

## **Bus Route Speed Tool Using AVL Data**

**Alex Cui  
Chicago Transit Authority**

**Planning and Development  
567 W. Lake Street 10<sup>th</sup> floor  
Chicago, Illinois 60661**

**Tel: 312-681-4226**

**Fax: 312-681-4297**

**[a\\_cui@hotmail.com](mailto:a_cui@hotmail.com)**

## **Bus Route Speed Tool Using AVL Data**

### **Abstract**

The current abundance of automatically generated transit data provides a unique opportunity to greatly improve our ability to understand the transit system in a far greater detail and to utilize such understandings in transit planning and operations in ways that is not possible before. This paper will introduce a newly developed tool in the Chicago Transit Authorities (CTA)—Bus Route Speed Tool that is based on the available Automatic Vehicle Location (AVL) data. The information of detailed average bus speed at practically any point along the bus route can be used in transit planning and operations to help evaluate measures such as rerouting, near side versus far side boarding, techniques to speed up boarding and alighting, and the locations for Traffic Signal Priority (TSP) among others. It can also help to identify traffic bottlenecks and to provide inputs on selecting time points.

This tool calculates the bus speed along a bus route in 500-foot intervals by tallying the distance traveled by the buses and the corresponding time as recorded in the AVL database. The bus speeds are then averaged hourly over selected date range to obtain average speeds for these 500-foot segments. The results are presented on Google Map through a intranet web server and viewable from any Flash enabled web browser within the internal network. The web Graphic User Interface (GUI) allows for customized displays of multiple bus routes color coded by speed.

### **I. Background and introduction**

The current abundance of automatically generated transit data provides a unique opportunity to greatly improve our ability to understand the transit system in a far greater detail and to utilize such understandings in transit planning and operations in ways that is not possible before. Currently every Chicago Transit Authority (CTA) bus is equipped with Automatic Voice Annunciation System (AVAS) in compliance with the ADA requirements which tracks and records the location of the bus. Such Automatic Vehicle Location (AVL) data is saved in an Oracle database for later analysis.

There have been a lot of transit applications utilizing AVL data to obtain the bus running time distribution, to ensure compliance of terminal departure policy, to measure the bus on time performances and to analyze the headway distributions, besides many other valuable applications.

Estimates of the bus speed along the route can be obtained by measuring either the end-to-end running time or the time point to time point running time and the distances between the terminals or the time points. However, the bus speed data thus obtained are usually not detail enough to form a meaningful speed profile along the route.

The City of Chicago's Office of Emergency Management and Communications (OEMC) is currently developing a project that will in part be using the CTA's Bus Tracker data to provide real time information on arterial traffic conditions.<sup>1</sup>

An application to estimate Stop Level Bus Speed using AVL data was developed previously for the CTA to obtain the stop level speed for transit planning and operations purposes. However, the Stop Level Bus Speed application has several disadvantages. First, bus almost never stop at the exactly same location, and sometimes stopping both before and after an intersection will both yield valid data. As a result, the resulting speed data will require further analysis to avoid misleading results. Second, for the portion of the routes with no stops or with few stops such as the express portion of a route, the Stop Level Bus Speed application will not yield enough details. Stop Level Bus Speed will be discussed in further detail in part II of this paper.

This paper will introduce a newly developed Bus Route Speed Tool that provides bus speed data along a route in 500-foot segments. The outputs allow us to see the variation of the speed along a bus route in much finer detail. Such information can be used in transit planning and operations to help evaluate measures such as rerouting, near side versus far side boarding, techniques to speed up boarding and alighting and the locations for Traffic Signal Priority (TSP) among others. It can also help to identify traffic bottlenecks and to provide inputs on selecting time points.

## **II. Stop Level Bus Speed project review**

The Stop Level Bus Speed application is a predecessor to the Bus Route Speed Tool.

