Special edition paper

Study of a Transportation Impact Index Based on the Impact on Passengers

Hiroshi Fukuyama*

The current index indicating the scale of transportation disruptions is based on train operators' data such as the number of cancelled train operations or train delays (total time and the number of delayed trains) and also on data such as the total number of affected passengers. Such an index, however, is never based on how individual passengers view the impact of transportation disruptions. Therefore, based on the impact of transportation disruptions on passengers, we have been studying a method of evaluation of overall affected people-minutes based on the number of affected passengers and lost time, and also quantification of how the affected passengers perceive our transportation services. This paper introduces the concept of overall affected people-minutes, quantification of understanding of delay time and passenger dissatisfaction due to various transportation disruption causes, and the POINT system using intuitive expressions.

Keyword : transportation disruption, index, queuing theory

1 Introduction

When a transportation disruption (disruption in a train schedule for some reason) occurs, its scale is expressed based on train operators' data such as the number of cancelled trains or train delays, and also based on data of the total number of affected individuals. Meanwhile, one of the basic management approaches announced in JR East's medium-term business plan "New Frontier 2008" is "Provide services taking passengers' viewpoints into account," and, we have therefore been pursuing the establishment of passenger-oriented services. In response to this, the Safety Research Laboratory has been studying quantification of the scale of transportation disruptions and establishment of an index that will allow evaluation of the impact of transportation disruptions from the passenger's point of view. By establishing such a transportation impact index, we believe that we will be able to quantitatively understand our transportation safety and implement effective measures based on such understanding.

2 Transportation impact index

2.1 Concept of transportation impact index

First, the concept and calculation method of a transportation impact index, which is expressed in the unit of a "person-minute," will be described here. In general, when establishing an index for events having various sizes or occurrence frequencies, an analysis is carried out by multiplying the event's occurrence frequency by the post-disruption impact. In other words, when the transportation disruption frequency is expressed as "F" and the impact of the disruption is expressed as "C" then the overall transportation impact "I" can be obtained from equation (1):

$$I = F \times C \quad \text{... equation (1)}$$

In this study, train delays and the number of affected passengers are analyzed to obtain the post-disruption impact. Also, in order to evaluate the impact on passengers, it is necessary to consider the number of affected passengers and the time that passengers had to lose instead of delays or cancellation of train operations. Since the time necessary for passengers to arrive at the destination is usually increased when a transportation disruption occurs, if that increment is multiplied by the total number of affected passengers, then it will become possible to qualitatively understand the impact on the passengers. This is the foundation of the transportation impact quantification.

Now, as the basic concept of a new transportation impact index which allows understanding of the transportation impact from the passengers' point of view, overall affected people-minutes will be used. Overall affected people-minutes can be defined as equation (2) below. When the number of passengers on an affected train i is expressed as "Pi", and when a delay of that train is expressed as "Ti", then the overall affected people-minutes "R" can be expressed as follows:

$$R = \Sigma\,(P_i \times T_i) \quad \text{... equation (2)}$$

For example, assuming there are 100 passengers on a train that is 30 minutes late, and there are another 200 passengers on a train that is 20 minutes late, then the overall affected people-minutes can be calculated as 30 (minutes) x 100 (people) + 20 (minutes) x 200 (people) = 7,000 (people-minutes).

* Safety Research Laboratory, Research and Development Center of JR East Group

2.2 Approach to calculation of the transportation impact index

Unfortunately, it is impossible to actually measure the number of affected people and minutes that were defined above. Therefore, they are estimated based on a set of assumptions.

The assumptions are as follows:

(1) As the basic data indicating the number of affected people, the data from the annual line traffic volume survey is used to calculate the number of passengers on each train, inbound and outbound, for weekdays, Saturdays, and holidays.

(2) If passengers were first affected by a transportation disruption during the time between the occurrence of the disruption and train operation resumption, then it is classified as impact type (A), and if passengers were first affected after train operation resumption, then it is classified as impact type (B).

