Inferring Usage Characteristics of Electric Bicycles from Position Information

Johannes Paefgen  
University of St. Gallen  
johannes.paefgen@unisg.ch

Florian Michahelles  
ETH Zurich  
fmichahelles@ethz.ch

ABSTRACT
This paper describes an experimental setup for the analysis of e-bike usage characteristics based on GPS data. Usage characteristics include parameters such as average and maximum velocity, trip lengths and distribution over daytime. Based on high resolution position measurement these parameters are extracted and compared to other studies on both e-bikes and conventional bicycles. We show that applying location technology to concurrent monitoring of a fleet of e-bikes yields higher quality in terms of resolution and accuracy (1), and is less intrusive (2) than obtaining these data by conventional user surveys. The findings form a proof-of-concept for the adoption of location technology to transportation and behavioral sciences and suggest further interdisciplinary collaboration in these fields.

Categories and Subject Descriptors
J.4 [Computer Applications]: Social and Behavioral Sciences – economics, psychology.

General Terms
Economics, Experimentation, Human Factors.

Keywords
Electric bicycle, e-Bike, GPS data, e-mobility.

1. INTRODUCTION
While four-wheel e-mobility continues to dominate academic and press publications on sustainable mobility, electric bicycles -- also known as e-bikes -- have become a successful trend in recent years. E-bikes are two-wheel vehicles of the dimensions of conventional bicycles that are either fully powered by an electric motor, or supply additional torque to human pedal power in a design variant called the Pedelec. Reasons for their increasing adoption in Europe, Asia and North America are low energy costs when compared to alternative modes of transportation, low or no costs of insurance, registration and parking, the avoidance of traffic jams and their perception by users as environmentally friendly [1]. Electric bicycles have been tremendously successful in China, where more than 20 million units were sold in 2009. For large parts of the population in China, four-wheel vehicles -- conventional fuel or electrical -- are not yet affordable for a majority of the population, so that electric bicycles form a viable option in personal transportation [2]. In Western countries, the situation is somewhat different. Here, electric bicycles do not substitute cars, but are perceived as an additional mode of transportation for close-distance commuting or leisure activities, being more convenient than non-electric bicycles but more environmentally friendly than scooters or subcompact-class cars. Furthermore, they enable elderly or handicapped people to operate bicycles also in hilly landscapes or with higher payloads, e.g., groceries [3]. In Switzerland, they are currently far more popular than electric four-wheelers with more than 30'000 electric bicycles on the road in mid-2010 according to statistics published by the Swiss federal office of energy.

While electric bicycle technology is well-researched and understood, user behavior and accident risk factors are yet relatively unknown. While it is statistically established that accidents involving electric bicycles are both more frequent and more severe, it is not yet clear to which extend different causes contribute to this trend. This lack of understanding of risk factors in e-Bike usage has also been acknowledged by a recently published report by the Swiss council for accident prevention. The authors specifically call for increased research effort that allows for an evidence-based evaluation of usage characteristics [4]. In China, growing concerns with increased accident risks due to higher velocities have even resulted in e-bike bans in several municipalities [5]. Our paper investigates how usage characteristics of e-bikes can be analyzed using location technology. This demands a spatial data-mining approach, where GPS data collected from individual e-bikes has to be aggregated, filtered and interpreted in order to arrive at higher-order data structures that allow for the inference of usage characteristics. Examples of such characteristics are average and maximum velocities, type of road, distribution over daytime and trip routes. Most of these parameters can be expected to be the same as for conventional bicycles so that the research objective becomes their adaption to our specific experimental setup.