### **1. Data sources**

The Stop Level Bus Speed project relies on the stop level AVL data. This data comes from AVAS system which will save a record of bus location and corresponding time for every stop event. There are three types of stop events: serviced stop, unserviced stop, and unknown stop. The serviced stop is where the bus actually stopped to let passengers on or off. The unserviced stop is a designated stop that was passed because there is no passenger boarding or alighting. And the unknown stop is when the bus stopped to let the passengers get on or off, but the location was more than 150 feet away from a designated stop. This could be due to, for instance, when a bus stopped mid block to pick up a passenger or when a bus in traffic let passengers off too far before it reached the stop. For the Stop Level Bus Speed, only serviced stop and unserviced stop events are used.

### **2. Methodology**

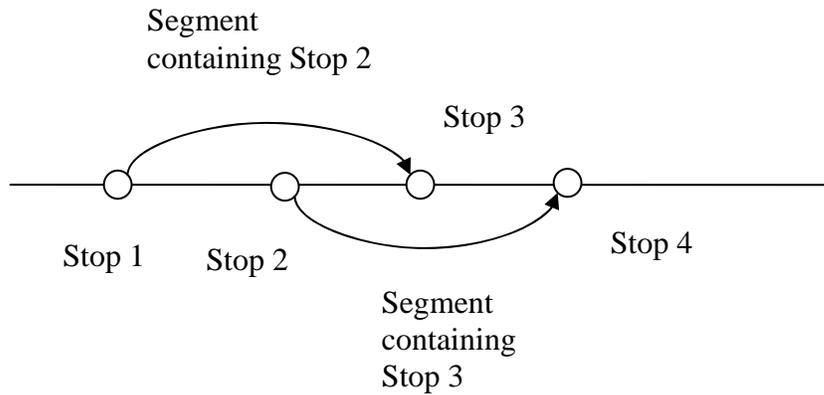
The Stop Level Bus Speed application is initially developed in order to identify which intersections along a bus route<sup>i</sup> are the major choke points. Therefore, the primary goal is to be able to provide some indicative information on major intersections, and the

---

<sup>i</sup> Route in this paper is considered to be one direction of a bus route with a particular routing. This is equivalent to a pattern in the terminology of many transit agencies.

segments upon which the speed is calculated is defined as from the previous stop to the next stop. In other words, for each stop, the resulting speed is over a two-stop segment. See figure below.

**Figure 1. Segment definition for Stop Level Bus Speed**



The first advantage of this segment definition is that, for the main intersections, the variation of where the bus actually stop is less important because the previous and next stops of a main intersection are mostly minor intersections with higher chance of being skipped by the bus. A more important advantage is that if the segments are defined as between to consecutive stops, a far side stop that have a lot of bus actually stop near side (or vice versa) may give very misleading results as the bus speed passing the intersection. With the segment as currently defined, the corresponding speed of a stop can indeed be an indicator of how fast a bus passes through the intersection near the stop without having to know first whether the stop is a near side stop or a far side stop.

The speed of one bus over each segment is calculated by dividing the difference of odometer readings at the end points of the segment by the difference of event time at the end points. Usually five weekdays of data will be aggregated to arrive at an average speed for this segment over a selected time period such as the AM peak.

### **3. Applications and issues**

Stop Level Bus Speed is an improvement over route level or time point level bus speed data by providing much more detailed speed variation along the route. Therefore, the output data from the Stop Level Bus Speed can be used for a variety of purposes. It has been used to help evaluate bus routing options, to help evaluate backdoor boarding initiative, and to help identify bottlenecks along a bus route.

From these and other potential applications, we have identified several issues with the Stop Level Bus Speed application as follows:

First, since it is often used for purposes not specifically designed for when it was first implemented, the segment definition of from the previous stop to the next stop rather than

just between two consecutive stops does cause some inconveniences for some applications despite its benefits described above. Sometimes, running time over a portion of the route covering several stops are desired, due to the segment definition, such information may or may not be readily and accurately obtained depending whether there are even or odd number of stops involved. This problem can be relatively easily solved by changing the segment definition.

