The Use of Visual Cues to Determine the Intent of Cyclists in Traffic

Paul E. Hemeren[1], Mikael Johannesson[1], Mikael Lebram[1],
Fredrik Eriksson[1], Kristoffer Ekman[2] and Peter Veto[1]

Interaction Lab, School of Informatics[1] and Dept. of Cognitive Neuroscience and Philosophy[2]
University of Skövde, 54 128 Skövde, Sweden
Email: paul.hemeren@his.se

***Abstract*— The purpose of this research was to answer the following central questions: 1) How accurate are human observers at predicting the behavior of cyclists as the cyclists approached a crossing? 2) If the accuracy is reliably better than chance, what cues were used to make the predictions? 3) At what distance from the crossing did the most critical cues occur? 4) Can the cues be used in a model that can reliably predict cyclist intent? We present results that show a number of indicators that can be used in to predict the intention of a cyclist, i.e., future actions of a cyclist, e.g., "left turn" or "continue forward" etc.**

Results of empirical studies show that humans are reasonably good at this type of prediction for a majority of the situations studied. However, some situations seem to contain conflicting information. The results also suggested that human prediction of intention is to a large extent relying on a single "strong" indicator, e.g., that the cyclist makes a clear "head movement". Several "weaker" indicators that together could be a strong "combined indicator", or equivalently strong evidence, is likely to be missed or too complex to be handled by humans in real-time. We suggest this line of research can be used to create decision support systems that predict the behavior of cyclists in traffic.

I. Introduction

With the increasing occurrence of cyclists in traffic, there is a greater need for operators of motor vehicles to predict cyclist behavior (as well the behavior of pedestrians). Current active safety systems in automobiles can detect pedestrians and to some extent cyclists. Future active safety systems will likely be more adaptive in the sense of being able to predict the behavior of, for example, cyclists in traffic within a time frame that allows appropriate action to avoid a collision.

A significant part of operating a motor vehicle in traffic is to predict and adapt to the movements of fellow motorists and vulnerable road users. People in general are quite good at predicting the behavior of other people, i.e., reading their intentions. As cognitive systems, humans are predictive rather than merely reactive. The research presented here shows that observers can reliably predict the intentions of cyclists who are approaching a crossing and will either turn left or continue straight on. We present results from experiments that indicate the cues that observers use to make the reliable predictions. We also present a logistic regression model based on the identified salience of the cues that participants reported. The performance of the model achieves a higher level of accuracy than the human observers.

A. *Biological Motion and Social Cues*

Movement patterns play a crucial role when it comes to human prediction of intention [1-5]. Furthermore, recent results show the importance of social cues, e.g., eye contact when motorists and cyclists meet at a crossing, in terms of perception of others' intentions. Such results are typically exploited in different types of artificial systems in order to detect different types of interesting behavior of humans (e.g. [6-8]). These results add to the evidence [9-10] which shows that some biologically determined movements, e.g., eye direction and foot movement attract human attention.

Walker and Brosnan [8] describe the importance of social cues between motorists and cyclists when they meet at an intersection, not least the presence of eye contact. If social cues are missing, motorists seek other relevant indicators that might reveal the intention of the cyclists [8]. They [8] investigated the question of what indicators can be used to read a pedestrian's intentions. The results showed that the indicators associated with body language were necessary for a reliable assessment of the pedestrian's intentions. In this case the body language consisted of leg or head movement and rotation of the body. A reasonable assumption is that a similar reliance on bodily movement and posture hold also for cyclists.

The purpose of this research was to answer the following questions: 1) How accurate are human observers at predicting the behavior of cyclists as the cyclists approached a crossing? 2) If the accuracy is reliably better than chance, what cues were used to make the predictions? 3) At what distance from the crossing did the most critical cues occur? 4) Can the cues be used in a model that can reliably predict cyclist intent? An experiment and a logistic regression analysis were conducted to answer to these questions and present complementary methods for verifying the cues that people use to detect cyclist intent and that those same cues can be used in model to reliably predict cyclist intent.

978-1-4799-3564-2/14/$31.00

II. EXPERIMENT: PREDICTION, ACCURACY AND CUE IDENTIFICATION

This experiment investigated the extent to which observers could reliably predict the behavior of cyclists as the cyclists approached a crossing where they could either turn left or continue straight on and to identify critical cues that observers use. This experiment was carried out in order to establish a proof-of-concept concerning the ability of observers to predict cyclist behavior. As such, the conditions of the experiment were selected to minimize the influence of extraneous variables and to create favorable viewing conditions.

A. Method

1) Participants

Twenty students from the University of Skövde participated in the experiment (17 males), mean age = 21. All participants provided informed consent after receiving information about the experiment and their specific task. The recording of stimuli and participation in the experiment were carried out in accordance to Swedish law and the World Medical Association Declaration of Helsinki.