The remainder of this paper is organized as follows. First we briefly review related work on usage characteristics of both e-bikes and conventional bicycles, and on the role of ubiquitous data in the bicycling context. The subsequent section describes the experimental setup and data processing methodology. After some preliminary empirical results we conclude with some remarks on limitations of the presented work and outline general insights gained in our research.
2. RELATED WORK

Much of the research on e-bike usage focuses on China as the largest market worldwide. Within this local scope, researchers have analyzed technical, economic and political factors that facilitated the transition from conventional bicycles to e-bikes [2]. One study specifically investigates usage characteristics and travel patterns by questionnaires. Users of e-bikes are found to travel distinctively longer distances than users of conventional bicycles, and at higher velocities. Build-in speed limits imposed by regulators are seen as the main control lever for risk factors of e-bikes [6]. Another study examining future potential of e-bikes in China arrives at the conclusion that due to increasing battery range and decreasing market price, e-bikes usage will spread from large cities to small and medium size cities and rural areas as an attractive alternative to unreliable public transportation systems [7]. None of the studies mentioned above discusses usage characteristics for e-bikes based on hard experimental data.

The analysis of ubiquitous, in the sense of automatically and continuously generated, data is a prevalent subject in transportation research. In the bicycling context, one remarkable paper investigates shared bicycle usage characteristics in the city of Barcelona based on RFID sensor readings. Bicycles are provided by distributed network of stations, where they can be checked-out by users at a station kiosk using a RFID chip card. From the sensor readings, several temporal and spatiotemporal patterns are extracted. These visualize transportation dynamics in the city and allow for a prediction of bicycle station usage [8]. Another recent paper presents algorithms for the inference of transportation modes, including bicycle use, from raw GPS data based on supervised learning. However, the paper has a strong methodological focus especially on the classification of transportation modes and does not discuss usage characteristics from an individual-centered behavioral perspective [9].

For conventional bicycles, the body of research on individual usage patterns is comparatively broad. We consider two recent publications as particularly relevant to our research. One study describes a large-scale survey amongst bicyclists in Texas which is evaluated through the application of several econometric models. Research questions include perception of safety and quality by bicyclists, frequency of bicycling for commute and non-commute purposes and the importance of bicycling infrastructure [10]. Another study in fact employs GPS monitoring of 164 bicyclists in the area of Portland, OR and focuses on the bicycling routes taken by participants, how these routes depend on infrastructure and how they differentiate from a calculated optimal path. In terms of data quality, this study appears to be superior to comparable research that does not rely on position information [11]. To our knowledge, no similar studies have been carried out with a focus on e-bikes, especially for Switzerland where geographic, cultural and regulatory boundary conditions can be expected to be quite different from China or the US.

We would like to remark that the research outlined below can also be understood as a bounded socio-technical experiment for the sustainability transition of personal mobility in the sense of [12]. Higher-order learning takes place amongst both the participants in the experiment due to their personal experiences, and society as a whole due to the generation of public awareness of e-bikes.

3. RESEARCH METHODOLOGY

3.1 Experimental Setup

The e-bikes used in the experiment are manufactured in Switzerland by the Austrian company Thoenus and distributed under the model name Stromer. They are equipped with a 250 W power system resulting in 20 Nm torque output by the motor. Power supply is a Li-Ion battery with 11 Ah capacity and 36 V nominal voltage. The motor is designed to support the bicyclist (Pedelec-design) and can recuperate brake energy, e.g. during downhill rides. Every Stromer bicycle was equipped with a Picotrack GPS/GSM device by German company Telic. These devices are equipped with a GPS and a GSM module and integrated antennas, an accelerometer, and an internal Li-Ion battery for independent power supply. With 54x37x20 mm its dimensions make it suitable for mounting on a bicycle frame. As the internal battery allows only for continuous operation of approx. 24h, the Picotrack device was connected to the power supply of the Stromer bicycle. Data is transmitted to a server’s IP address using the GPRS communication standard. The GPRS capability of the Picotrack device also provides A-GPS functionality, which allows for increased accuracy and a shorter time to first fix (TTFF) after initialization.

A total of 17 e-bikes were fitted with a Picotrack device and distributed amongst branches of an insurance company in 11 different cities in Switzerland. In all these branches, employees were encouraged to make use of the available e-bike free of charge, without limitation of time or distance and for work purposes as well as for leisure activities. Though this does not constitute a statistically valid sample, it is sufficient for a thorough testing of hardware and the development of appropriate evaluation algorithms.