A slightly more difficult issue is caused by the fact that the locations of stop events are scattered within 150 feet of the designated stop. If such distribution is truly random with similar degree of variation, increasing the size of the sample would be able to achieve desired accuracy. However, there tend to be more scattering for stops with higher boarding and more impacted by vehicular traffics. Factors such as the location of a stop can also cause the average location that buses stop to be away from the designated stop even though it is within 150 feet. This can create errors that are more systematic in nature. This may not be a serious concern if the accuracy requirements of the data are not high.

An issue that is much harder to mitigate is for the express route or, in general, for any portion of the route where there is no stop or has a very low stop density. Because only AVL stop event data is used in the Stop Level Bus Speed application, the minimum segments possible are between stops. For the express portion of the routes along Lakeshore Drive, for example, the segments will be very long and the resulting data would lose its details.

These deficiencies as discovered in practical applications, and are the main reasons for the new Bus Route Speed Tool which will be discussed in detail in the next parts of this paper.

We also need to keep in mind that the speed for each segment is still an average over the whole segment and can not provide an accurate estimate of the extreme values or the variations of the instantaneous speed that the bus has experienced.

### **III. Bus speed estimation methodology**

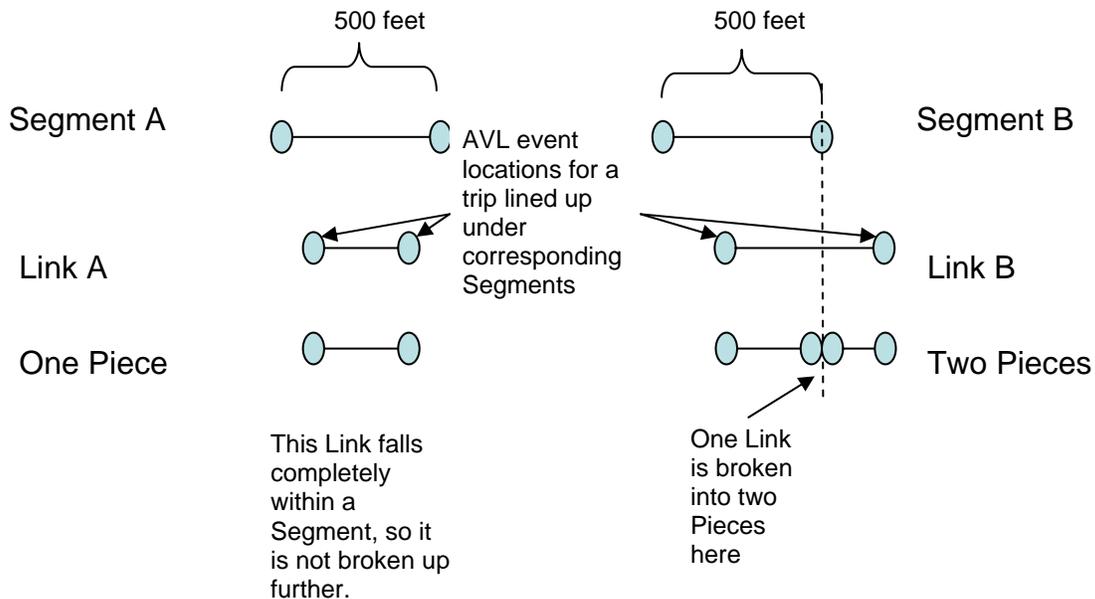
#### **1. Data sources and introduction**

Besides the stop level AVL data used in the Stop Level Bus Speed application, the new Bus Route Speed Tool also uses the distance type of AVL data (Distance Data). An distance type event is triggered by AVAS approximately every 500 feet along the route, and a data record is entered and saved with the corresponding odometer readings and event time. However, the actual distance between two consecutive Distance Data events can vary as much as 200 to 300 feet. For the stop level data, instead of only using the serviced stop event and unserved stop event data, the unknown stop event data is also used as well.

