International Journal of Architectural, Civil and Construction Sciences
ISSN: 2415-1734
Vol:12, No:6, 2018

Estimation of Train Operation Using an Exponential Smoothing Method

Taiyo Matsumura, Kuninori Takahashi, Takashi Ono

***Abstract*—**The purpose of this research is to improve the convenience of waiting for trains at level crossings and stations and to prevent accidents resulting from forcible entry into level crossings, by providing level crossing users and passengers with information that tells them when the next train will pass through or arrive. For this paper, we proposed methods for estimating operation by means of an average value method, variable response smoothing method, and exponential smoothing method, on the basis of open data, which has low accuracy, but for which performance schedules are distributed in real time. We then examined the accuracy of the estimations. The results showed that the application of an exponential smoothing method is valid.

***Keywords*—**Exponential smoothing method, open data, operation estimation, train schedule.

I. Introduction

In Japan, each railway in suburban areas, where populations continue to grow, is seeking to enhance its transportation capability on commuter lines in order to improve convenience for train users. The methods for this include measures to increase the number of trains, speed up trains, diversify the train types, and increase the length of trains. However, these have at the same time led to a prolonging of the time for which level crossings are closed, and this is becoming a social problem in the form of bottleneck level crossings where the crossing gate bars do not rise. In particular, level crossings that are closed for more than 40 minutes per hour during peak hours in the morning and evening are defined as "unopened" crossings. Currently, there are 600 "unopened" crossings in Japan, and they are having a significant impact of traffic that is concerned with level crossings [1].

Against this background, there are demands for countermeasures regarding prolonged closure of level crossings. As a measure by the government, the Act on Promotion of Railway Crossings has been revised in order to prevent accidents and promote smooth traffic at level crossings, and progress is being made with regard to conversion to crossings with overpasses, structural improvements, and the promotion of maintenance of crossing safety equipment. However, conversion to crossings with overpasses and structural improvements require substantial costs and time and are therefore difficult to carry out, so there is an increasing need for measures that rely on improvement of crossing safety equipment.

The conventional measures taken to optimize crossing closure time through improvement of crossing safety equipment are warning control measures [2], [3], for which factors such as train velocity, stops at stations, and early warning by means of the following train have been taken into account. However, none of these methods is able to keep up with train operation patterns completely, and in reality, achieving adequate effectiveness is difficult [4]. Therefore, the purpose of our research is to estimate the time until it will next be possible to cross a level crossing and provide the information to level crossing users, in order to prevent level crossing users from forcibly entering level crossings, before they attempt to do so. Through this method, level crossing users are provided with information to tell them when the level crossing will open, so they can select whether to wait until the gate opens or use an alternative route.

The travel times, stopping times, and intervals between trains necessary to estimate train operation are obtained through a travel simulation for trains. The information required to conduct this in detail includes the signal system (train control, blocking boundary positions, signal indication system, gradients, etc.), train stopping positions (stopping positions at stations and between stations, etc.), train schedules (times, platforms, train types, vehicle operation plans, etc.), vehicle information (acceleration, number of carriages, number of doors, wide-door vehicles, etc.), and operations at stations, etc. (departure inhibition, number of platform staff, etc.) In addition, circumstances such as direct operation with other lines, transfers to other means of transport, time periods, the operation direction, the weather, the day of the week, and the season need to be ascertained. For example, to estimate train stopping times, a method is used whereby the number of boarding, and alighting passengers is estimated by simulating passenger flow, and then the time required for boarding and alighting is estimated using that number [5]. When a study is carried out using a several such detailed estimation models, there are issues such as requiring a large amount of calculation time and inputting of information to perform the estimation, and requiring additional equipment in order to introduce the results into existing railways. Against this background, the government in Japan has been promoting a shift to open data in order to promote the use of public data, and is carrying out a trial publication of real-time train information [6]. These open data can generally be obtained from existing devices. However, the accuracy of the information is low from the viewpoint of

Taiyo Matsumura and Takashi Ono are with the Department of Electrical Engineering, College of Science & Technology, Nihon University, Tokyo, 101-8308 Japan (phone: +81-3-3259-0761; fax: +81-3-3293-8265; e-mail: matsumura.taiyou@nihon-u.ac.jp, ono.takashi30@nihon-u.ac.jp).
Kuninori Takahashi is with the Graduate School of Science and Technology, Nihon University, Tokyo, 101-8308 Japan (phone: +81-3-3259-0784; fax: +81-3-3259-0784; e-mail: takakuni1162.keisoku@gmail.com).

crime prevention. Therefore, against a background in which the obtainable information has limitations, with regard to the estimation accuracy when operation estimation was performed, we carried out respective examinations using an average value method, a variable response smoothing method, and an exponential smoothing method.

