# A Study of Comfort Measuring System Using Taxi Trajectories

Li-Ping Tung[1], Tsung-Hsun Chien[2,3], Ting-An Wang[3], Cheng-Yu Lin[3]
Shyh-Kang Jeng[2], and Ling-Jyh Chen[3]
[1]Intelligent Information and Communications Research Center, National Chiao Tung University
lptung@nctu.edu.tw
[2]Graduate Institute of Communication Engineering, National Taiwan University
d99942014@ntu.edu.tw, skjeng@cc.ee.ntu.edu.tw
[3]Institute of Information Science, Academia Sinica
{balamark, ntuaha, cclljj}@iis.sinica.edu.tw

***Abstract*—The comfort of rides has been identified as one of the top criteria that affects passengers' satisfaction with public transportation systems. Conventional comfort measurement approaches rely on professional measure tools or interviews from passengers, which are costly, time-consuming, and not yet feasible. The concept of Internet of Things (IoT) is a new solution to answer this problem. The idea of IoT is to interconnect state-of-the-art digital products in physical world to provide more powerful applications. Vehicles equipped with GPS devices and wireless access technologies are parts of the IoT elements in traffic networks. We use the GPS data to measure the comfort level of vehicle rides, and provide a detailed comfort statistics as a value added service. Using real data collected from one of the Taipei taxi service providers, we show that over $95\%$ taxi trajectories are viewed as comfortable. In addition, rides without passengers get higher comfort scores than with passengers. We also give the rankings of all taxi drivers according to a number of criteria, such as the comfort score and the number of loads. With the ranking results, we can track back to the trajectories and infer drives' driving behaviors, road conditions, and traffic conditions. We believe that the proposed solution has the potential to provide a representative comfort measurement service for taxi services and additional value-added services for public transportation systems.**

## I. Introduction

The rapid development of the Internet, evolving from a network of a few hundred hosts to a platform connecting billions of "things", makes the Internet of Things (IoT) become a novel design paradigm. The fundamental concept of IoT is to interconnect diverse objects in the physical world, such as vehicles and cellphones, through low-cost information gathering devices, which are like RFID tags, sensors, actuators, machine-to-machine devices, that facilitate interactions among the objects themselves as well as the objects and persons in any place and at any time.

IoT enables a wide range of intelligent applications and services that can help people in their daily life; for example, intelligent transport systems will help reduce traffic congestion; remote healthcare systems will help doctors have a review and analyze the patients' condition; smart grid systems will help respond to many conditions in power supply and demand, and so on. These applications would dramatically change the way of our life.

Nowadays, vehicles, such as taxis and buses, equipped with GPS devices have been applied to collect the instant traffic condition data of some specific areas. They can be viewed as many pervasive and mobile sensors to form a large scale dynamic sensor network for the purpose of gathering traffic information. The collected data could be real-time transmitted to the data server via wireless technologies, such as WiMAX and 3G. Thus, vehicles are viewed as parts of Internet of Things. The collected data generally include longitude, latitude, speed, and etc.. Based on these trajectory data, many useful applications are developed during the past years [5, 11, 12]. For example, Balan *et al.* [5] used the historical taxi data to provide passengers with the expected trip time and fare of a given itinerary. Yuan *et al.* [11] used the taxi trajectory data to predict driving directions. And Yue *et al.* [12] recorded passengers' desirable areas and trajectory patterns for the purpose of supervising urban traffic or serving location-based services (LBS).

In this study, we focus on not only the analysis of taxi drivers' driving behaviors but also the comfort measurement of the driving. The *comfort* of rides has been identified as one of the top criteria that affect customers' satisfaction with public transportation systems, and it has been shown that *comfort* is an important consideration for passengers that use public transportation[4, 7, 8]. There are a lot of factors that would affect the comfort level of rides, such as road conditions, traffic jam or not, drivers' driving behaviors, vehicles' age, and so forth. However, conventional comfort measuring approaches rely on personal interview [9] or literature surveys [6], which are generally labor-intensive and time-consuming, and are thus limited in terms of scalability and timeliness.