This application will use a linear coordinates system with only one coordinate starting from the originating stop and running along the bus route. In other words, any point on the route can be specified by using its distance from the originating stop.

For the purpose of this application, a Segment is defined for every 500 feet along the route starting from the originating stop. A Link is defined as a portion of the route between two the locations of two consecutive AVL events of any acceptable type for a given bus trip. A trip is a bus traveling from the originating stop to the terminal. A Piece is a Link or a portion of a Link that lies completely within a Segment. It is possible to define a Segment to be shorter than 500 feet, but a shorter segment will only be useful if there is enough data to support the increased details. See figure below.

**Figure 2. Segment, Link, and Piece definition**



The methodology detailed below is implemented in PL/SQL language in an Oracle database, and the entire process is fully automated feeding from AVL data and the electronic form of the schedule and route data. Minimal manual maintenance is required.

## 2. Trip length variation

The Segments for each route is created based on the measured route length or the standard route length. However, each actual trip may have a different end-to-end length due to odometer errors and the variation of actual distance traveled. Different bus trip may make different number of stops which may increase the actual distance traveled slightly. The presence or absence of various traffic maneuvers may also cause slight variation in the actual distance traveled. These are considered to be the acceptable variations. On the other hand, the difference of distance traveled due to detours, short turns, or navigation problems are not acceptable, and trips with these issues will have to be discarded.

To eliminate trips with unacceptable variations, a threshold of 2% was set, and any trips that has an end-to-end actual odometer distance that is more than 2% greater or less than the standard route length is discarded. The great majority of the trips fall within the 2% threshold.

For the trips that are acceptable, a scaling factor is applied, so the end-to-end length of each trip is normalized to be the same as the standard route distance.

### **3. Segment speed data**

A Link is defined using all valid AVL event types in order to use as much data as possible. The AVL events tend not to be evenly distributed along the route because the stop level AVL data most likely will be at every block which is approximately 600 feet apart while the Distance Data is about every 500 to 700 feet apart. There is a certain degree of scattering of the data so the coverage is reasonably good, but if Segments are defined small enough, some will have relatively significantly fewer AVL events than others due to the location. This presents a practical limit on how small the Segments can be. Segment that does not contain an AVL event will get a speed through interpolation that is similar to the Segments that are next to it, and in this sense having short segments does not give wrong results. However, the resulting lack of speed variation across different short Segments is actually due to the lack of data which can be quite misleading to ordinary observers.

A trip is broken down into many Links and compared against the route with the standard Segments. Each Link will then be either completely within one Segment or run across two or more Segments. For the Links that are not completely within one Segment, they will be cut at the Segment boundaries into two or more Pieces, such that each Piece is completely contained within one Segment. Reviewing Figure 2, we can observe that even though there may be multiple Segments corresponding to multiple Links, there only one Segment corresponding to each Piece and only one Link corresponding to each Piece as well. Therefore breaking the Links into Pieces allows us to join the Pieces with the Segments.

The distance relative to the originating stop and the time relative to the start of the trip are known for the end points of each Link. When a Link is cut into Pieces, the newly created end points of the Pieces will have interpolated distance and time using the end points of the Link assuming the travel speed is constant over the Link. Therefore, the distance to originating stop and the time elapsed from the start of the trip for the end points of any Piece are known, and the length of the Piece and the time spend on it can be obtained by taking the difference of respective values of two end points of the Piece. Speed in MPH over the Piece can then be calculated.

Each Piece with distance and time know is therefore a data point for the corresponding Segment. The GIS coordinates of the Segment is also kept through the process to enable mapping capability.