In this paper, we first present a method for creating operation performance on the basis of information that can be obtained from open data. Next, we present methods for estimating the departure time from a station on the basis of the operation performance. We then propose estimation methods that use an average value method, a variable response smoothing method, and an exponential smoothing method, and present the comparison results. At the end, we have performed and evaluated estimations of the operation of delayed trains using each of the methods, on the basis of actual data. From these results, we have shown that the estimation method using an exponential smoothing method is effective, and have clarified that train operation estimation that uses open data can be performed.

II. Creation of Operation Performance

The railways studied are actual lines for which statistics are kept by the Ministry of Land, Infrastructure, Transport and Tourism in the form of congestion rates for main sections [7]. The range studied is between stations 1 to 17 in double-track sections, the direction is one-way from station 1 to station 17, and the train type is local trains only.

The obtainable open data include information such as times, train positions, train numbers, and direction of travel, and they are updated every minute. Train positions are obtained in the form of stations or station intervals that include station names. Train numbers are a unique value in a single day, and starting points, terminating points, and planned departure times for each station can be obtained by referring to basic schedules that can be obtained from time tables. On the basis of the obtained open data, it is possible to tell which trains were in which sections at a certain time. The definition of "when the train was in a certain station" is from when the train starts to enter, or has finished entering, a range including the station that is defined separately from the actual position of the station, until when the train starts to leave, or has finished leaving. Therefore, the situations and times with regard to a train's entering, stopping at and leaving a station cannot be distinguished between. Also, if the travel time between stations is under 1 minute, information about traveling between stations cannot be obtained, since the update interval is 1 minute. Therefore, we decided at the beginning to create an operation performance on the basis of open data. For the departure time, we used the first time that can be obtained when the position of the train is between the current station and the next station, or the time at which it had transitioned from the current station to the next station. Here, we define the time from the departure time at the current station until the departure time at the next station as the dep-dep time. The operation performance includes the departure time of each train from each station and the dep-dep time between stations, in addition to the content in the open data.

Fig. 1 shows the result of creating operation performance for six trains that departed from station 1 between 9:53 and 10:10, on the basis of the open data for Monday, November 7, 2016. The plots show the departure times at each station. From this result, it can be seen that the operation performance of the second train is almost the same as the first train, and that on-time operation is broadly being performed. With regard to the third train, a delay arose at station 5, and after that, it departed station 17 without recovering from the delay. The fourth train was delayed from station 4 as a result of the third train's delay. Similarly, the fifth train was delayed from station 3. The sixth train was not affected by the delay, because there was some leeway in the basic schedule for the departure time from station 1, and it is conceivable that the delay was recovered from at this time.

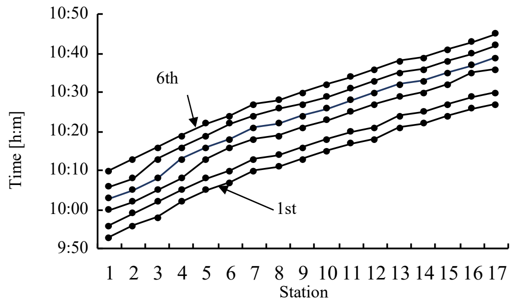


Fig. 1 Created operation performance (Monday, November 7, 2016)

III. Operation Estimation Method

We considered it to be the case that a delay of a group of trains on a given day occurs because the dep-dep times grow longer in a chain-like fashion, starting from the departure time of the train that is the cause of the delay. Therefore, we studied the relationship between the departure times and dep-dep times of the preceding trains.

Fig. 2 shows, using black circles, the dep-dep times from station 1 to station 2 for the preceding trains at station 2, created from the operation performance of a train that departed from station 1 after 7:50 and had departed from station 2 by 8:10, on Monday, November 20, 2016. Since the open data can be obtained per minute, the dep-dep times are in minutes. In this way, we examined methods for calculating the departure times for a given station, by estimating the dep-dep times (white circles) of trains that depart a given station, on the basis of the dep-dep time from the previous station until the given station for each train that had run up until the present at the given station, and then adding the departure time for the previous station. For the estimation method for the dep-dep times, three kinds of methods were examined: an average value method, a variable response smoothing method, and an exponential smoothing method. This was because operation performance follows basic schedules during on-time operation, and delays due to station congestion have a chronological tendency.