Comfort Measuring System (CMS) proposed by Lin *et al.* [10] exploits the GPS and 3-axis accelerometer functions of modern smart phones to measure the comfort levels of rides on public transportation systems. Using personal mobile devices to collect data via down-top gathering model is called *Participatory Sensing*. Everyone who owns a smart phone can contribute the instant collected information and join the IoT network. In this paper, we use the GPS data provided by one of the Taipei taxi service providers to do comfort measurements. The GPS data are regarded as the elements of IoT, and are transformed into our defined comfort scores. With our analysis, the comfort measurement results could be useful to understand drivers' basic driving behaviors, traffic conditions, and human beings' daily behaviors for further research, like social sciences.

The remainder of this paper is organized as follows. In Section II, we present the CMS system; and in Section III, we describe the time features of taxi services. In Section IV, we provide a preliminary set of comfort measurement results, and we investigate the factors that affect the comfort levels of the taxi services in detail. We then summarize our conclusions in Section V.

## II. The Comfort Measuring System

Specifically, a trajectory is the path of a moving object (i.e., a vehicle) through space. It is usually represented by a set of discrete sample points on the path with a fixed time interval between every two contiguous data points. Each data point contains a time-stamp of the sample and its geographical location information (i.e., the latitude and the longitude). In addition, the vibration measures contain a sequence of accelerations. We let $v_{t_1}$ and $v_{t_2}$ denote the driving speed of the taxi recorded at time $t_1$ and $t_2$ respectively. The acceleration of the taxi can be obtained by $a_f = \frac{v_{t_2} - v_{t_1}}{t_2 - t_1}$; and following the ISO 2631 standard [2], we obtain the value of $a_w = \sqrt{1.4a_f^2}$. Then, the weighted root-mean-square acceleration is obtained by Eq. 1 in meters per second squared ($m/s^2$) for translational vibration, where $T$ is denoted as the duration of the measurement, in second.

$$a_t = \frac{1}{T}[\int_0^T a_w^2(t)dt]^2 \quad (1)$$

With the value of $a_t$,we adopt the ISO 8041 standard to calculate the *acceleration level* at time t, i.e., $L_t = 20 \log \frac{a_t}{a_0}$, where $a_0$ is a normalization factor with a constant value equal to $10^{-5} m/s^2$ [3]. Then, following [1], we obtain the comfort index at time t, i.e., $CI_t$, by Eq. 2.

$$CI_t = \begin{cases} 1, & \text{if } L_t \leq 83dB \\ 2, & \text{if } 83dB < L_t \leq 88dB \\ 3, & \text{if } 88dB < L_t \leq 93dB \\ 4, & \text{if } 93dB < L_t \leq 98dB \\ 5, & \text{if } 98dB < L_t \leq 103dB \\ 6, & \text{if } 103dB < L_t \end{cases} \quad (2)$$

For simplicity, we define the *Comfort Score* as $20 \times (6 - CI_t)$, which indicates that the larger the comfort score, the more comfortable will be the transportation experience. We calculate the comfort score of a trajectory by averaging all the comfort sores that belong to that trajectory.

Table I
The Entries in the Dataset

| Column | Datatype | Description |
|---|---|---|
| id | int | Sequence Number |
| micmac | char | Taxi Number |
| longitude | double | Longitude of Trajectory |
| latitude | double | Latitude of Trajectory |
| speed | double | Driving Speed |
| datatime | datatime | Driving Time |
| clientontaxi | bool | Load/Unload Passenger |

## III. Taxi Trajectory Dataset

This study is based on taxi service. This dataset was collected by one of the Taipei taxi service providers for operational purpose, such as taxi dispatch, real-time vehicle tracking, and many others. Each taxi is equipped with one recorded device, which logs GPS data and helps taxi drivers keep records of loading and unloading, and then real-time transmits back to taxi service provider by network. Taxi service provider received all the data from taxis and then for subsequent analysis. Table. I shows the part of fields of the dataset including the ID of the taxi, time, longitude, latitude, velocity, and load. The load bit is used to indicate the status of the taxi. When a passenger takes a taxi, the load bit is set to one. Otherwise, the load bit is set to zero to indicate the unloading. The load bit is also used to partition the trajectories. All data are logged every second. The dataset here summaries 21 days of activity from November 8 to November 28 in 2010, which consists of near 200,000 trajectories among about 700 taxis. If the trip time is less than four minutes, we thought this trajectory is not reasonable and then deleted these error data.