(3) For impact type (A), the maximum transportation capacity of one train after train operation restarts is assumed to be 250% of its passenger capacity. Passengers who are unable to get on that train due to its crowdedness are assumed to have to wait for the next train which is assumed to arrive with the regular train interval.

(4) For impact type (B), although it is difficult to accurately know the train interval, the queuing theory can be used to estimate a train interval and then to calculate the impact imposed on passengers before the regular train schedule is reestablished.

(5) Passengers are assumed to keep arriving at a station at the average rate.

Figure 1(A) shows the impact observed before train operation resumption, and Figure 1(B) shows the impact observed after train operation resumption. In order to facilitate the understanding of the concept of the index, in Figure 1, time is divided into increments of 10 minutes, and the regular train occupancy rate is maintained at 67%. In this example, it is assumed that a transportation disruption occurred at 13:30, train operation restarted at 15:00, and the regular train schedule was reestablished at 16:30.

In both Figures 1 (A) and (B), for each time band the number of affected passengers is indicated by the difference created by subtracting the number of passengers who actually got on a train from the number of passengers that usually get on the train. The increase in wait time for a train can be obtained by the number that is shifted towards the right on the graphs, and this number is the difference between the number of passengers that actually got on the train and the number of passengers that usually get on the train.

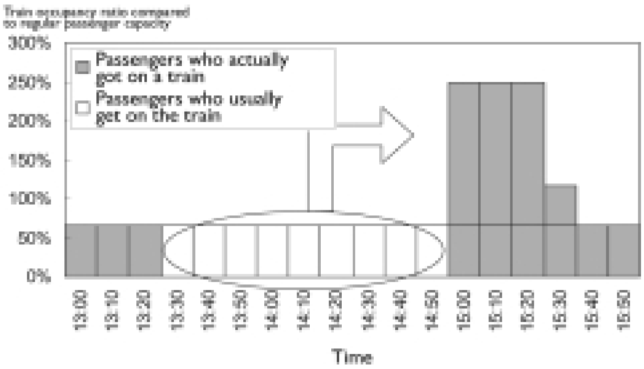

Fig. 1 (A): Impact before Train Operation Resumption

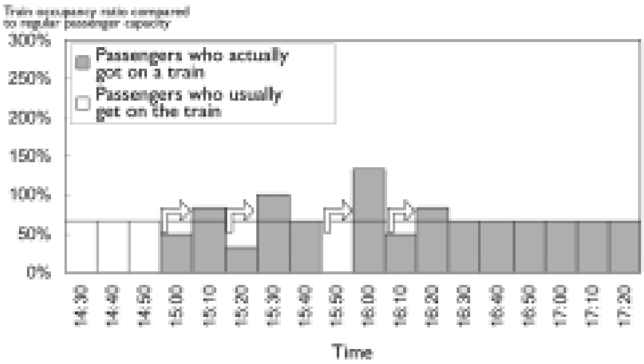

Fig. 1 (B): Impact after Train Operation Resumption

2.3 Detailed calculation of transportation impact after train operation resumption

In this section, a detailed explanation of the method for calculating impact-type (B) will be provided. In the Tokyo metropolitan area where trains run at intervals of every few minutes, a transportation disruption seems to cause little impact from the passengers' point of view since trains are operated in sequence even if they are behind schedule. Passengers are believed to enter a station and get on a train that just arrived to go to their destinations without being much aware of train departure time, and therefore, in terms of wait time, the actual impact of the transportation disruption is small. Hence, as described above, for the case of impact-type (B), the amount of affected time is obtained based on the amount of time the passengers have to wait for the next train, which is the time between the occurrence of a transportation disruption and the time that the regular train schedule is reestablished, to understand the reality of the transportation impact.