3.2 Data Collection and Processing

The Picotrack device transmits the most recent position measurement together with a timestamp, a smoothed velocity value, and some additional status information. Transmission can be triggered by several predefined events. In our setup, we distinguish four types of events: Movement after a period of inactivity detected by the accelerometer, change of direction in terms of a course deviation of more than 30 degree with a velocity larger 6 km/h, driven distance of more than 500 m since last position, and 300 s passed since last transmission. These flexible criteria guarantee for a sufficient position resolution to identify streets and intersections, while avoiding the generation of vast amounts of data that would not add additional informational value. Raw positions are stored in a database from which they can be exported in standard comma separated value format.

Data processing and analysis is carried out in Matlab. All records of an individual e-bike for a given period are consolidated into a matrix, where timestamps are converted to Unix time format. In order to calculate the distance D between two data points we use the approximation:

\[ D = \sqrt{ \left[ l \left( \lambda_2 - \lambda_1 \right) \right]^2 + \left[ l \cos \left( \frac{\lambda_1 + \lambda_2}{2} \right) (\theta_2 - \theta_1) \right]^2}, \]
where $\lambda$ denotes latitude, $\theta$ longitude, and $l$ is the length of a circle segment corresponding to one latitudinal degree on the equator, i.e., about 111'120 m. The cosine correction term adapts longitudinal lengths to a given latitude, consequently reaching 1 at the equator at zero degrees latitude.

Total distance for a single trip is computed by accumulating distances between individual data points. Trip beginning and end points can be identified by the uninterrupted succession of valid event triggers. Furthermore, if a trip exceeds a certain velocity threshold, it is marked as invalid as this points to the transportation of an e-bike via automobile or train. For each trip, we store total distance, duration, average velocity and maximum velocity. Furthermore, for every e-bike elapsed driving time and driven distance are accumulated according to the hour of the day.

4. PRELIMINARY RESULTS

Below, we discuss the evaluation of a data set collected over a duration of four months. Only two selected e-bike records are discussed as a proof-of-concept, a similar evaluation for all e-bikes of a statistically valid sample will be addressed in later work.

One of these records comes from the city of Basel, with 7043 position recordings and a total driven distance of 186 km. The other sample comes from the Zurich area with 2786 position recordings and a total driven distance of 79.5 km. These two were selected due to the fact that Zurich and Basel are the largest Swiss cities in the experiment and e-bike usage in the considered period was particularly high.

4.1 Evaluation of Trip Characteristics

For an evaluation of trip characteristics, we consider the average duration of a trip and the average driven distance. These values are given for the two selected e-bikes in table 1, together with results from the e-bike study in two Chinese cities mentioned above [6]. Furthermore, a comparative figure for bicycling is given from [11]. Comparison of distances does not yield substantial insights regarding e-bike use in our setup, but results lie within the range of expected values. The increasing trip lengths may be attributed to both geographical features of the cities and differences in trip purposes. Zurich and Basel are relatively small cities where distances between destinations can be assumed to be shorter than in Shanghai or Kunming. The high average trip length for the Portland dataset follows a trip purpose rationale, due to the fact that bicycling trips are often done as physical exercise and for recreational purposes.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Average trip duration</th>
<th>Average trip distance</th>
<th>Average velocity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zurich</td>
<td>4.92 min</td>
<td>1556 m</td>
<td>18.98 km/h</td>
</tr>
<tr>
<td>Basel</td>
<td>7.06 min</td>
<td>3504 m</td>
<td>29.78 km/h</td>
</tr>
<tr>
<td>Shanghai</td>
<td>18.75 min</td>
<td>4250 m</td>
<td>13.60 km/h</td>
</tr>
<tr>
<td>Kunming</td>
<td>11.29 min</td>
<td>2080 m</td>
<td>11.05 km/h</td>
</tr>
<tr>
<td>Portland</td>
<td>23.89 min</td>
<td>6920 m</td>
<td>17.38 km/h</td>
</tr>
</tbody>
</table>

From the average velocities, it becomes evident that e-bikes in our experiment where driven faster than e-bikes in the Chinese study and then conventional bicycles in the study from Portland. A possible explanation is that e-bike velocities in China are limited to 20 km/h by regulation. The relatively high average velocities for conventional bicycles again points to physical exercise as a trip purpose. When used for commuting, e-bikes can be slowed down by surrounding traffic. The exceptionally high average velocity for Basel seems unreasonable and will be subject to further analysis.