First, we shows lots of statistical results of our dataset. The average number of taxis among 24 hours is shown in Figure 1. During the time interval from midnight to dawn, there are very scarce taxi drivers working at this period, which is a normal phenomenon. After then, we observe that the number of taxis is increasing till 10:00 in the morning. Taxi drivers know the large demand of services in this working period. Moreover, in the afternoon, 15:00 is another peak and then gradually decreases. We can infer that taxi drivers' working types also follow the daily schedule.

Figure 2 shows the total number of loads per hour. The curve of the number of loads is close to that shown in the Figure 1. The number of loads from 2:00 to 6:00 is small, but there are higher loads at 1:00. Perhaps passengers might finish their night activities and go back to home. In addition, we observe a first local maximum about 10:00. This maximum corresponds to the typical starting hours in offices, which varies normally between 8:00 to 10:00 in Taipei city. A second maximum is observed around 18:00, which corresponds to the typical finishing time of many working schedules (typically between 17:00 and 18:00).

Next, we show the average number of taxis and the total

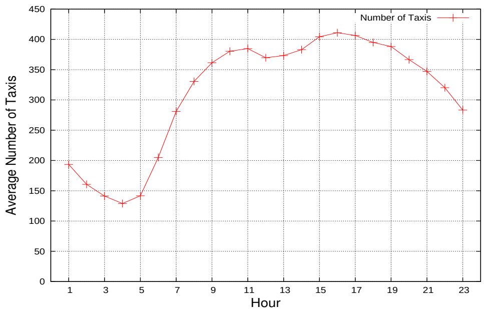


Figure 1. The average number of taxis among 24 hours

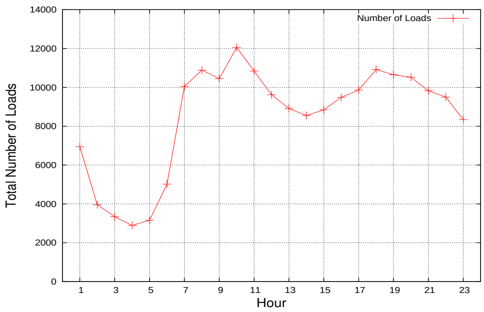


Figure 2. The total number of loads among 24 hours

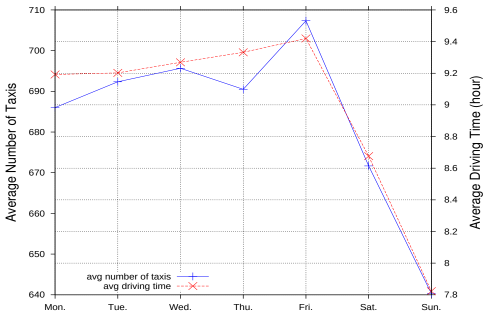


Figure 3. The average number of taxis and the average driving time among a week

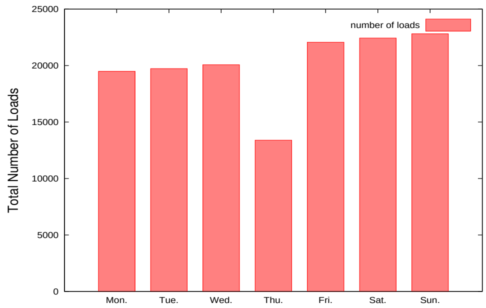


Figure 4. The total number of loads among a week

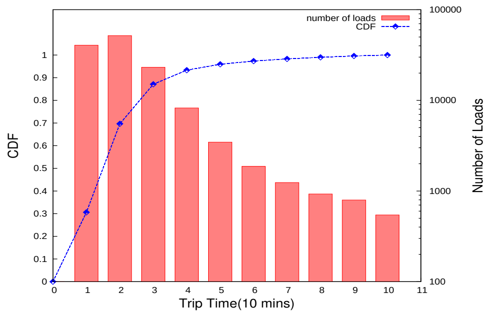


Figure 5. The distribution of trip time

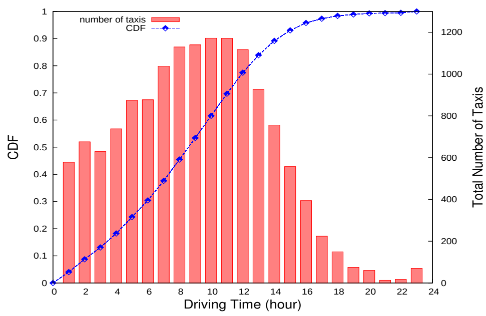


Figure 6. The distribution of driving time

number of loads among a week in Figure 3 and Figure 4 respectively. Figure 3 shows the property that it has most drivers in work on Friday. Besides, the average driving time is at least nine hours in the weekday, which is higher than in weekend. Perhaps taxi drivers might also need to take a rest on weekend. In Figure 4, it shows that loads gradually increases from Friday, and the maximum loads happens on Sunday. Friday is an end day of the working day; therefore, there are more demands for public transportation systems both from local travelers or visitors from other places than in weekday. Thus, it is obvious that the higher loads in Friday. Furthermore, there are around total 20,000 loads for all taxis except Thursday, and the total number of loads is even more than 20,000 from Friday to Sunday.

The distribution of trip time is illustrated in Figure 5. It shows an interesting result that more than $85\%$ is under $30$ minutes in a trip. Passengers' destinations almost can be reached in $30$ minutes. It means that passengers choose to take taxis for short trips in urban area for the purpose of saving time. Otherwise, a long distance trip follows a high fare. Passengers would rather choose other public transportation systems. In addition, taxi drivers are also not willing to leave too far for the sake of reducing risk of no load in returning trip. Besides trip time, the driving time distribution is shown in Figure 6. We observe that most of drivers work seven to twelve hours per day, which corresponds to the typical working time.

