Journal of Research and Applications in Mechanical Engineering
ISSN: 2229-2152 (Print); 2697-424x (Online)
(2024) Vol. 12, No. 1, Paper No. JRAME-24-12-001
[DOI: 10.14456/jrame.2024.1]

Research Article

Research on Communication Analysis between Vehicles and Pedestrians at Unsignalized Crosswalks Using Online Video

G. Bamba
D. Misaki*
Department of Mechanical Systems Engineering, School of Engineering, Kogakuin University Shinjuku Campus, Tokyo 1638677, Japan

Received 20 January 2023
Revised 14 May 2023
Accepted 20 May 2023

Abstract:
In recent years, considerable attention has been paid to research on communication in self-driving vehicles. However, few studies have quantitatively evaluated and clarified the relationship between modern vehicles and pedestrians. Additionally, communication methods and gestures are perceived differently in different countries and regions. In the past, quantitative research on communication that differed across countries and regions often required research teams to travel across countries to conduct research, which entailed enormous costs and a heavy burden. In this study, we propose an online video-based pedestrian behavior analysis and compare pedestrian communication in Japan and U.S. based on a questionnaire survey. Results with the proposed algorithm reveal that Japanese pedestrians tend to use eye contact as a form of communication at unsignalized crosswalks, whereas American pedestrians tend to use hand-raising movements. Moreover, the gesture choices of pedestrians in both countries vary based on the strength of their authority on the road.

Keywords: *Autonomous vehicle, Communication, Online video, Unsignalized, Crosswalks, Motion analysis*

1. Introduction

As research and development in automated driving technology continues to advance rapidly, expectations for the spread of driverless, fully automated vehicles are growing. In addition to "advanced control technology" and "technology that recognizes the vehicle's position and surrounding information," "technology that makes decisions on behalf of humans" is required for the development of fully automated vehicles. The "technology that makes decisions on behalf of humans" is required make both simple decisions, such as going straight, turning, and stopping, and complex decisions, such as cooperative behavior with pedestrians and other road participants. However, no clear rules regarding the type of behavior or gestures pedestrians should use to coordinate among traffic participants at uncontrolled pedestrian crossings without signals have been established, even in the current situation where fully automated vehicles are not widely used. For automated vehicles to function safely and efficiently, efforts must be made to understand the intentions of pedestrians' behaviors and gestures, and to reduce uncertainty.

Therefore, this study defines gestures that encourage this behavior and interaction on the part of pedestrians as "communication" [1-3] between automobiles and pedestrians. Zhuang and Wu attempted to empirically understand the relationship between automobile deceleration behavior and pedestrian gestures in Beijing, China [4]. In addition, studies that assume communication when driverless self-driving cars are widely used [5-7], such as using the Ghost Driver to obtain pedestrian reactions, have been conducted. Research is also being conducted to explore the potential

* Corresponding author: D. Misaki
E-mail address: misaki@cc.kogakuin.ac.jp

of applying gaze-based intentionality to autonomous vehicles. This is aimed at reducing dangerous road crossings by pedestrians [8-10], and enabling autonomous vehicles to understand the intent behind pedestrian gestures [11]. By contrast, non-verbal communication varies across different countries and regions [12].

Furthermore, this suggests that in the study of self-driving cars, research is required on the differences in communication between pedestrians in different countries and regions for driverless, fully self-driving cars to be widely used worldwide [13].

However, the current situation of restricted mobility owing to the outbreak of the coronavirus disease limits the ability of research teams to collect local sample data; additionally, the cost of experiments involving comparing communication in different countries is high.

One way to solve these problems is to conduct research using online videos. A method for analyzing human behavior from online videos, rather than videos taken by a research team [14], has already been developed. However, to the best of our knowledge, few studies have quantitatively compared communication between countries or regions by conducting pedestrian behavior analyses using online videos.

Therefore, the purpose of this study was to compare pedestrian communication in Japan and the USA based on a proposed pedestrian behavior analysis using online videos and a questionnaire survey. The proposed method enables the collection of a wide range of sample data worldwide when determining the role of communication, which does not fit within the laws between traffic participants. Additionally, it is expected to extract communication that includes the cultural characteristics of each country or region. This makes it possible to collect a wide range of sample data worldwide when determining the role of communication, which does not fit within the laws between traffic participants.