The increase in amount of wait time after train operation resumes ("a wait time increment" hereinafter) is the difference between wait time after train operation resumes and wait time in normal operations. By obtaining the probability that the wait time becomes equal to or longer than the average train operation interval, and also by obtaining its expectation value, it is possible to estimate the accumulated wait time increments when trains are late. This is the total wait time increment after train operation resumes, and the product of this value and the number of passengers is the increase in the impact scale after train operation resumption. Note, however, that there is almost no actual measurement data of train intervals after train operation resumption, and also, recording of such data at the time of transportation disruption is not quite realistic. Therefore, cases where the wait time becomes equal to or longer than the average train operation interval and the probability for these cases to occur are logically examined in order to obtain approximated values.

Generally, when the frequency of a certain event occurring during a certain period of time is consistent with the Poisson distribution, the probability distribution y of the time intervals x for that event to occur is known to take the form of the exponential distribution $\lambda e^{-\lambda x}$ [e is the base of a natural logarithm] (the queuing theory). The average wait time is $1/\lambda$ of the exponential distribution. In order words, assuming that the train operation interval that exceeds the shortest train operation interval that is mechanically possible forms the exponential distribution, the definition of the exponential distribution is used to obtain that interval and the probability of that interval to occur.

When

t_0: the shortest train operation interval that is mechanically possible (in minutes),

t_u: the average train operation interval under normal operation (in minutes),

W: a train operation interval at an arbitrary time after train operation resumption (in minutes),

t_m: the average train operation interval after train operation resumption (in minutes), and

$t_m - t_0$: the average value of the interval that is exponentially distributed,

then t_m is the average value of W and is $\frac{1}{\lambda} = t_m - t_0$ in accordance with the definition of the exponential distribution; therefore, λ can be obtained in equation (3).

$$\lambda = \frac{1}{t_m - t_0} \qquad \text{... equation (3)}$$

Also, the increment ratio when waiting for the train interval to become longer than the minimum interval and become W is the area ratio s expressed as a shaded area in relation to the area of triangle n_1; therefore, the value s when $n_1 = 1$ can be obtained by using the next equation.

$$(s + n_1 + \cdots + n_i) : n_1 = (x + t_0) : t_u \qquad \text{... equation (4)}$$

The value n_j changes as $x+t_0$ becomes larger than multiples of an integer of t_u $i-1$ as follows:

$$n_j = \begin{cases} 1 & (j=1,2,\cdots,i-1) \\ \left\{\mathrm{mod}\left(\frac{x+t_0}{t_u}\right)/t_u\right\}^2 & (j=i) \end{cases} \qquad \text{... equation (5)}$$

Therefore,

$$s = \left(\frac{x+t_0}{t_u}\right)^2 - \left[\mathrm{int}\left(\frac{x+t_0}{t_u}\right) + \left\{\mathrm{mod}\left(\frac{x+t_0}{t_u}\right)/t_u\right\}^2\right] \qquad \text{...equation (6)}$$

Note, however, that int (x) is the largest integer value that is smaller than the value x, and mod (x) is the surplus of a division x.

Assuming the expectation value of s is S_T in the case where the value W changes before the regular train operation interval is reestablished, S_T is the total area s and is the wait time increment ratio expected by S_T. Also, since the probability of the train interval becoming W at an arbitrary time is $\lambda e^{-\lambda x}$, S_T can be obtained as follows.

$$S_T = \int_{t_u - t_0}^{\infty} [\lambda e^{-\lambda x} \times s] dx \qquad \text{... equation (7)]}$$

If the value s is obtained based on Figure 2, meaning the concept of equation (6), then the integration value S_T must be separately obtained for different cases. For this reason, a simple equation (8) can be established.

$$s' = \left(\frac{x+t_0}{t_u}\right)^2 - \left(\frac{x+t_0}{t_u}\right) \qquad \text{... equation (8)}$$

Here, as seen in Figure 3, triangle n_i changes from being homologous to triangle n_1 to being a triangle having the same height as triangle n_1, and therefore, s' is undervalued as $s' < s$. The difference between these two however is small as seen in Figure 4, and thus the value s'can sufficiently substitute for the value s.

The wait time increment ratio S_T after train operation resumption then can be obtained as equation (9), using the value s'.

