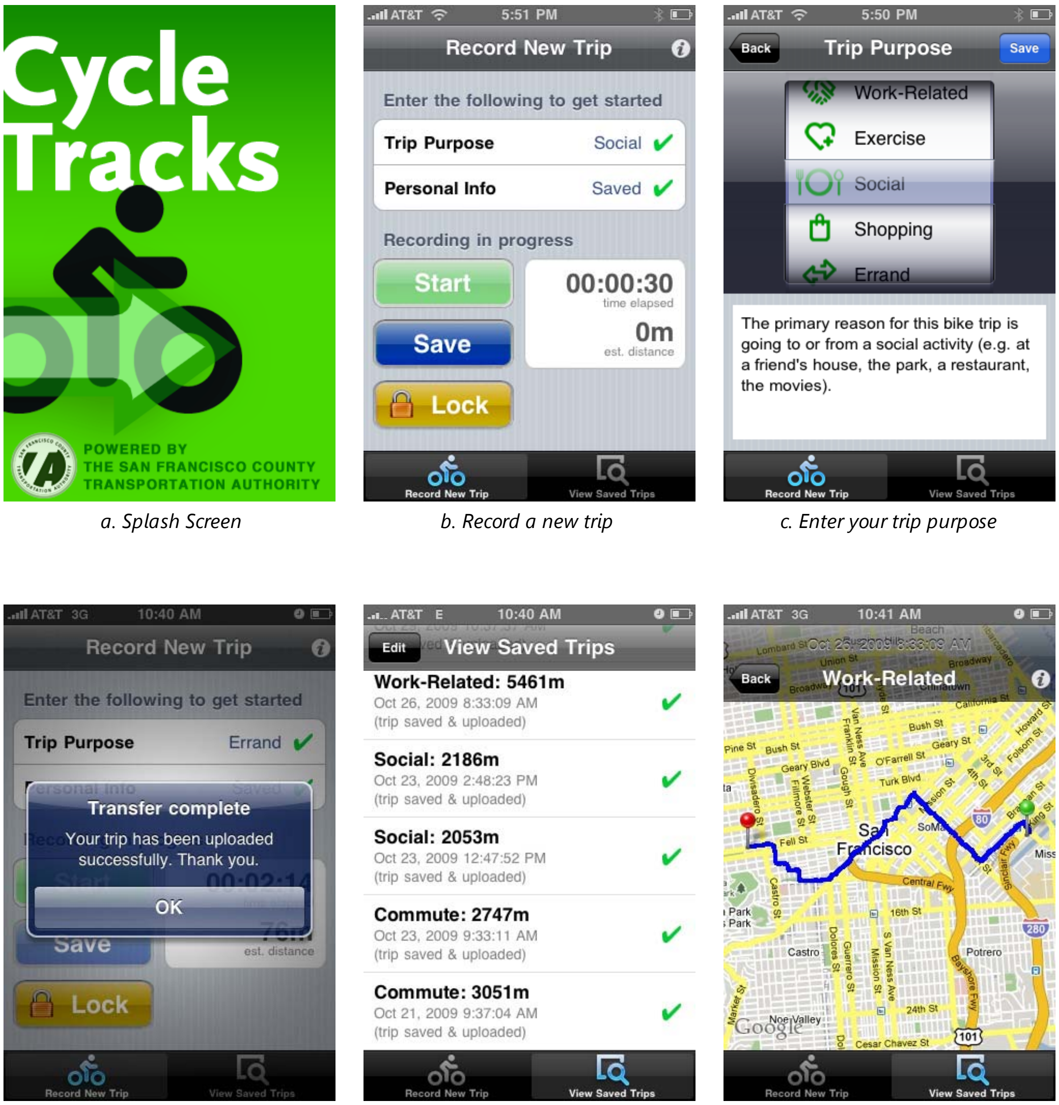
CycleTracks iPhone Application

Available on the iTunes "App Store" for free, any user can download and install CycleTracks. To make the app something cyclists would want to download, care was taken to make the app not just useful for our research purposes but also useful and possibly fun for cyclists. To that end, in addition to recording and sending GPS data to SFCTA servers, the app also allows users to view maps of all the routes they've recorded, and track their distance and speed on each trip.