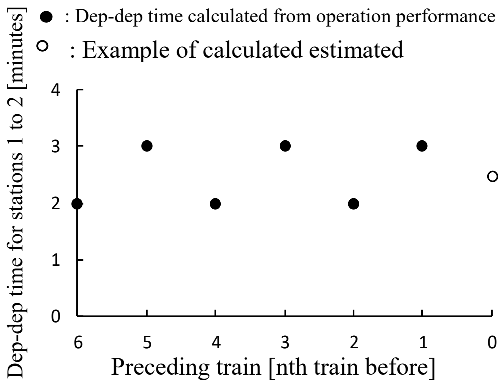


Fig. 2 Example of estimation for dep-dep time for station 2

First, for the average value method, we calculated an estimated value for the dep-dep time using a moving average for the dep-dep time of each train that had run up until the present, for each station. Here, if we assume that the dep-dep time of the *n*th train before is T_n, the estimated value for the dep-dep time, S_0, can be calculated from (1):

$$S_0 = \frac{\sum_1^n T_n}{n} \tag{1}$$

Next, for the variable response smoothing method, we obtained the estimated value for dep-dep time, S_0, from

$$S_{n-1} = \gamma_n T_n + (1-\gamma_n) S_n \tag{2}$$

$$\gamma_n = \left|\frac{\delta_n}{\Delta_n}\right|$$
$$\delta_n = \frac{\sum_{m=1}^{k}(T_{n+m} - S_{n+m})}{k} \tag{3}$$
$$\Delta_n = \frac{\sum_{m=1}^{k}|T_{n+m} - S_{n+m}|}{k}$$

repeating the calculation from *n* to 1, *n* times. Here, S_n is the estimated value for the dep-dep of the *n*th train before, γ_n is a smoothing variable for the *n*th train before, and *k* is the initial value that is required to obtain γ_n. *k* is 4, taking into account the occurrence of a new delay and the number of trains affected by it. For the initial values S_n to S_{n-k+1}, we used vales that were 1.001 times each of T_n to T_{n-k+1}, so that they would each be close to the estimation targets T_n to T_{n-k+1} and (3) would not diverge. Also, γ_n was calculated for each station.

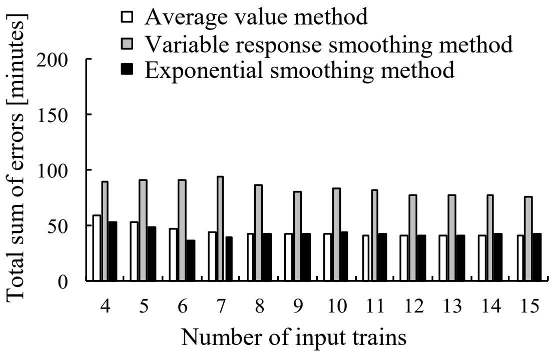
For the exponential smoothing method, we calculated values for each station, whereby the root mean square of the estimation error calculated from (2) was smallest when γ_n in (2) was set to range from 0.1 to 0.9 in steps of 0.1. We decided on this value for the estimated value for dep-dep time.

We performed the estimation on the basis of operation performance when the number of trains (hereinafter, the number of input trains) that can be obtained at each station before the estimation target train departs from station 1 is set to 4. The flow for the estimation is as follows: [1] On the basis of the operation performance of the trains from the 1st train before to the 4th train before, create a dep-dep time for preceding trains at each of the stations 2 to 17. [2] Calculate the estimated value for the dep-dep time for stations 16 to 17 for preceding trains at station 17. [3] From the operation performance, extract the departure time at which the train that will depart from station 17 next departed from the station 16. [4] To the operation performance, add the value obtained by adding [2] to [3], as the departure time from station 17, and also [2] as the dep-dep time for station 17. [5] Repeat [2] to [4] for stations 16 to 2, in the same way. [6] Perform [2] to [5] for the trains that will depart from each station from now onward. The departure time at which the estimation target train departs from the station is extracted from the basic schedules. In the methods above, we changed the number of input trains from 5 to 15, then performed the estimation in the same way.