Maximum velocities are not given in the two cited sources, but in our experiment were 56 km/h and 46 km/h for Zurich and Basel, respectively. Because these values are hardly reached by leisure bicyclists on conventional bicycles, they already hint towards higher accident risks for e-bikes.

4.2 Distribution over Daytime

In the following, we discuss the distribution of e-bike usage over different times of the day. For this purpose, we aggregate e-bike usage time during every hour of the day and normalize the results to 1. The corresponding plots are displayed in figure 1 & 2 for Zurich and Basel, respectively.
Several insights can be gained from these histograms. Firstly, a large fraction of usage time can be attributed to commuting. For Zürich, the 9th, 18th and 19th hour of the day constitute approximately 60% of total usage time. For Basel, this trend is less distinct, but still clearly visible. Secondly, when comparing distribution shapes, it becomes evident that the distribution for Basel has a large fraction of usage time allocated to afternoon hours. Apparently this e-bike was used for leisure activities to a greater extent than the Zürich e-bike. The e-bike study from China does not provide any data on the distribution of e-bike usage over daytime. The study from Portland, on the other hand, provides the distribution of trips over trip start times by hour, for conventional bicycles (cf. [11] Figure 3.3, page 36). Here, afternoon trips account for more than one quarter of all trips, which is consistent with the assumption above that conventional bicycles are used for leisure activities instead of commuting more often than e-bikes.

5. CONCLUSION
In this paper we have described an experimental setup for the inference of e-bike usage characteristics from position data and delivered a proof-of-concept analysis of first empirical results. Our research demonstrates a use-case how transportation and behavioral sciences can employ location technology in their respective fields. E-bikes are especially attractive research subjects in this sense, as their battery provides sufficient energy for an uninterrupted operation of location technology. The technical infrastructure used in our setup is reliable and can be installed with acceptable efforts. The concurrent monitoring of a fleet of e-bikes is preferable to data gained from user surveys in two ways. Firstly, it is of higher quality in terms of accuracy and accuracy as it does not depend on the assertions of individuals. Secondly its collection affects user behavior to a lesser extent, as the Picotrack device is not recognized during everyday use. The parameters extracted from the data in our setup are consistent with existing literature in the field and allow for a comparison of e-bike usage characteristics with those of conventional bicycles.

Based on these findings, we are preparing a study for the investigation of the effects of monetary incentives and normative instruction on usage characteristics of e-bikes. Therefore, a set of test subjects will be provided with e-bikes for a duration of several months. Based on concepts from behavioral theory, we will than compare usage characteristics from a reference group with two other groups that are exposed to the mentioned stimuli.

We acknowledge several shortcomings in our research presented so far. The most prominent limitation is the sample considered in the analysis. While 17 e-bikes may be sufficient to determine fundamental usage characteristics, the experimental setup should focus on one city at a time and data collection should cover a longer period, even several seasons in the optimal case. Furthermore, only a very limited number of usage parameters, especially for individual trips, where considered in our analysis. As a next step, more information should be extracted from the dataset in the trip aggregation phase, such as start and end times. Based on this additional information, further classifications of trips become feasible, e.g., using cluster analysis. In addition to refined information extraction from our sample, external information such as weather conditions and route characteristics should be taken into account as well. Lastly, it appears reasonable to extend data sources by an additional questionnaire to be filled out by participants. This would add valuable information in terms of user demographics and bicycling behavior to future analyses.

As a final thought we would like to express our opinion that social sciences in general, and transportation and behavioral sciences specifically in the case of e-bikes can achieve substantial improvements in data quality through the use of location technology. Where applicable, the tracking of locations poses a valuable alternative to conventional questionnaire surveys when movement patterns have to be measured and deducted. This insight calls for increased collaboration between the location technology community and researchers from social sciences.

6. REFERENCES