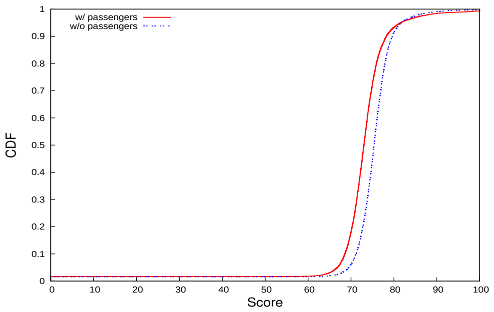


Figure 7. Comfort scores in CDF distribution with and without passengers

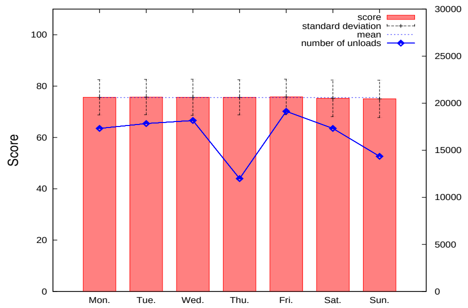


Figure 8. Comfort score performance among a week without passengers

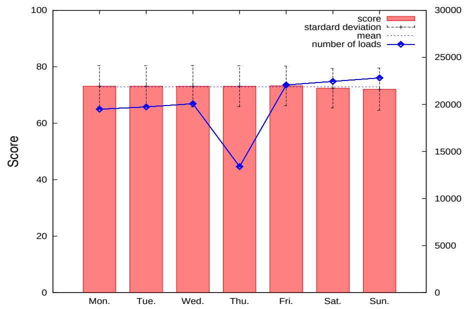


Figure 9. Comfort score performance among a week with passengers

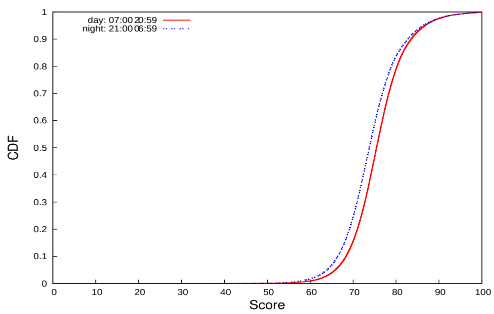


Figure 10. Comfort score during day and night without passengers

## IV. Analysis of Comfort Measurement

### A. Comfort Score

Statistics of comfort scores are summarized in this section. These results could further help us get insights into the drivers' behaviors. Figure 7 shows comfort scores with and without passengers in cumulative distribution function (CDF) curves. The distribution of comfort scores range around seventy to eighty. We can see that taxi drivers obtain higher scores if they have no passengers on taxis, compared to the conditions with passengers. When there is no passengers in the taxi, the drivers' goal is to search passengers on the road. Thus, they must drive slowly and smoothly to notice whether someone calls a taxi along the roadside. Driving slowly and smoothly makes the acceleration of driving more stable, thus it can obtain higher comfort scores. On the contrary, the taxi drivers hope to reach the destination as soon as possible if they have passengers in the taxis. Moreover, traffic conditions are more complex in the urban area. Taxi drivers may accelerate the car to pass through an amber light for saving time. But if there is no enough time to pass through, they must do emergency stop. Changing driving speed suddenly introduces the high variation on accelerations, which incurs low comfort score. According to the ISO standard and our defined comfort scores, scores above $60$ are regarded as comfortable. We can observe that over $95\%$ trajectories are considered as comfort.