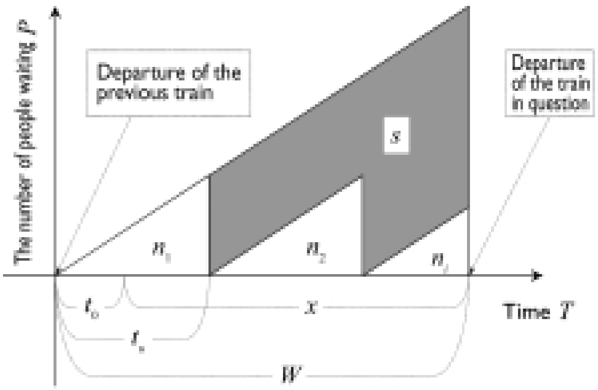

Fig. 2: Conceptual Diagram of the Increment Ratio

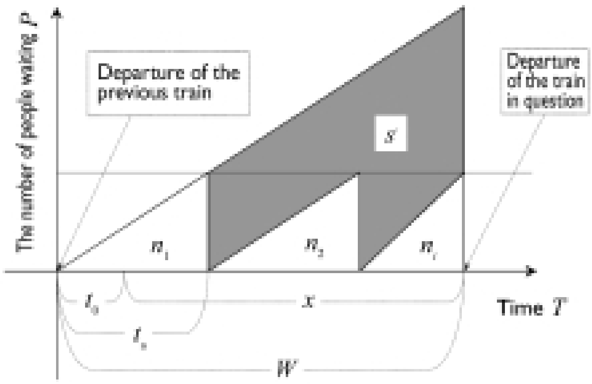

Fig. 3: Conceptual Diagram of s-Value Replacement

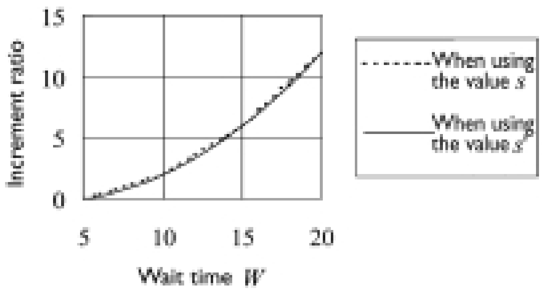

Fig. 4: Difference between Increment Ratios due to Different Calculation Methods

$$S_T \cong \int_{t_u-t_0}^{\infty}\left[\lambda e \left\{\left(\frac{x+t_0}{t_u}\right)^2-\left(\frac{x+t_0}{t_u}\right)\right\}\right]dx$$
$$= \frac{e^{-\lambda(t_u-t_0)}}{t_u^2}\left(\frac{t_u}{\lambda}+\frac{2}{\lambda^2}\right) \quad \text{... equation (9)}$$

The concept of the obtained increment ratio S_T should be briefly explained here. The value S_T is an increment ratio, indicating the increase in wait time based on the regular wait time which is set to the value 1. For example, when the increment ratio S_T becomes 0, the value s in Figure 2 becomes 0. This clearly means that there is no increased wait time. On the other hand, when the increment ratio becomes 1, the wait time becomes twice as long as the regular wait time. Also, assuming that the shortest train operation interval is 2 minutes and the regular train operation interval is 5 minutes, Figure 5 shows the relationship between the operation ratio after train operation resumption and before reestablishment of the regular train schedule, and the increment ratio S_T. In this case, even though the operation ratio is 80%, the wait time increment ratio is over 1; therefore, the resulting wait time is twice as long as the regular wait time or longer.