## 4. Aggregation

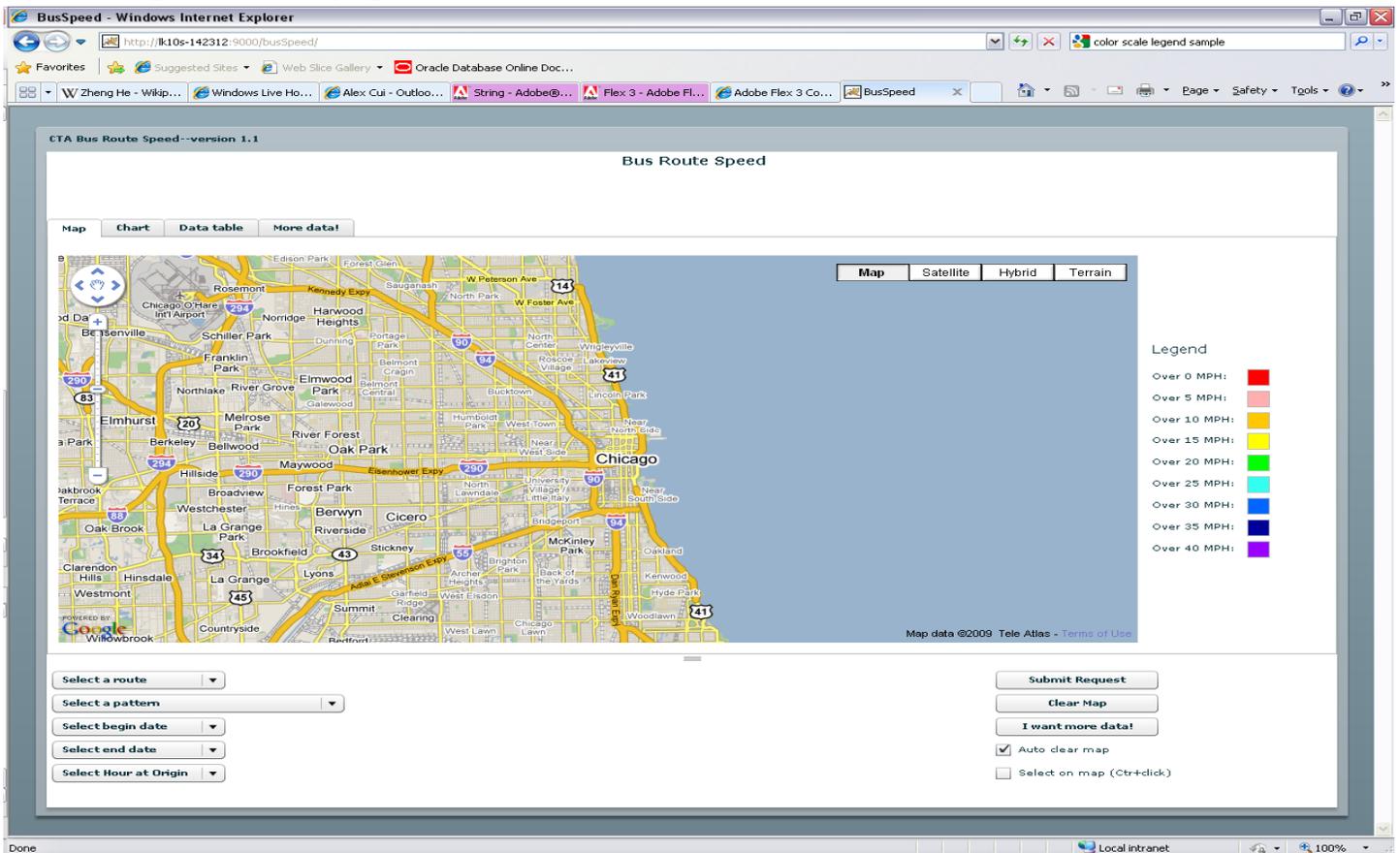
To obtain accurate average speed for the Segment, data from many trips are needed. Because there are usually significant variations of the travel speed over different time of the day, it usually makes little sense to aggregate all the trips over a day. In fact, the shorter the time period of the day the sharper the speed variation can be shown. Therefore, for this application, all the Pieces for a segment over each hour are aggregated. To ensure there is enough data, usually five weekdays' data is used. Average speed for the Segment for the time period is obtained by aggregating and calculating using the data from all these Pieces. Repeating these steps for all the Segments will yield a full speed profile.