Figure 8 and Figure 9 illustrate the comfort score performance without and with passengers respectively. The mean comfort score is $82$ under the condition of no passengers while the score is $74$ under the loaded case, which is compliant with the previous analysis. In addition, the comfort score among a week follows a uniform distribution regardless of load or unload. It also shows that the total number of loads is over 20,000 from Friday to Sunday, which is much higher than that from Monday to Thursday in Figure 9. It conforms to the high demand of taking a taxi in the weekend.

Next, we shows the comfort scores analysis during day and night. We define time segment from 7:00 to 20:59 as day, and that from 21:00 to 6:59 as night. Figure 10 and Figure 11 show the results without and with passengers respectively. We observed that it gets better scores when driving in the day than in the night no matter whether there are passengers in the taxi or not. It is because that taxi drivers are tired and can not drive a car well in the night. Furthermore, comfort scores without passengers would higher than that with passengers, which also corresponds to the former analysis.

Figure 12 shows the comfort score performance correlated with the trip time. We can easily see that the longer trip time, the higher comfort score. Perhaps the longer trip time for a passenger is a trip from urban to suburban or on the highway. Taxi drivers could drive more steadily without bad traffic conditions in these areas. Thus, it achieves higher comfort score. Besides, Figure 13 shows the comfort score correlated with the trip distances, which has similar results to Figure 12. As we know, the trip time is proportional to the trip distance if there is no heavy traffic conditions. Thus, the

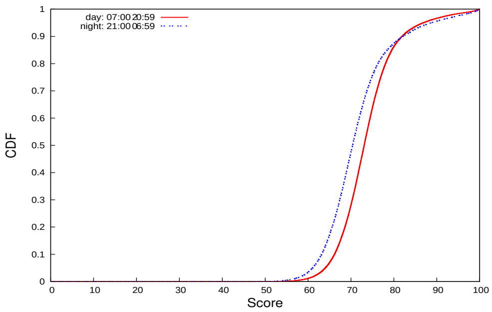


Figure 11. Comfort score during day and night with passengers

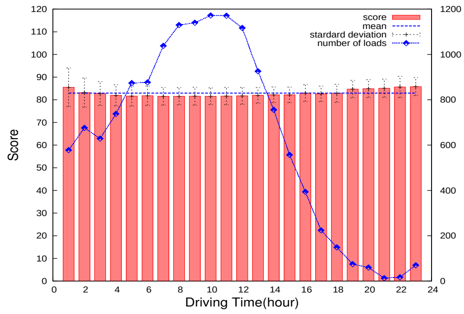


Figure 14. Comfort score analysis with the driving time

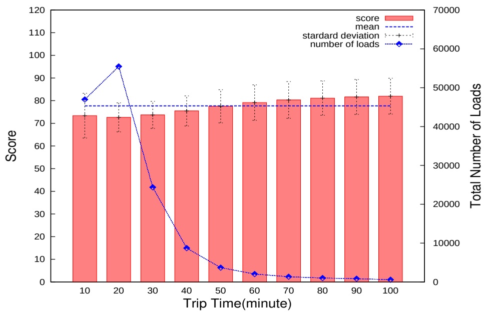


Figure 12. Comfort score analysis with the trip time

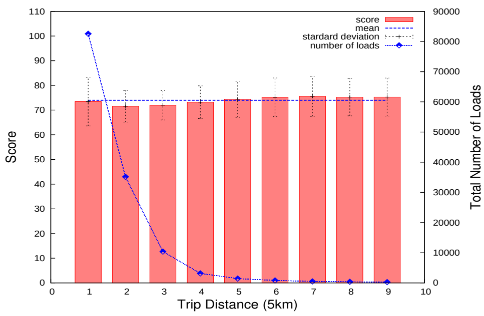


Figure 13. Comfort score analysis with the trip distance

longer trip distance has the better comfort score. We can also observe that most of the trip distances are below five kilometers, and it correlates to our dataset which are collected in Taipei city.