2. Communication between automobiles and pedestrians

Figure 1 shows a series of communication processes between vehicles and pedestrians. Pedestrians communicate intentionally or unintentionally with motor vehicle drivers through their movements and postures when crossing pedestrian crossings. Pedestrians intentionally or unintentionally make eye contact, raise their hands, or bail motor vehicle drivers when crossing pedestrian crossings. In addition, car drivers make decisions, such as stopping or passing through, based on the pedestrian's gestures and postures; additionally, car drivers express their intentions to pedestrians. The following definitions and processes were based on previous studies [15].

- Pre-communication: Communication that occurs before cooperative behavior, as typified by eye contact.
- Cooperative behavior: Behaviors such as slowing down and pausing between car drivers and pedestrian’s/bicycle users (limited to cooperative behavior on the car side in this study).
- Post-communication: Communication occurs as a result of cooperative behavior, such as bowing and thanking.

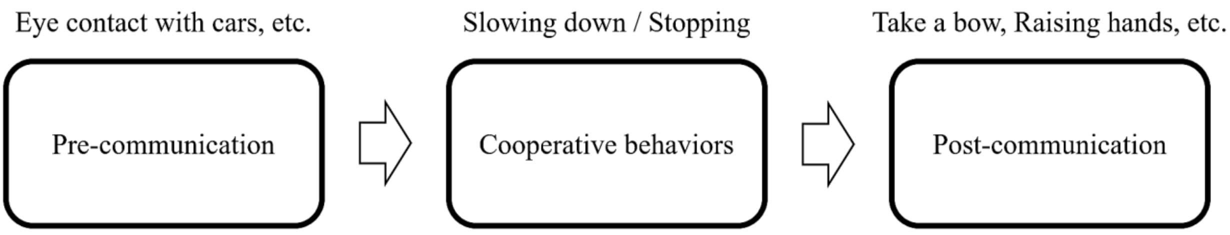

Fig. 1. Communication and cooperative behavior flow on the road.

3. Pedestrian motion analysis using online video

3.1 Collecting data from YouTube

In this study, we collected data from YouTube, which has the most users and videos in the world and has been used for research purposes.

Figure 2 shows an example of YouTube video that was used in this study. We collected videos using keywords, such as "Unsignalized crosswalks" and "Crosswalk pedestrians, which are associated with pedestrians at unsignalized crosswalks on YouTube. "Unsignalized crosswalks" and "Crosswalk pedestrian" on YouTube. Videos of unsignalized crosswalks were used because we believe that they provide a very meaningful phenomenon for extracting the relationship between cars and pedestrians. In addition, in this study, only pedestrians were included in the analysis, whereas bicyclists, skateboarders, and kickboard users were excluded.

The following four conditions are used in this study. We sampled two videos on YouTube based on the following four criteria [16, 17].

- The angle of view of the online video must be from the pedestrian side of the crosswalk.
- The video must be of an unsignalized pedestrian crossing that is not controlled by a traffic signal.
- The scene must show cooperative behavior, such as stopping or slowing down vehicles when pedestrians cross the street.
- Pedestrians crossing the street must cooperate with other pedestrians by stopping or slowing down.

In addition, we will describe the YouTube videos used in this experiment. Both videos depict traffic conditions at two-lane, unsignalized crosswalks in each direction. The featured roadways are state highways and major arterials in Los Angeles, California, indicating that they are relatively busy streets. The number of pedestrians traversing the unsignalized crosswalks is approximately 2 to 3 individuals per minute in both videos.

Fig. 2. Examples of videos on YouTube that satisfy the experimental conditions.