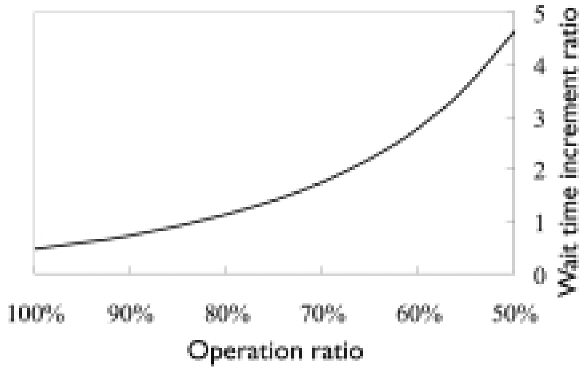

Fig. 5: Relationship between the Operation Ratio and the Wait Time Increment Ratio when Train Operation Intervals are Exponentially Distributed

As described above, the affected people-minutes can be obtained by multiplying the number of passengers who wait for train schedule reestablishment by the product of the increment ratio S_T and the average wait time before train schedule reestablishment. Therefore, when

R_B: impact after train operations resumed (in people-minutes),

T_R: average wait time (= average train operation interval / 2) (in minutes), and

P_m: the number of passengers who wait for the train schedule to be reestablished (in people),

it is possible to use the following equation (10) to obtain the impact after train operation resumption.

$$R_B = T_R \times S_T \times P_m \quad \text{... equation (10)}$$

By using the methods described so far, it is possible to calculate the overall affected people-minutes based on transportation disruption data. Although very few pieces of actual transportation disruption data are available at present, we used some of them to test the calculation methods.

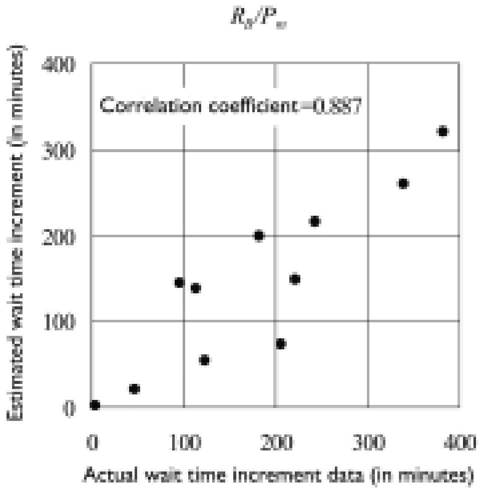

Fig. 6: Comparison between the Actual Data and Estimated Values

As for the actual data, train intervals measured at a station was directly read from the actual train schedule when a transportation disruption occurred. Since it is the train interval that needs to be tested here, the number of passengers is excluded from the scope of the test. Then, R_B/P_m was compared between the actual data and estimated values, and the result is shown in Figure 6. Although some of the estimated values are slightly off from the actual data, the correlation coefficient was 0.887, and this value is considered high for social events and is therefore believed to be a practicable value.

2.4 Case examples for calculation of overall affected people-minutes

Table 1 shows the detailed evaluation results of a transportation disruption that occurred on one of the lines operated in the Tokyo metropolitan area, and the evaluation has been made by using the abovementioned calculation methods.

Cases 2 and 3 each show a transportation disruption that occurred in the city during rush hour. As the "Current indication" column on the table indicates, these cases are about the same scale having almost the same overall affected people-minutes. Case 1 is a transportation disruption that occurred at the end of this line during the night time. As a result of operation adjustment, a pendulum operation was carried out to exclude that site from the route; therefore, the impact on the passengers was small, and accordingly, the overall affected people-minutes obtained was also small. Cases 4 and 5 each show a transportation disruption that occurred in a suburb during the day time, and the scale of the impact is somewhere between the night-time case and the rush-hour cases. Based on the above, therefore, the status of a transportation disruption conventionally expressed by many kinds of values, such as the number of cancelled train operations and the values indicating the train delay status, can be uniquely expressed in the unit of "people-minutes."