The CycleTracks user experience is designed to be as unobtrusive as possible. To record a trip, cyclists select a trip purpose from a revolving wheel and tap **Start** (see Figure 1c). The iPhone then automatically locks itself to prevent the participant from accidently tapping something, dims the backlight to save battery life, and displays a timer to let the user know how long it has been since the trip started. When a participant finishes a bike trip, they slide-to-unlock the iPhone and tap **Upload** to upload the route to the SFCTA server.

GPS data is all saved locally throughout the trip, and only uploaded at completion. Participants have the option to change their trip purpose or add a note about the trip when they choose to upload. Since the GPS data is locally saved, participants can review the trips they have taken in a Google Maps mashup (Figure 1f) and delete trips they no longer wish to keep locally. Any trips that did not successfully upload can be re-uploaded and are marked with an exclamation point. The participant will be prompted to upload them.

Figure 1. Screenshots of the CycleTracks iPhone Application



a. Splash Screen *b. Record a new trip* *c. Enter your trip purpose*

d. Success! Trip uploaded. *e. List of saved trips for review* *f. Map of your bike trip*

A beta testing period revealed battery usage issues correlated with heavy GPS radio usage. Recognizing that participants are voluntary, several modifications avoid draining the iPhone battery. To remind participants that it is collecting data, CycleTracks makes a "bicycle bell" noise and vibrates every five minutes, after an initial 15 minutes of GPS data collection. Additionally, CycleTracks turns itself off if the iPhone battery life gets below 20%, allowing participants enough battery life to make phone calls.

Server Side Database

CycleTracks uploads trip data to a MySQL database on an SFCTA server, via an Apache webserver and the "Javascript over network" (JSON) data transfer protocol. Three data tables are stored: Person (identified by their iPhone unique device ID), Trips, and GPS Coordinates. Data stored in each of these tables is outlined below.

Note that the only user-identifiable fields in the collected data are the phone hardware "IMEI" number and a voluntarily-provided email address. The IMEI field needs to be collected so that multiple trips by the same person can be identified and linked to the anonymous user ID, and is scrubbed from the final analysis data.

Trip data typically takes just a few seconds to upload from the phone to the database. All route-based analysis can be performed on the raw GPS coordinate data at a later time.

Table 1: Person Table

User ID	Numeric identifier for the person record
Created	Creation date/time for this user record
Device ID	IMEI number of phone hardware (stripped from final data)
Home ZIP	Home ZIP code*
School ZIP	School ZIP code*
Work ZIP	Primary workplace ZIP code*
Gender	Male/Female*
Age	Age in years of this person*
Biking Frequency	Daily / Several per week / Several per month / <1 per month*
Email	Email address for raffle & future contact*

** Fields with an asterisk are optionally provided by user; response rate to be provided upon completion of study*

Table 2: Trip Table

Trip ID	To match to GPS Coordinates Table
User ID	To match to Person Table
Start Time	Time stamp for when user taps "start"
End Time	Time stamp for when user taps "stop"
Number of Coordinates	Number of non-null coordinates in this trip
Trip Purpose	Selected and confirmed by user (see Figure 1c)

Table 3: GPS Coordinates Table

Trip ID	To match to Trip Table
Time	Time stamp of when the GPS coordinate was taken. In addition to travel time and speed calculations, it is used to order points to determine route.
Latitude	Latitude and Longitude are stored with ten decimal points
Longitude	
Altitude	In meters
Estimated Accuracy	Provided by GPS system, in meters

CycleTracks Usage Promotion/Self Recruitment

From the very early stages of the project, SFCTA performed outreach to local bicycle advocacy organizations including the Bay Area Bicycle Coalition and the San Francisco Bicycle Coalition, along with sister agencies with an interest in the data. The campaign to generate interest includes bike coalition newsletters and blog entries, press releases to local media and transportation-related blogs.