3.2 Pedestrian motion analysis

In this study, videos of pedestrians at actual unsignalized pedestrian crossings were recorded and information on their posture and movements were acquired to clarify the communication between cars and pedestrians. We used Open Pose [18], to acquire such pedestrian posture and motion information. We proposed a method to identify the posture and movements of a pedestrian by obtaining seven points of information from an Open Pose: the head, neck, waist, hands, and feet. However, OpenPose's analysis accuracy drops when the analyzed part overlaps an obstacle, such as a vehicle. Therefore, to obtain more accurate joint information from the analysis results, we excluded joint point frames with a confidence level of less than 70 % [19], which indicates analysis accuracy. Figure 3 shows the analysis procedure and pre-processing for the motion analysis of pedestrian videos. When performing the pedestrian motion analysis, the position of the subject's neck was corrected such that it was at the origin. When the neck position was considered as the origin, the neck coordinates were $\boldsymbol{Xneck}$ ($\boldsymbol{xneck}$, $\boldsymbol{yneck}$), the posture information $\boldsymbol{Xi}$ ($\boldsymbol{xi}$, $\boldsymbol{yi}$), and position correction was performed. The posture information $\boldsymbol{X}$ ($\boldsymbol{x}$, $\boldsymbol{y}$) is shown in equation (1) as follows.

$$X(x, y) = Xi - Xneck = (xi - xneck, yi - yneck) \tag{1}$$

In addition, a normalization process [20] was performed using the following equation such that the length from the neck to the waist, which is relatively stable among the various regions, was 100.

$$N = \frac{X(x,y)}{(y\_hip - y\_neck)} \times 100 \tag{2}$$

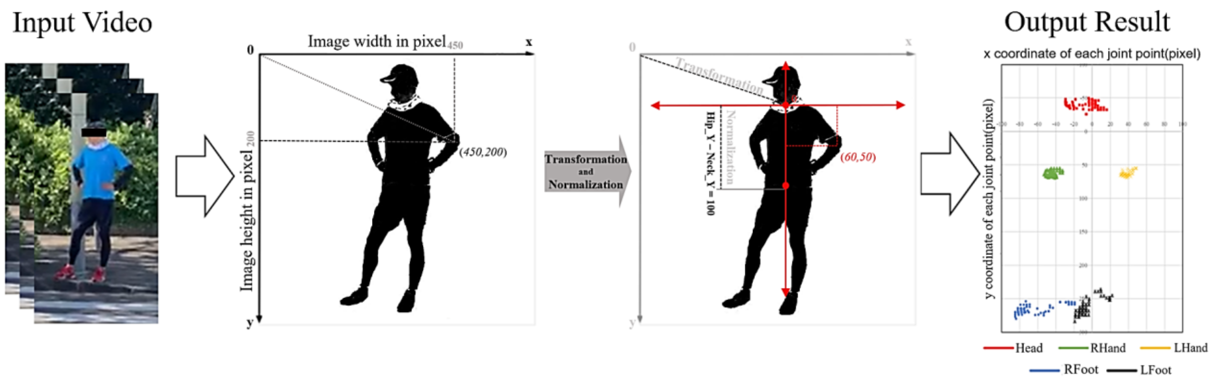

Fig. 3. Analysis procedure and preprocessing.

4. Experiment

We conducted a comparative study with 10 participants, five from Tokyo, Japan (Japan A-E), and five from California, USA (USA F-J), filmed at two locations. To analyze the differences in communication between cars and pedestrians at the two sites and the regional differences, we used OpenPose analysis for pre-communication. We analyzed the pedestrians' movements at five points: head, both hands, and both feet. Preliminary experiments revealed that the standard deviation of Head_x varied from 9.0 to 15.0 when indicating the movement of the pedestrian's head (left and right). As is well-known, the standard deviation of RHand_y or LHand_y, which indicates the Y-axis movement of the hand, varies from 20.0 to 40.0 owing to the pedestrian's hand-raising motion. Communication methods were compared based on the size of the standard deviation of the head and both hands.

Figure 4 shows the questionnaire conducted, in addition to image analysis. The questionnaire consisted of a choice-type question: "What action would you take as a pedestrian when you want to cross a pedestrian crossing and a car is coming near you?" In addition, the subjects were also asked regarding the reason for their choice. The subjects were residents of Japan and the USA.

Fig. 4. Questionnaire on pedestrians at unsignalized crosswalks.