## IV. Web based integrated tool

### 1. Introduction

The results from the application can be viewed directly or exported for further analysis. A web Graphic User Interface (GUI) is provided to allow user control of the processing and to display the resulting data in an integrated platform.

Figure 3. Web page overview



The GUI allows the user to select the route and pattern for a particular date range, and speed profiles will be automatically calculated for every hour of the day. Requests from multiple users can be handled simultaneously. See figure below.

**Figure 4. GUI new data request**

Select a route

Beginning date:

Ending date (not included):

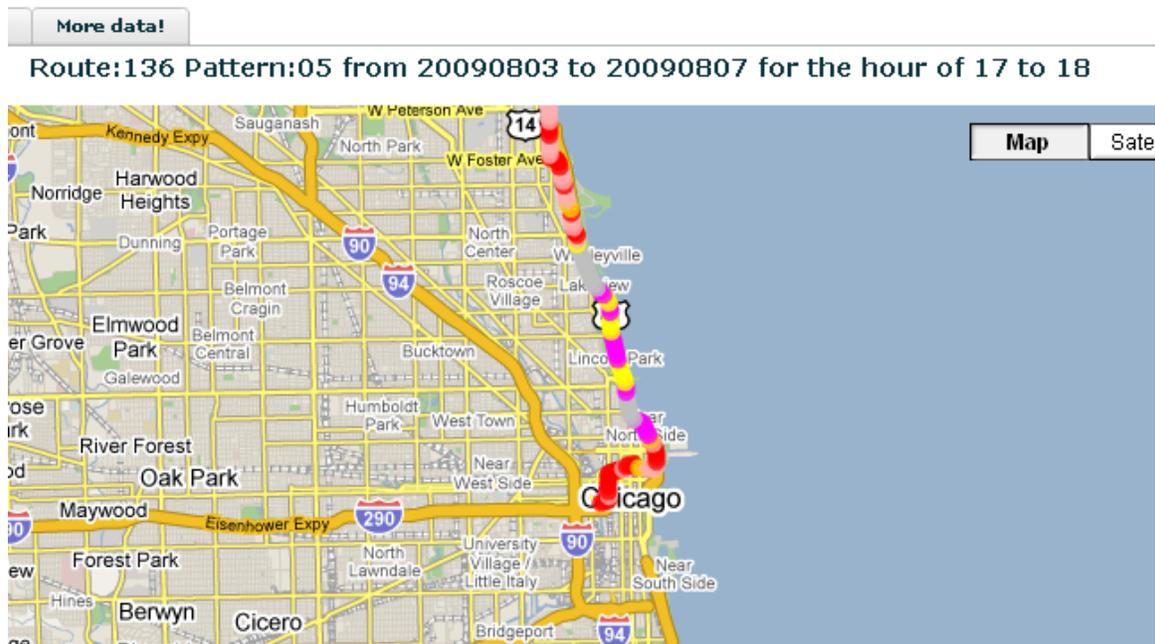
You can optionally leave your email address here to receive a notification when your request is complete.

After you submit the request, check back in 30 minutes or after you receive email notification.

Select a pattern

The resulting data is displayed in three ways: on a map, in a speed profile chart and in a tabular form. Usual Google Map controls such as zooming and moving are allowed. Segments can be selected on the map to automatically calculate the total running time and the corresponding speed in MPH for all the Segments selected. See figures below.