To encourage use, these announcements included notice of multiple iTunes gift cards for users who submit at least one valid trip using the app. iTunes gift cards were a "no-brainer" incentive since fully 100% of users are known to already use the iTunes program.

Testing Results

In beta testing of the app, accuracy is usually sufficient in most Bay Area locations to identify the street being traversed. However, two cases repeatedly caused GPS problems:

- At the start of a trip, the GPS receiver may not have a fully accurate lock on enough satellites to provide high-quality coordinates. The iPhone attempts to place the location anyway, resulting in a noisy start to many trips. This problem is usually rectified after about one minute of use.
- A downtown "urban canyon effect" is noticeable in the high-rise portion of downtown San Francisco. The shadows of skyscrapers definitely block out some satellite signals, resulting in less accurate pinpointing. Fuller analysis of this effect is forthcoming.

Data Collection and Analysis

Developing the app took about three weeks longer than anticipated – which for a software development project isn't too unexpected. Unfortunately, as of the TRB paper submission deadline, CycleTracks is submitted to the iTunes App Store, but data collection will not begin until early November.

Analysis is expected to include several cleaning tasks: screening for non-Bay Area users (since there are no restrictions on who can download the app from the iTunes store); screening for false paths including non-bicycle trips, and eliminating "gamed" trips by the same person re-treading a particular route to skew results.

After cleanup, point-based paths will be snapped to the street network and analysis of the chosen routes can begin in earnest. We are particularly interested in identifying patterns of route choice that vary by time of day (or night) and weather, along with road characteristics such as number of vehicle lanes, speed of traffic, on-street parking, slope, and bicycling facilities.

Data cleaning issues, response rates, and methodology for conversion of point-based GPS to a "snapped" path which follows streets will be described in detail during presentation and in further revisions to this paper.

Conclusions and Next Steps

Depending on response rates, further application revisions and platforms may be added for GPS-equipped Google Android and BlackBerry phones.

The final database of revealed paths cyclists choose to ride will be used as the selected path (of several generated) in an newly estimated bicycle route-choice for the CHAMP model.

Obviously, it's difficult to augur the success of the study without the dataset in hand, so that will be added upon completion of the study by spring 2010. It is already clear from the beta testing and from excitement in the advocacy organizations that the basic premise is solid. We'll know soon if the gamble paid off.

Work Cited

Sener, I.N., N. Eluru and C.R. Bhat (2008) An analysis of bicycle route choice preferences using a web-based survey to examine bicycle facilities, Working Paper, Department of Civil Engineering, University of Texas, Austin.

Dill, J. and J. Gliebe (2008) Understanding and Measuring Bicycling Behavior: A Focus on Travel Time and Route Choice, Final Report. Oregon Transportation Research and Education Consortium Portland, OR.

Doherty, Sean T. (2009) Multi-sensor Monitoring of Human Activity and Physiology in the Built Environment. Presented at the Annual Meeting of the Transportation Research Board, Washington, DC.

Hood, J., Sall, E., & Charlton, B. (n.d.), A GPS-based Bicycle Route Choice Model for San Francisco,California. Submitted for presentation at the 3rd Conference on Innovations in Travel Modeling 2010, a Transportation Research Board Conference.

McDonald, C.H. and E.K. Burns (2001) Cycling to Work in Phoenix: Route Choice, Travel Behavior, and Commuter Characteristics (Paper 01-2526). Presented at the Annual Meeting of the Transportation Research Board,Washington, DC.

Menghini, G., N. Carrasco, N. Schüssler and K.W. Axhausen (2009) Route choice of cyclists: discrete choice modelling based on GPS-data, Arbeitsberichte Verkehrs- und Raumplanung, 544, IVT, ETH Zurich, Zurich.

Outwater, M. & Charlton, B. (2006), The San Francisco model in practice: Validation, testing, and application, in `Innovations in Travel Demand Modeling: Summary of a Conference', Vol. 2, Transportation Research Board, Washington, DC.