5. Experimental result

The analysis results for the USA, CA, and Japan, Tokyo, which were collected on YouTube and filmed, are summarized in Table 1 and 2. To quantitatively compare the head and hand movements, comparisons were made using the standard deviation to show the movements of each pedestrian, making eye contact with the vehicles, and raising/holding up their hands, respectively. First, we compared the movements in which pedestrians moved their heads to the left and right to make eye contact with vehicles. Of the Japanese pedestrian samples (Japan A-E), 4/5 participants made eye contact with the vehicles (Head_x standard deviation is approximately 9.0-15.0), and 3/5 participants raised or held their hands up (standard deviation of RHand_y or LHand_y is approximately 20.0-40.0). In the American pedestrian sample (USA F-J), 2/5 participants made eye contact with the vehicles (Head_x standard deviation is approximately 9.0-15.0), and 5/5 participants raised or held up their hands (standard deviation of RHand_y or LHand_y is approximately 20.0-40.0).

Table 1: Japan A-E uncorrelated pedestrian sample.

Subject	Gender	Flame	SD									
			Head_x	Head_y	RHead_x	RHead_y	LHead_x	LHead_y	RFoot_x	RFoot_y	LFoot_x	LFoot_y
Japan A	Male	103	12.30	6.09	10.60	25.20	6.07	5.22	7.91	8.09	12.90	8.69
Japan B	Female	69	8.13	5.17	6.73	4.86	8.40	5.72	19.28	7.11	12.44	8.54
Japan C	Male	48	12.74	3.83	9.53	22.88	7.96	6.57	29.31	14.49	31.59	13.92
Japan D	Male	55	9.88	5.11	20.82	45.88	5.83	3.73	32.35	10.09	28.93	11.79
Japan E	Female	66	14.60	3.10	12.09	6.61	8.11	6.96	18.60	12.08	19.15	11.79

Table 2: USA F-J uncorrelated pedestrian sample.

Subject	Gender	Flame	SD									
			Head_x	Head_y	RHead_x	RHead_y	LHead_x	LHead_y	RFoot_x	RFoot_y	LFoot_x	LFoot_y
USA F	Male	235	3.87	4.41	6.55	4.47	21.68	33.58	12.79	10.45	12.79	9.28
USA G	Male	89	3.36	3.19	13.23	5.60	9.11	33.05	30.19	10.40	30.58	15.56
USA H	Male	218	9.84	6.32	7.09	4.53	9.01	30.80	10.24	9.62	14.85	8.32
USA I	Female	110	13.08	3.52	10.06	6.18	35.73	35.87	6.86	18.32	5.67	14.29
USA J	Male	80	5.54	2.78	19.65	36.35	7.59	6.29	8.28	8.98	9.29	6.73

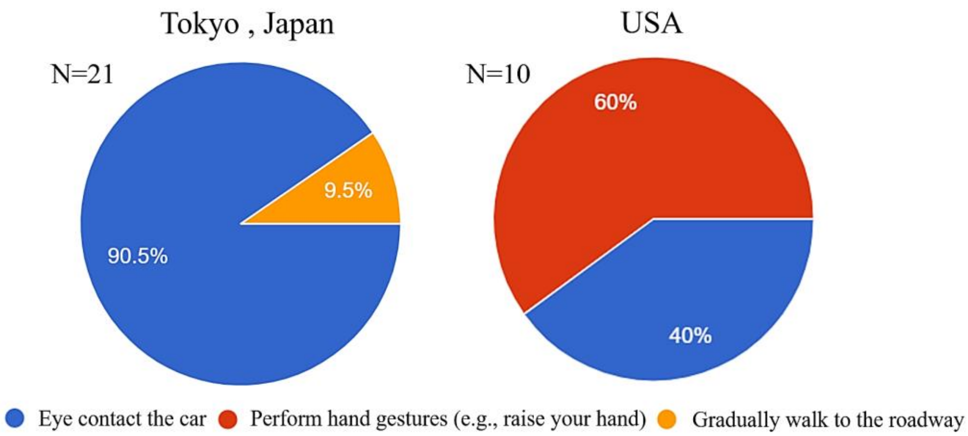

Fig. 5. "What action would you take as a pedestrian if you wanted to cross an unsignalized crosswalk with a car nearby?"