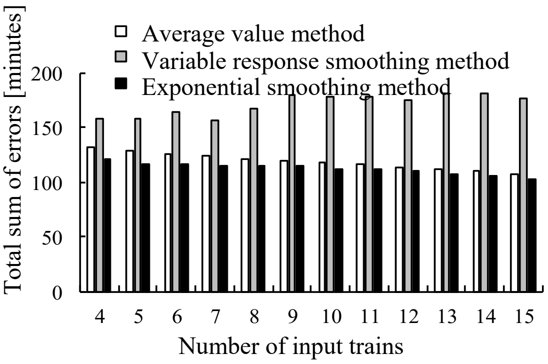
IV. Comparison and Examination of Estimation Methods

In Fig. 3, each train that departs from station 1 from 7 a.m. until 9 a.m. (a rush hour on account of commuting to work and school, etc.) was treated as an estimation target, and the number of input trains was changed from 4 to 15. The figure shows the results of calculating the total sum of the absolute values of the departure time errors at each station for each estimation target. (a) is for Monday, November 14, 2016, and (b) is for Monday, November 21, 2016. From (a) in the same figure, the total sum of errors with the variable response smoothing method is large for all numbers of input trains. Also, with the average value method and variable response smoothing method, the values show a tendency to decrease when the number of input trains increases, and for the exponential smoothing method, the lowest value when it is six trains. In (b) in the same figure, the total sum of errors decreases in the order variable response smoothing method, then average value method, then exponential smoothing method, for all numbers of input trains. Also, with the variable response smoothing method, the values show a tendency to increase when the number of input trains increases, whereas with the average value method and exponential smoothing method, they decrease. Compared to the values in (a) in the same figure, the total sum of errors is generally larger in (b). This is thought to be due to the fact that there were many delayed trains. In this manner, with regard to the total sum of errors, the differences and tendencies with each estimation method vary depending on the congestion situation on the day in question. Therefore, with regard to comparison of the estimation methods and finding the optimal value for the number of input trains, the total sum of errors was calculated using multiple days, and a comprehensive examination conducted.

Fig. 4 shows the results of adding the total sums of errors for five business days. With the variable response smoothing method, the total sum of errors is 1.3 to 1.6 times larger than with the others, and with the exponential smoothing method, it is from 1 to 24 minutes smaller than with the average value method. From the above results, the exponential smoothing method can be considered to be the optimal estimation method.



(a) Monday, November 14,



(b) Monday, November 21, 2016

Fig. 3 Total sum of errors for each departure time for stations 1 to 17

Next, we examined the optimal value for the number of input trains for each estimation method. The total sum and tendency of the errors varies depending on the day-by-day congestion situation. Therefore, for each estimation method, we calculated the percentage accounted for by the total sum of errors for each number of input trains in relation to all of the numbers of input trains, averaged them over multiple days, and established the minimum value as the optimal value for the number of input trains.

Fig. 5 shows the results of calculating the average value for the percentage of the total sum of errors for five business days. The number of input trains for which the total sum of errors is smallest was 12 for the average value method and variable response smoothing method, and 8 for the exponential smoothing method.

V. Operation Estimation

We carried out operation estimation for each of the following trains: one whose departure time from station 17 was delayed by 3 minutes compared with the operation performance schedule after departing from station 1 at 8:20 on Monday, December 5, 2016, and one whose departure time from station 17 was delayed by 5 minutes after departing from station 1 at 8:26 on the same day. Fig. 6 shows the results of the operation performance created for the operation estimation. For the train whose departure time from station 17 was delayed by 3 minutes and the train that was delayed by 5 minutes, an operation performance was obtained that was similar to the other trains with almost no delay.

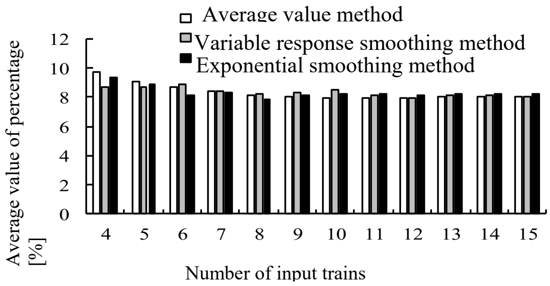


Fig. 4 Results of adding the total sums of errors for five business days

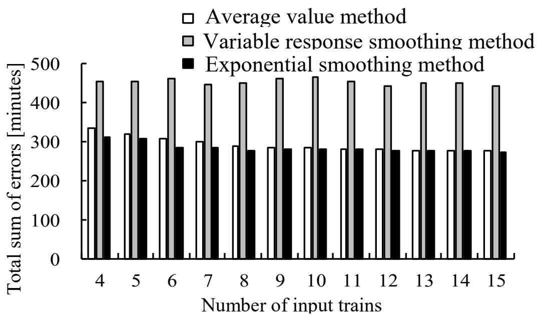


Fig. 5 Average value of the percentage of the total sum of errors for five business days