**Figure 5. Color coded speed on a map**



**Figure 6. Speed profile for the same data**



**Figure 7. Data in tabular form for the same data**

Distance from terminal	Route	Time	Speed
0	136	211.96	1.599424984783167
500	136	149.6	2.278804083616918
1000	136	82.8	4.117259552042161
1500	136	59.12	5.766391930126707
2000	136	96.6	3.529079616036138
2500	136	82.44	4.135238851396057
3000	136	95.88	3.555580839685971
3500	136	186.88	1.8242138854296388
4000	136	156.92	2.1725024911361897
4500	136	86.68	3.932961362587574
5000	136	96.84	3.520333445984003

## 2. Software used and program structure

Google Map Flex API is used to display the map on an Adobe Flash based web page with dynamic displays. Open Flash Chart is used to display graphs in Flash. Java Server Page (JSP) on Tomcat is used to serve the data to the web page, and JDBC is used to connect to the backend Oracle database where the processing is done.

The whole program is a standard three-tiered web application with a backend database layer, a mid-tier using JSP, and a client based on web page with Flash.

## 3. Walkthrough of the software

This section is a distilled version of the first scenario walkthrough of *Tutorials—Bus Route Speed Tool*<sup>2</sup> which is a walkthrough instruction on how to interact with the web GUI. Please refer to this document for complete graphic tutorials.

The scenario is that a planner wants to get the running time for Route 14 southbound between Jackson and Roosevelt during the peak of PM rush-hour. The following are the steps to get the requested running time.

First, submit the request. See figure below.

**Figure 8. Submit the request**

Select a route  
Route:136  
Route:14  
Route:146  
Route:147  
Select hour at origin

Route:14  
Select a pattern  
01:[14]WASH+JEFFRSN - 103+STONY ISL  
04:[14]103 STONY ISL - WASH+JEFFRSN  
Select Hour at Origin

1. Select route

2. Select pattern

Route:14  
Select an hour  
Hour:04-05  
Hour:05-06  
Hour:06-07  
Hour:07-08  
Select an hour

103+STONY ISL

Submit Request  
Clear Map  
I want more data!

Auto clear map  
 Select on map (Ctrl+click)

3. Select hour of the day

4. Submit the request

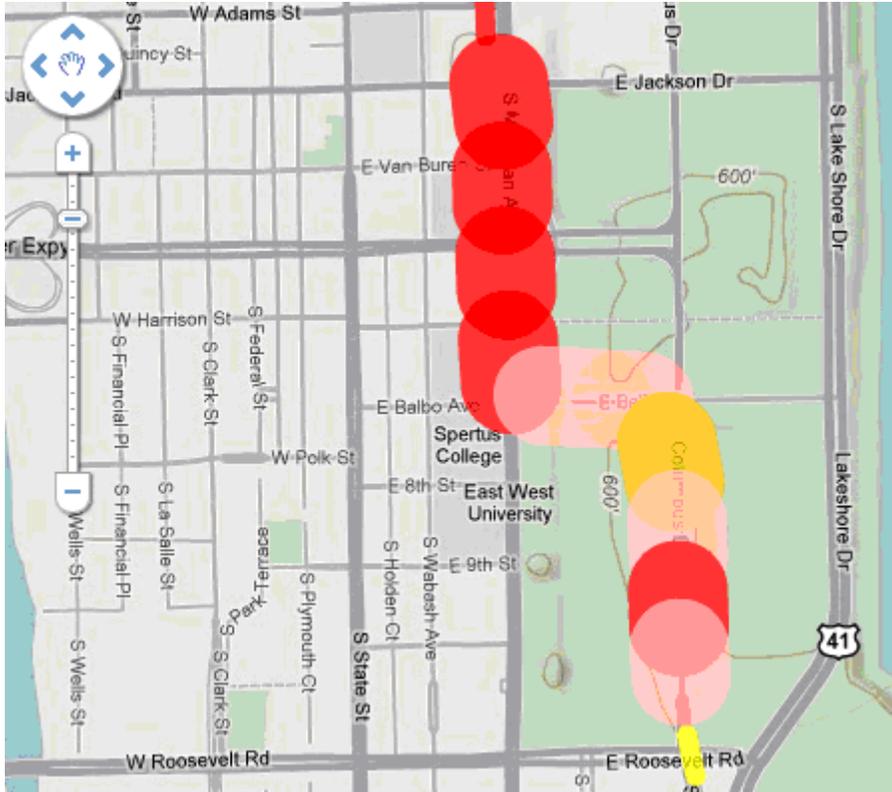
Next, zoom and span to the interested section of the map. See figure below.