Table 1: Case Examples for Calculation of Overall Affected People-Minutes

	Time of occurrence	Current indication	Overall affected people-minutes
Case 1	September, 2003 Around 21:00 Accident resulting in injury or death	23 minutes before train operation resumption, 6 trains were cancelled, and 10 trains delayed (ranging from 23 minutes to 5 minutes)	350,000 people-minutes
Case 2	September, 2003 Around 08:00 Accident resulting in injury or death	48 minutes before train operation resumption, 80 trains were cancelled, and 29 trains delayed (ranging from 44 minutes to 4 minutes)	23.6 million people-minutes
Case 3	August, 1999 Around 17:00 Accident resulting in injury or death	47 minutes before train operation resumption, 81 trains were cancelled, and 33 trains delayed (ranging from 47 minutes to 15 minutes)	23.34 million people-minutes
Case 4	August, 1999 Around 11:00 Accident resulting in injury or death	83 minutes before train operation resumption, 71 trains were cancelled, and 9 trains delayed (ranging from 94 minutes to 3 minutes)	2.04 million people-minutes
Case 5	April, 1999 Around 11:00 Accident resulting in injury or death	53 minutes before train operation resumption, 58 trains were cancelled, and 29 trains delayed (ranging from 48 minutes to 1 minutes)	2.99 million people-minutes

2.5 Issues regarding overall affected people-minutes

It has been pointed out that the indication of the overall affected people-minutes is not intuitive or that values are too large.

As for the former comment, this seems to be because the obtained values are large. However, the resulting values are accumulated values and thus are inevitably large. Note that, as shown as Cases in Table 1, for the two accidents having approximately the same scale that occurred during rush hour, the results of index calculations were almost equal to each other. For the accident that occurred in the day time (for this case in Table 1, approximately 1/10 of the crowdedness compared to the rush hour), the calculation result is smaller than these two cases above. Therefore, we believe it is an appropriate index for comparing the scale of transportation disruptions and for considering the impact on passengers.

As for the latter comment, although a set of assumptions are used when calculating the index, there are still some factors, such as passengers transferring to alternate lines, which are not included in the assumptions. When major lines operated in the Tokyo metropolitan area are analyzed in this regard, the number of people transferred was as low as 10% of the total number of affected people. As for the other items in the assumptions, the values may become large due to accumulation of small errors; however, it is not very likely that estimated values would become significantly different from the actual values.

Therefore, as a solution to these issues, we examined an evaluation method that incorporates passengers' perspectives.

3 Transportation impact index from passengers' perspectives

3.1 Survey on delay time

Based on awareness of a problem that one of the reasons for evaluated values being too large may be overestimation of wait time, we examined an evaluation method that incorporates passengers' perspectives on time spent waiting for a train.

In general, cognition of train delay time is experienced as "frustrations against delays in train arrival time or train operations." Therefore, we tried to evaluate the level of delay time cognition based on the proportion of frustrated people (frustrated people ratio). We conducted a survey and the respondents answered for each time range whether or not they would be frustrated if a train is delayed by that time range (N = 267). We then approximated the average percentages of frustrated people by the logarithmic normal distribution, and multiplied that approximated value by the actual time in order to estimate the level of delay time cognition. The result indicated that almost all respondents answered that they would be frustrated when the actual time is 30 minutes, and if this statement alone is used, the level of delay time cognition after 30 minutes will be the same as the actual time.

In response to this, we used the Weber-Fechner law stating that in general "the magnitude of a subjective sensation increases proportional to the logarithm of the stimulus intensity." We decided to use this law after 30 minutes where the frustrated people ratio reaches 100% and connect this with the approximation curve obtained from the survey. Figure 7 shows the quantified level of delay time cognition.

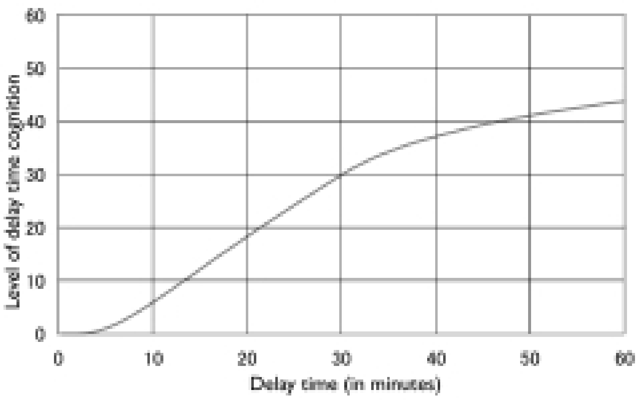

Fig. 7: Relationship between the Actual Delay Time and the Level of Delay Time Cognition