Fig. 7 is the operation estimation results for each estimation target train in Fig. 6 when each estimation method was applied on the basis of the optimal value for the number of input trains calculated in the previous chapter. Fig. 7 (a) is the 3-minute delay, and (b) is the 5-minute delay. From (a) in the same figure, it can be seen that in the case of the 3-minute delay, the results for each estimation method are almost the same. Also, at station 13 to 15, around 2-minute error from operation performance occurred in all estimation methods, but errors in the departure time at station 17 is within 17 seconds. This is thought to be due to the fact that the traveling of a train is a continuous phenomenon, and the effects from variations in the dep-dep times calculated from open data that is obtained per minute are small. From (b) in the same figure, in the case of the 5-minute delay, the error is larger with the variable response smoothing method than with the other methods. This is significant in the operation performance after station 13, but this is because the train in question was delayed at station 13 in Fig. 6. The errors in departure time for station 17 were, in the order average value method, then variable response smoothing method, then exponential smoothing method: -16.8 seconds, -8.4 seconds, and -5.4 seconds in (a) in the same figure, and -2.5 minutes, -3.7 minutes and -2.3 minutes in (b) in the same figure. All errors are smallest with the exponential smoothing method, and in particular, in (a) in the same figure, a maximum of a 67.9% reduction effect was obtained. From these results, it could be seen that the estimation method using the exponential smoothing method is effective.

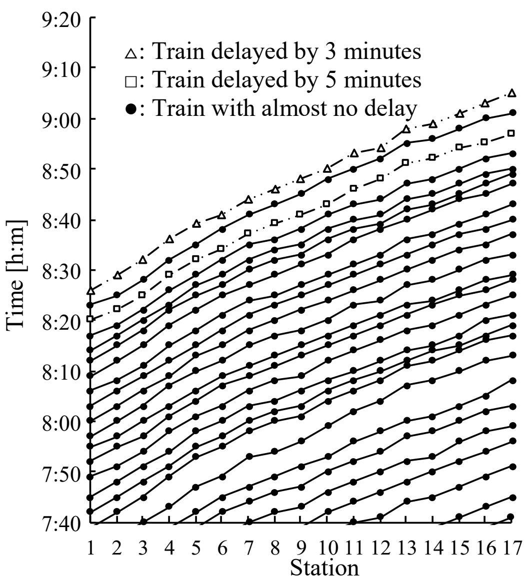


Fig. 6 Operation performance created for operation estimation

VI. Conclusion

In this paper, we proposed methods for estimating operation on the basis of open data that includes real-time train information. Against a background in which the obtainable information has limitations, we performed and then compared estimations of train operation using an average value method, a variable response smoothing method, and an exponential smoothing method. These results have led to the following conclusions.

1) We have proposed methods for obtaining operation performance from open data.
2) We have proposed method for estimating operation on the basis of dep-dep times of preceding trains.
3) We have proposed an average value method, a variable response smoothing method, and an exponential smoothing method as methods for estimating dep-dep times.
4) Operation estimation was performed for each train that departed station 1 from 7 a.m. to 9 a.m. (rush hour) on the basis of operation performance for five days, and the results were that the exponential smoothing method had the smallest errors.
5) The number of input trains for which the total sum of errors is the smallest is 12 with the average value method and variable response smoothing method, and 8 with the exponential smoothing method.
6) Operation estimation was performed using each method with regard to trains delayed by 3 minutes and 5 minutes. The results indicated that the errors in departure time at station 17 were smallest with the exponential smoothing method in both cases, and a maximum of a 67.9% reduction effect was obtained compared with the other estimation methods.