**Figure 9. Interested section of the map**



With “Select on map” checkbox selected, holding down the Control Key and select all Segments between Jackson and Roosevelt. See figure below.

**Figure 10. Selected Segments**



The selected Segment can be reviewed in the tabular data display form and any missing Segments be selected. The distance, running time and average speed in MPH for all the Segments selected can be read from the same page. See figure below.

**Figure 11. Resulting data**

Selected Distance (feet):	5000
Selected Average Runtime (sec):	630.4545454545455
Average Speed Over Selected Segments:	5.407354001441961

## **V. Discussions and further improvements**

Detailed bus speed profile data is previously not generally available. Its use in transit planning and operations is thus not established. Besides the applications mentioned in this paper, there are possibly other applications or better ways to utilize the data produced that may become evident later on.

There are quite a few potential improvements that can be made to the Bus Route Speed Tool, and these are discussed below. If demand warrants, they can be implemented in a later version.

### **1. Integrate closely with stop, time point and route information**

One issue with the data from the Bus Route Speed Tool is that the data must be mapped to see where a Segment is at since all that is known for the Segment is the distance to the originating stop and the GIS coordinates of the Segment. It will be difficult to say this Segment is between which and which stops without mapping the data. A complete solution to this would be to add linear coordinate information to the stops and time points. In other words, we can add the distance to the originating stop for each stop and time point. Currently the Tool is mostly viewed through Google Map, but the ability to view data without a map is also valuable.

If the route shape data can be transferred to use linear coordinates, a standard set of Segment definition common across different directions and patterns of a route can be used. Data from different patterns can then be easily compared.

### **2. Fine grain data acceptance**

Currently, only the data within a trip whose end-to-end length is within 2% of the standard route length is treated as acceptable. A trip that has a small detour or otherwise become too long or too short would completely be discarded. Even though increasing the number of days in the samples will easily make up for the discarded data, the potential for bias can not be fully discounted. It is possible that a very crowded bus making all the stops will have a higher chance of being discarded. And this, if true, can make the estimated average speeds higher than they actually are.

To counter this potential bias, we should discard at stop interval level rather than at the trip level as currently. We will only discard data if the corresponding odometer difference between two stops is too different from the standard distance between these two stops. This way, even if a portion of the trip is not acceptable, the rest are still retained. By reducing the discarded data, potential bias will be reduced as well.

### **3. Improvement of GUI**

There are several possible improvements to the web GUI as well. First, the color coding can be made more intuitive. Also because a sizable portion of the CTA's route network has an average speed of below 10 MPH, the color coding should provide relatively more details to the lower speed Segments.

### ***Acknowledgements***

I have received valuable advices from several people especially in our group of Planning Analytics at the CTA. My thanks go to all who has made suggestions or has given feedbacks on the Bus Route Speed Tool.

### ***References***

---

<sup>1</sup> Chicago Transit Authority. 2009 Sep. 9<sup>th</sup>. “CTA to Introduce Developer Tools Web Page”. CTA press release 9/9/2009.

<http://www.transitchicago.com/news/default.aspx?Archive=y&ArticleId=2450>.

<sup>2</sup> Cui, Alex. 2009. Tutorials—Bus Route Speed Tool. Unpublished training material.