Figure 5 shows the results of a survey conducted on men and women aged between 17 and 58 living in Japan (N = 21) and the USA (N = 10). Of the Japanese subjects, 90.5 % chose "Eye contact with the car." Of the American subjects, 60 % chose "Perform hand gestures" and 40 % chose "Eye contact with the car. The questionnaire revealed that American subjects tended to select "Perform hand gestures" more compared with the Japanese subjects. When the subjects were asked, "Why did you choose that action?," those in Japan who chose eye contact were more likely to choose eye contact "To confirm the driver's intention to stop" and "Indirectly to confirm the driver's intention to stop." Those living in the USA chose hand gestures, such as raising/holding up their hands answered, "Because I want to make sure that the driver sees me and reads my body language." For those who chose eye contact, responses included "Make sure they knew you were there at the crosswalk.".

6. Discussion

The experimental results confirmed that pedestrian pre-communication could be analyzed using online video, and the differences in pre-communication methods between the two locations could be observed quantitatively through behavior analysis. However, the background and context of why the gesture was selected could not be confirmed based on only the behavioral analysis results. Therefore, by examining the questionnaire results in conjunction with pedestrian analysis using online video, we could observe the meaning of each pedestrian in the two countries attached to the gesture.

One fact from the analysis results was that Japan used eye contact as a form of pre-communication. The questionnaire also revealed that respondents from the USA tended to use hand movements, such as raising/holding their hands, in the form of pre-communication. The differences in pre-communication between the two countries were similar to the results of the questionnaire. One factor contributing to the difference in pre-communication methods between the two countries may be related to the differences in the authority and position of motorists and pedestrians.

In the questionnaire, we asked, "Why did you choose that action?" The responses of those living in Japan were "To confirm whether the car is willing to stop" and "To indirectly convey my intention to the driver." This suggests that Japanese pedestrians choose gestures for pre-communication, based on the assumption that the driver has authority over the road.

By contrast, the questionnaire from residents in the USA yielded responses, such as "Because I want to make sure that the driver sees me and reads my body language" and "Make sure they know you are there at the crosswalk." All of these responses indicate that pedestrians are making their presence known on the road. This suggests that American pedestrians choose their gestures based on the assumption that they have authority on the road. Thus, based on the results of the questionnaire and analysis, we believe that the difference in pre-communication methods between Japan and the USA can be attributed to the concept of the strength of authority between pedestrians and cars at unsignalized pedestrian crossings.

In addition, this study used online videos taken from common perspectives and conditions. Therefore, the applicability of the results obtained in this study is limited. However, the algorithm for analyzing communication from pedestrian motion analysis was appropriate; therefore, future work is required to expand the range of applicability of the online videos used for analysis by increasing the conditions and number of samples of online videos.

7. Conclusion

We analyzed the differences in communication between the residents in Japan and the USA as a sample survey of videos on video-posting sites. A questionnaire survey was conducted, which allowed us to examine the relationship between motorists and pedestrians at unsignalized pedestrian crossings. In addition, future work includes examining the extent to which videos posted on video-posting sites are representative samples of each country. The sample size was small for this study. Additional verification of whether the same results of a study can be obtained when the number of pedestrians analyzed and surveyed is increased to 1,000 or 10,000, will also be future work.

References

[1] Rasouli A, Tsotsos JK. Autonomous vehicles that interact with pedestrians: a survey of theory and practice. IEEE Trans Intell Transp Syst. 2020;21(3):900-918.

[2] Domeyer JE, Lee JD, Toyoda H. Vehicle automation–other road user communication and coordination: theory and mechanisms. IEEE Access. 2020;8:19860-19872.

[3] Risto M, Emmenegger C, Vinkhuyzen E, Cefkin M, Hollan J. Human-vehicle interfaces: the power of vehicle movement gestures in human road user coordination. The Ninth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design; 2017 Jun 26-29; Iowa, United State. p. 186-192.

[4] Zhuang X, Wu C. Pedestrian gestures increase driver yielding at uncontrolled mid-block road crossings. Accid Anal Prev. 2014;70:235-244.

[5] Rothenbücher D, Li J, Sirkin D, Mok B, Ju W. Ghost driver: a field study investigating the interaction between pedestrians and driverless vehicles. 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN); 2016 Aug 26-31; New York, USA. USA: IEEE; 2016. p. 795-802.