Finally, in Figure 14, we can see that taxi drivers who drive eight to twelve hours per day have the most number of loads, but the comfort scores for these drivers are under mean score.

### B. Ranking

Based on the collected data, we can give the ranking lists of all taxi drivers according to some criteria, such as the number of loads, the comfort scores, and so on. With these ranking lists, taxi drivers' behaviors can be analysed. We can track back to the trajectories to understand what happened, such as drives' driving behaviors, road conditions, traffic conditions, and etc.. For example, a driver with the higher number of loads may have excellent driving habit. However, drivers with bad comfort scores may not represent that they are bad drivers. If we further add on the spatial analysis, we may correlate the road conditions with the comfort scores, and get more ideas from these results. In the meantime, the analysis results could be extended to added-on value services in smart phone for passengers. Passengers could call a taxi with higher comfort score since most of people look for a comfort trip.

Table II shows the ranking lists of the top ten drivers in terms of load frequency. The number of loads frequency is above thirty-five, and load time is over eight hours[1]. Besides, the top one driver has load time as 655 minutes in one day; that is, he works more than ten hours. Driving time also includes unload time for searching passengers. It means that the driver has driven more than 20 hours in that day, and taken the most number of loads. These results are really amazing and break common thinking.

Another ranking list is concerned about comfort scores with passengers. Comfort score is defined by $20 \times (6 - CI_t)$. When $CI_t$ value is smaller than 3, we describe the trip as comfortable [2]. That is, if the comfort score is larger than 60, passengers would feel comfortable with the trip. From Table III, we can see that the top ten comfort scores are around 99, which means that passengers in taxis feel comfort almost during the whole trip.

Lastly, we see the ranking of the worst ten uncomfortable feeling when taxi drivers have passengers in a car. The comfort scores are reduced to around forty. Bad comfort scores might be resulted from heavy traffic conditions or tired taxi drivers. From these ranking lists, we can track back to the drivers' driving bahaviors analysis, spatial-temporal analysis of road/traffic conditions, and so on for the improvement of comfort levels.

## V. Conclusion

In this paper, we have proposed a simple yet effective approach for comfort measurement by exploiting real-time

[1]We mark three number as "xxx" on the taxi number for the privacy concerns.

Table II
THE RANKING OF LOAD AMONG A DAY

| Ranking | The Number of Load | Taxi Number | Date | Load Time(Min.) | Driving Time(Min.) |
|---|---|---|---|---|---|
| 1 | 47 | 9002xxx86 | 2010/11/13 | 655 | 1296 |
| 2 | 42 | 9002xxx65 | 2010/11/13 | 602 | 1367 |
| 3 | 41 | 9002xxx61 | 2010/11/20 | 565 | 1151 |
| 4 | 41 | 9002xxx90 | 2010/11/10 | 533 | 1379 |
| 5 | 39 | 9002xxx65 | 2010/11/27 | 593 | 1379 |
| 6 | 39 | 9002xxx99 | 2010/11/20 | 639 | 1379 |
| 7 | 38 | 9002xxx93 | 2010/11/12 | 575 | 812 |
| 8 | 37 | 9002xxx99 | 2010/11/14 | 556 | 1250 |
| 9 | 37 | 9002xxx99 | 2010/11/10 | 480 | 1281 |
| 10 | 35 | 9002xxx90 | 2010/11/21 | 485 | 1009 |