3.2 Analysis of passenger opinions

In this study, the level of frustration felt by passengers was defined as "level of dissatisfaction," and then based on the "passenger opinions" data (comments and requests from passengers to JR East), levels of dissatisfactions were set. Since the "passenger opinions" data would probably be received in larger volume as the scale of a transportation disruption increased, the number of passenger opinions was considered to indicate the level of dissatisfaction towards transportation disruptions. Here, three JR East lines, the Tokaido line, the Chuo line, and the Joban line were selected as typical examples, and for these lines, we examined whether the transportation disruption data recorded from FY2001 to FY2003 and the "passenger opinions" data also recorded during these years would correlate with individual transportation disruptions. Then, the data was categorized according to the causes of transportation disruptions.

In order to examine the quantified relationship between delay time and the level of dissatisfaction, the average number of passenger opinions was examined for each transportation disruption scale. First, transportation disruptions were divided into three types: "disruptions caused by the company" in which disruptions are caused by JR East's own problems such as vehicle malfunctions; "disruptions caused by third parties" in which disruptions are not caused by JR East nor by natural disasters but rather by examples such as wayside fires; and "disruptions caused by disasters" in which disruptions are due to natural disasters such as typhoons. Then, using the "passenger opinions" data associated with transportation disruptions, we first divided that data for each 60 minutes within the total delay time caused by transportation disruptions, and then within each 60-minute range, we plotted the number of passenger opinions in relation to the number of transportation disruptions. Finally, for each disruption type, we conducted regression analysis (Figure 8).

As a result, when transportation disruptions caused by the company occurred the number of passenger opinions rose three times faster than in the case of the other two disruption types. This means that, based on the increase in the number of passenger opinions on the increase in delay time, the impact of disruptions caused by the company is three times as strong as disruptions by other causes. This