[6] Currano R, Park SY, Domingo L, Garcia-Mancilla J, Santana-Mancilla PC, Gonzalez VM, et al. ¡Vamos! observations of pedestrian interactions with driverless cars in Mexico. Proceedings of the 10th International conference on automotive user interfaces and interactive vehicular applications; 2018 Sep 23-25; Toronto, Canada. p. 210-220.

[7] Li J, Currano R, Sirkin D, Goedicke D, Tennent H, Levine A, et al. On-road and online studies to investigate beliefs and behaviors of Netherlands, US and Mexico pedestrians encountering hidden-driver vehicles. 15th ACM/IEEE International Conference on Human-Robot Interaction (HRI); 2020 Mar 23-26; Cambridge, United Kingdom. USA: IEEE; 2020. p. 141-149.

[8] Gui X, Toda K, Seo SH, Eckert FM, Chang CM, Chen XA, et al. A field study on pedestrians' thoughts toward a car with gazing eyes. Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems (CHI EA '23); 2023 Apr 23-28; Hamburg, Germany. p. 1-7.

[9] Chang CM, Toda K, Gui X, Seo SH, Igarashi T. Can eyes on a car reduce traffic accidents?. Proceedings of the 14th International conference on automotive user interfaces and interactive vehicular applications; 2022 Sep 17-20; Seoul, Republic of Korea. p. 349-359.

[10] Gui X, Toda K, Seo SH, Chang CM, Igarashi T. "I am going this way": Gazing eyes on self-driving car show multiple driving directions. Proceedings of the 14th International Conference on Automotive User Interfaces and Interactive Vehicular Applications; 2022 Sep 17-20; Seoul, Republic of Korea. p. 319-329.

[11] Holzbock A, Kern N, Waldschmidt C, Dietmayer K, Belagiannis V. Gesture recognition with keypoint and radar stream fusion for automated vehicles. Computer Vision–ECCV 2022 Workshops; 2022 Oct 23-27; Tel Aviv, Israel. p. 570-584.

[12] Patterson ML, Iizuka Y, Tubbs ME, Ansel J, Tsutsumi M, Anson J. Passing encounters east and west: comparing Japanese and American pedestrian interactions. J Nonverbal Behav. 2007;31(3):155-166.

[13] Wolf I. The interaction between humans and autonomous agents. In: Maurer M, Gerdes J, Lenz B, Winner H, editors. Autonomous driving. Berlin: Springer; 2016. p. 103-124.

[14] Dong J, Shuai Q, Zhang Y, Liu X, Zhou X, Bao H. Motion capture from internet videos. 16th European Conference on Computer Vision; 2020 Aug 23-28; Virtual platform. p. 210-227.

[15] Taniguchi A, Yoshimura T, Ishida H. A study on generation of cooperative behavior by communication between vehicles and pedestrians/bicycles. J Jpn Soc Civ Eng. Ser. D3 (Infrastructure Planning and Management) 2012;68(5):1115-1122. (In Japanese)

[16] BUSted Los Angeles. BUSted! Crosswalk Sting at Western & Serrano, Koreatown [Internet]. 2019 [cited 2022 Aug 25]. Available from: https://youtu.be/BsKmw_bmi2k.

[17] BUSted Los Angeles. BUSted! Los Angeles - Afternoon West Hollywood Crosswalk Sting - Santa Monica Blvd [Internet]. 2019 [cited 2022 Aug 25]. Available from: https://youtu.be/Q58JaCRKMHo.

[18] Cao Z, Simon T, Wei SE, Sheikh Y. Realtime multi-person 2D pose estimation using part affinity fields. IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2017 Jul 21-26; Honolulu, USA. USA: IEEE; 2017. p. 1302-1310.

[19] Nakai M, Tsunoda Y, Hayashi H, Murakoshi H. Prediction of basketball free throw shooting by openpose. In: Kojima K, Sakamoto M, Mineshima K, Satoh K, editors. New Frontiers in Artificial Intelligence. JSAI-isAI 2018. Cham: Springer; 2018. p. 435-446.

[20] Thomanek R, Rolletschke T, Platte B, Hosel C, Roschke C, Manthey R, et al. Real-time activity detection of human movement in videos via smartphone based on synthetic training data. IEEE Winter Applications of Computer Vision Workshops (WACVW); 2020 Mar 1-5; Snowmass, USA. USA: IEEE; 2020. p. 160-164.