Table III
THE RANKING OF TOP 10 COMFORT SCORE

| Ranking | Comfort Score | Taxi Number | Date |
|---|---|---|---|
| 1 | 99.999521 | 9002xxx49 | 2010/11/20 |
| 2 | 99.997113 | 9002xxx04 | 2010/11/10 |
| 3 | 99.996921 | 9002xxx22 | 2010/11/14 |
| 4 | 99.995154 | 9002xxx70 | 2010/11/16 |
| 5 | 99.995148 | 9002xxx44 | 2010/11/27 |
| 6 | 99.994319 | 9002xxx93 | 2010/11/8 |
| 7 | 99.992993 | 9002xxx22 | 2010/11/11 |
| 8 | 99.992316 | 9002xxx44 | 2010/11/27 |
| 9 | 99.988476 | 9002xxx06 | 2010/11/20 |
| 10 | 99.987900 | 9002xxx16 | 2010/11/27 |

Table IV
RANKING OF WORST 10 COMFORT SCORE

| Ranking | Comfort Score | Taxi Number | Date |
|---|---|---|---|
| 1 | 39.623528 | 9002xxx14 | 2010/11/23 |
| 2 | 40.398912 | 9002xxx06 | 2010/11/20 |
| 3 | 41.659462 | 9002xxx14 | 2010/11/21 |
| 4 | 42.674304 | 9002xxx14 | 2010/11/18 |
| 5 | 43.892154 | 9002xxx45 | 2010/11/20 |
| 6 | 44.283970 | 9002xxx86 | 2010/11/20 |
| 7 | 44.658356 | 9002xxx05 | 2010/11/28 |
| 8 | 45.529645 | 9002xxx93 | 2010/11/23 |
| 9 | 45.727199 | 9002xxx86 | 2010/11/27 |
| 10 | 46.843917 | 9002xxx86 | 2010/11/28 |

trajectory data reported by taxis in greater Taipei area. Using the realistic data collected by one of the taxi service providers in Taipei, we have conducted a comprehensive set of analyses, and the results shown that the comfort level varies with the trip time and the trip distance. We also give the rankings of all taxi drivers according to a number of criteria, such as the comfort score and the number of loads. With the help of ranking lists, we can track back to the drivers' driving bahaviors analysis, spatial-temporal analysis of road/traffic conditions, which may be used to apply on more IoT applications, such as traffic management and urban planing. We believe that simple and robust comfort measurements system, which exploits the GPS data can bring more convenience to enrich people's daily life.

## ACKNOWLEDGMENT

This study was supported by the National Science Council of Taiwan under grants: NSC 100-2219-E-001-001.

## REFERENCES

[1] Comfort index table (page 15). http://yctrtrc.ncku.edu.tw/site2/ocwCoursePPT/133trkslide-2009-III-3.ppt.

[2] Iso 2631-1-1997: Mechanical vibration and shock-evaluation of human exposure to whole-body vibration.

[3] Japanese industrial standard c1510: Vibration level meters.

[4] S. S. Andaleeb, M. Haq, and R. I. Ahmed. Reforming innercity bus transportation in a developing country: A passenger-driven model. *Journal of Public Transportation*, 10:1–25, 2007.

[5] R. K. Balan, K. X. Nguyen, and L. Jiang. Real-time trip information service for a large taxi fleet. In *ACM MobiSys*, 2011.

[6] M. Cantwell, B. Caulfield, and M. O'Mahoney. Examining the factors that impact public transport commuting satisfaction. *Journal of Public Transportation*, 12:1–12, 2009.

[7] J. Disney. Competing through quality in transport services. *Managing Service Quality*, 8:112–118, 1998.

[8] L. Eboli and G. Mazzulla. A new customer satisfaction index for evaluating transit service quality. *Journal of Public Transportation*, 12:21–88, 2009.

[9] B. Edvardsson. Causes of customer dissatisfaction - studies of public transport by the critical-incident method. *Managing Service Quality*, 1998.

[10] C.-Y. Lin, L.-J. Chen, Y.-Y. Chen, and W.-C. Lee. A comfort measuring system for public transportation systems using participatory phone sensing. In *ACM PhoneSense*, 2010.

[11] J. Yuan, Y. Zheng, C. Zhang, W. Xie, X. Xie, G. Sun, and Y. Huang. T-drive: driving directions based on taxi trajectories. In *ACM SIGSPATIAL GIS*, 2010.

[12] Y. Yue, Y. Zhuang, Q. Li, and Q. Mao. Mining time-dependent attractive areas and movement patterns from taxi trajectory data. In *Geoinformatics*, 2009.