2) Stimulus material: Video recordings

Forty-two naturalistic video sequences of cyclists approaching a crossing were chosen from recordings taken near the university. The cyclists were not aware they were being recorded (naturalistic observation recordings). In twenty-one of the sequences, the cyclists made a left turn, and in the other twenty-one sequences, the cyclists continued straight on. All sequences were recorded from the same position near the crossing. The recording position for the video camera was about 20 meters from the crossing and 6 meters above the path, which resulted in a view of the cyclist from a frontal three-quarter view (Fig. 1). The sequences were recorded on five different occasions in February, 2012, and resulted in 20 hours of recorded material.

Fig. 1. Picture of the crossing used in Experiment 1. The cyclist approaches the crossing from the upper right quadrant and continues towards the lower left or makes a left turn at the crossing.

Cyclists were riding on a shared bicycle-pedestrian path. The path was slanted slightly down-hill, and this affected the speed of the cyclists to some extent. The sequences with the most visible cyclists were selected. These sequences contained relatively little visual clutter in the form of other cyclists, pedestrians or heavy traffic. The purpose of this was to diminish visual disturbances while viewing the sequences. The cyclist occupied a mean of 6% of each image in each sequence.

In order to investigate the extent to which observers could predict the future behavior of the cyclists, each sequence was "cut" at six meters from the crossing. The border of the crossing is indicated by the white line in Fig. 1. Since there was no previous research to suggest an appropriate distance for being able to reliably predict cyclist behavior, six meters was determined to be a reasonable distance based on viewing the sequences. Each sequence started with the cyclist approximately 20 meters from the crossing. There was no intended bias in selecting the cyclists whose behavior was easiest to predict. We used all of the sequences that were recorded and that could be presented without too much visual clutter. The forty-two sequences were randomly ordered for presentation to the participants. Two of the sequences were used as practice trials to familiarize participants with the stimuli and task.

Cyclists had different speeds. In order to determine whether the two groups of cyclists (making a left-hand turn (TURN) or continuing straight on (STRAIGHT) had systematically different speeds, we calculated the mean sequence duration for each group: TURN = 3.12 seconds (sd = 0.65) and STRAIGHT = 2,82 seconds (sd = 0.64). The difference was not significant, $t(38) = 1.49$, $p = 0.14$.

3) Questionnaires

In addition to measuring accuracy, we also asked participants to list the cues they thought they used to make their predictions of cyclist behavior. A paper questionnaire was used to register the participants' predictions and three self-reported cues on which the predictions were based.

4) Procedure

Participants were tested in groups in a lecture hall, and the video sequences were projected on a large screen at the front of the hall. Two groups of ten students in each group were recruited after a lecture. After receiving information about the experiment, the students remained in the lecture hall to participate. The students then received more information about the experiment and provided formal written consent.

Participants were instructed that they would view forty sequences of a cyclist approaching a crossing and that twenty would turn left at the crossing and twenty would continue straight on. The order of the sequences would be randomly determined. Each sequence would be "cut" before the cyclist reached the crossing, and the screen would go black. At this point, they were asked to indicate whether or not the cyclist would turn left (YES or NO). They were then instructed to use the blanks on the questionnaire to write a maximum of three cues that they used to make their prediction. They were given 20 seconds to write the cues and then a tone signaled the start of the next sequence. The total participation time for the groups was approximately thirty-five minutes.

B. Results

1) Accuracy

Responses on the questionnaires from each participant were entered into a statistics program, and the mean percent correct predictions for the TURN and STRAIGHT behaviors were calculated. The mean percent correct for TURN was 78%, which was significantly better than chance at 50%, $t(19) = 14.09$, $p < 0.0001$. The corresponding result for STRAIGHT was 75%, which was also significantly better than chance, $t(19) = 8.50$, $p < 0.0001$. The difference between the prediction accuracy for the two behaviors was not significant, $t(19) = 0.88$, $p = 0.39$. These results show that the participants could reliably, and with the same level of accuracy, predict both behaviors of the cyclists based on the information available between six and twenty meters from the crossing.

2) Cues

On average, each participant provided 1.19 cues per sequence and thereby did not fully use the three spaces provided on the questionnaire. The total number of cues collected was 950, which were entered into a database. The cues were then sorted into seven broad categories such as *head movement*, *speed change*, *position*, *perceived speed*, *pedaling*, *blank* and *other*. These broad categories contained also subcategories. For example, the category *head movement* consisted of (cyclist) *looks ahead*, *looks behind* and *looks at crosswalk*. *Speed change* included cues about the cyclist *slowing down* or *accelerating*, and participants also specifically mentioned when there was *no speed change*. We then created a figure of the distribution frequencies of the cues according to the seven broad categories and accuracy for the two different behaviors. See Fig. 2 for the distribution of the cues according to the identified categories.

The results in Fig. 2, Panel A show that there were 411 cues provided when the participants correctly predicted the cyclist would turn. About 40% of the cues belonged to *head movement,* while *speed change* and *position* (cyclist leans or sits straight up) each constituted 15% of the reported cues. The other cue categories appeared to be listed much less frequently. The pattern for correctly predicting turning behavior seems to be indicated by *head movement* (looking behind), *speed change* (slows down) and *position* (cyclist leans or sits straight).