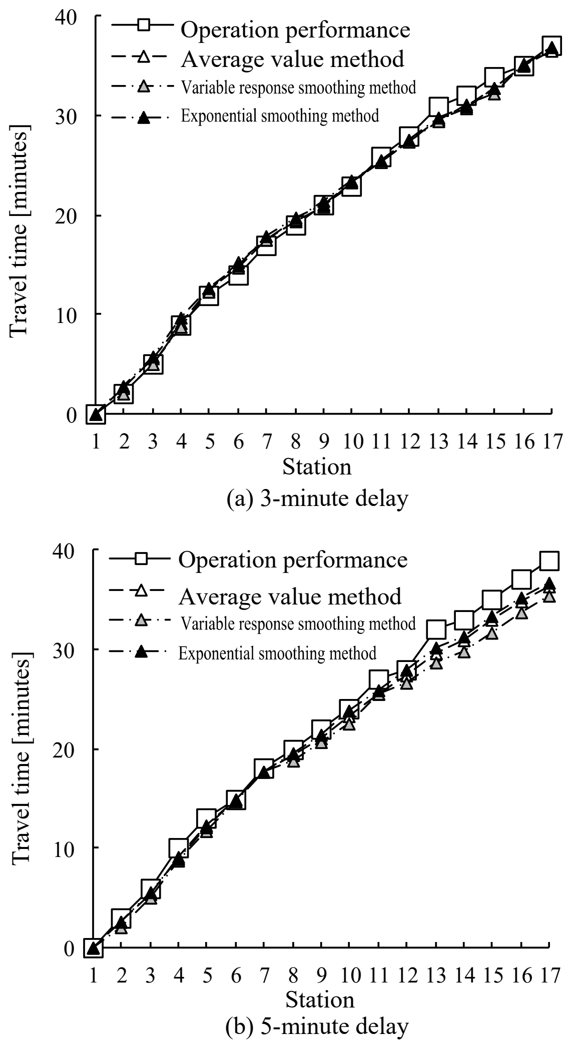


Fig. 7 Operation estimation results

From the above, we have shown that among the methods that we have proposed in this research, the exponential smoothing method is the optimal estimation method for trains using open data.

References

[1] Facilities Division of Railway Bureau, "Comprehensive inspection results regarding level crossing traffic accidents", Ministry of Land, Infrastructure, Transport and Tourism, 2007, p.1 and Attachment ＜ Definition of Terms＞.
[2] M. Miyachi, and Y. Hirao, "The constant warning time control installation used as countermeasures for extraordinarily long time closed level-crossings on urban railways", RTRI Report, Vol.2, No.2, pp.29-36, 1988.
[3] T. Kumagai, Y. Hirao, and Y. Hasegawa, "Level crossing control by

radio", RTRI Report, Vol.4, No.11, pp.42-49, 1990.
[4] T. Matsumura, and T. Ono, "Research into Warning Time Optimum Control for Level Crossing - Train Location Detector using Three-Inductive Wires –", IEEJ Transactions on Fundamentals and Materials, Vol.132, No.10, 2012, pp.893-900.
[5] K. Ohashi, T. Mori, and T. Koseki, "Modeling the Change in Passenger Flow Depending on Train Congestion in Train Rescheduling", The Institute of Electrical Engineers of Japan, Vol.135, No.4, 2015, pp.438-443.
[6] K. Murao, Initiatives on "Practical use of open data" of Tokyo Metro, Ministry of Internal Affairs and Communications, 2015, pp.3-47.
[7] Urban Railway Policy Division of Railway Bureau, "Congestion rate data", Ministry of Land, Infrastructure, Transport and Tourism, 2015, p.3.

Taiyo Matsumura was born on August 11, 1980. He received Dr. Eng. Degrees in the field of electrical from Nihon University, in 2010. He is working as Assistant Professor at Department of Electrical Engineering, College of Science & Technology, Nihon University. He is member of IEEJ, SICE and IEIJ. His major field of study is measurement engineering.

Kuninori Takahashi was born on April 25, 1994. He received Bachelor Eng. Degrees in the field of electrical from Nihon University, in 2017. He is a master-course student at Graduate School of Science and Technology, Nihon University. His research interests are in the field of signaling for railway.

Takashi Ono was born on March 15, 1948. He received Dr. Eng. Degrees in the field of electrical from Nihon University, in 1989. He is working as Professor at Department of Electrical Engineering, College of Science & Technology, Nihon University. He is member of IEEJ, SICE and IEIJ. His major field of study is measurement engineering.