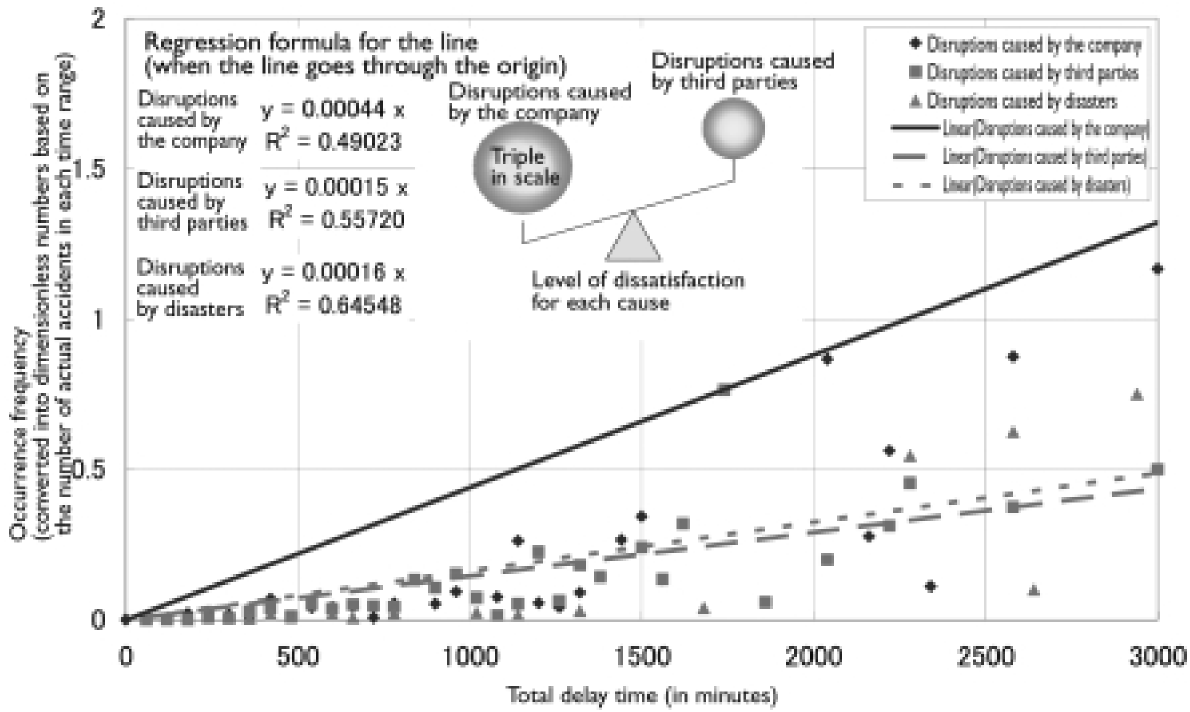

Fig. 8: Average Number of "Passenger Opinions" for each 60 Minutes within the Total Delay Time

tendency remained the same whether or not the regression line went through the origin.

3.3 Index indication method that appeals to passengers' senses

Seismic scales and wind-force scales are some of the examples of already-existing indices. These indices quantitatively indicate scales in values, and add intuitive expressions to describe the typical status of each scale. For example, a seismic level "intensity: upper 5" is a quantitative classification of a measured seismic level and it falls somewhere between 5.5 and 6.0. To this classification, the intuitive expression "there is extreme fear and many people have difficulty taking actions" is added. As for wind-force scales, wind-force level 6 (strong breeze) is a classification of a measured wind-force level and it falls somewhere between wind speed 10.8 m/s and 13.9 m/s. To this, too, the expression "large tree branches sway, electric wires make noises, and it is difficult to hold an umbrella" is added.

We then combined these indices and created the concept of "POINT (Personage Of Influence on Transportation)" that incorporates a scale system. In other words, we tried to classify and evaluate transportation disruptions as seen in Table 2. The descriptions in the "Sensory expression" column in the table are still in draft form and therefore may not precisely match with the actual data. If, however, it is possible to create appropriate expressions as indicated in this draft, then it will become possible to add intuitive expressions when calculating the overall affected people-minutes. With this, we believe that issues regarding overall affected people-minutes can be solved.

Table 2: POINT Classification (draft)

POINT	Impact (people-minutes)	Sensory expression
1	10 people-minutes	No impact
2	100 people-minutes	Slight impact
3	1,000 people-minutes	Some impact
4	10,000 people-minutes	Low dissatisfaction level
5	100,000 people-minutes	Middle dissatisfaction level
6	1 million people-minutes	Relatively high dissatisfaction level
7	10 million people-minutes	High dissatisfaction level during rush hours in the Tokyo metropolitan area
8	100 million people-minutes	Cancellation of train operations for a few hours in the Tokyo metropolitan area
9	1 billion people-minutes	Cancellation of train operations for the entire day in the Tokyo metropolitan area

Table 3: Trial Calculation Examples Using POINT

Date of occurrence	Descriptions	POINT
September 2003	Signal system problems affected train operations all day long.	8.4
March 2005	Vehicle malfunctions that occurred after the rush hours affected train operations during the morning.	7.9
September 2003	There was an accident resulting in injury or death (train operation was restricted for 48 minutes).	7.4
March 2005	There were strong winds over a bridge within a suburban section (train operation was restricted for 41 minutes)	5.6

Table 3 shows the results of rough POINT calculation for some of the specific transportation disruption examples. These POINT values were obtained by adding the results from Section 3.2 and Section 3.3 to the overall affected people-minutes of these examples. The results are almost the same as the sensory expressions listed in Table 2; therefore, POINT seems to be practical.

4 Future objectives

For the various approaches described in this paper, we will examine if they are applicable to the index by conducting various testing using actual data. We then plan to further develop expressions to be used in POINT classifications in order to fully implement the system throughout the company. Also, in accordance with the results of the internal discussion of the "safety improvement committee for transportation," we will use this index as our internal transportation disruption management index.