Panel B shows that there were 320 cues provided when the participants correctly predicted the cyclist would continue straight on. Approximately 40% of the cues belonged to the *speed category*, of which almost all indicated that the cyclist had a fast perceived speed. *Pedaling* (continue pedaling) made up 10% of the cues, which is consistent with the indication of a fast speed. There were also many participants who did not mention a specific cue (*no signal*, 15%). The remaining cues appeared much less frequently.

The results for Panels A and B indicate distinctly different patterns that participants seem to use when correctly

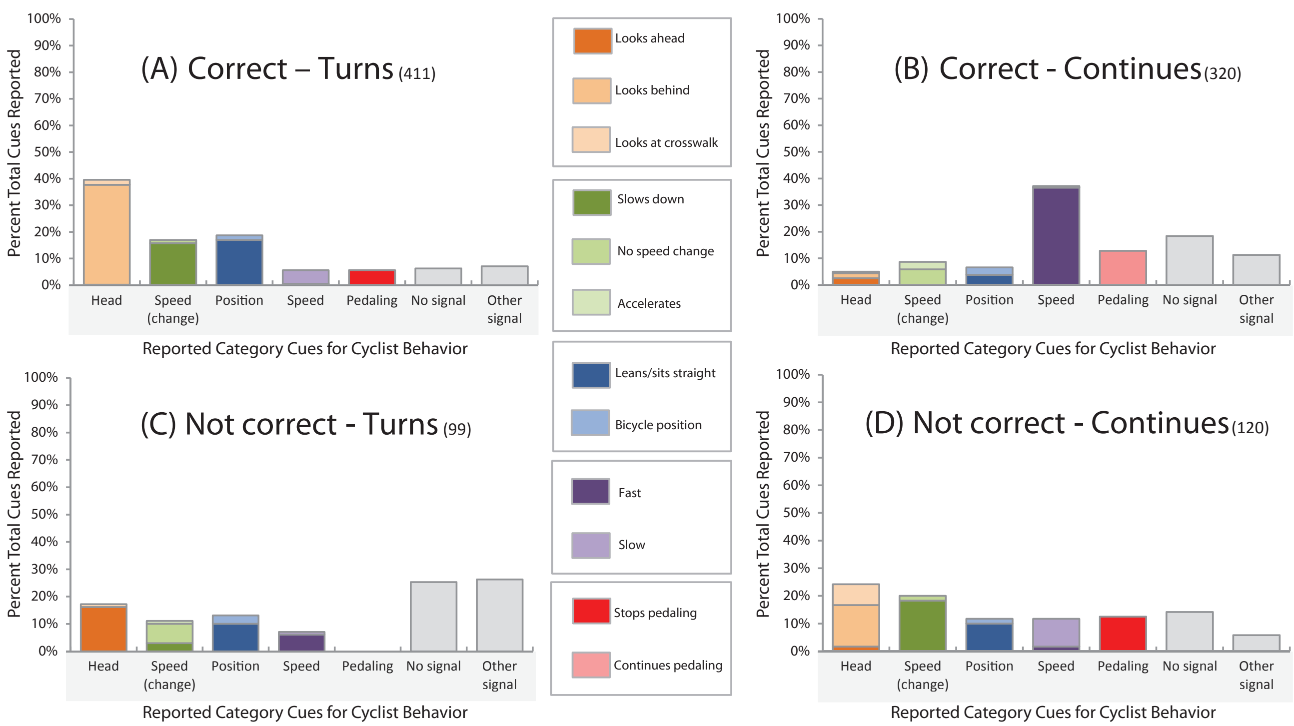

Fig. 2. Frequencies of reported category cues for cyclist behavior as a percentage of total reported cues.

predicting the behavior of cyclists. This result could reflect a combination of cues that are used to predict behavior, as indicated in Panel A. We are not claiming that there is one unique cue that reliably indicates the future behavior of cyclist. The methods in this experiment do not allow us to more clearly specify the role that each individual cue or combination of cues might play in predicting cyclist behavior.

Panels C and D show the distribution of cues for incorrectly predicted cyclist behavior. Panel C shows the distribution when participants *incorrectly* predicted the cyclist would continue straight on, and Panel D shows the results when participants *incorrectly* predicted the cyclist would turn. Although the patterns of frequencies are not as clear as for the correct predictions, there is some indication that the cues used to correctly predict cyclist behavior can also create incorrect expectations about the future behavior of cyclists.

In Panel C for example, *looking ahead* is inconsistent with a turning behavior. If not combined with any other salient cue for turning behavior, participants tended to predict that the cyclist will continue straight on. The majority of cues in Panel C consist of *no signal* or *some other signal* that does not fit into the other categories. Further analysis of the specific sequences and reported cues may provide a clearer account of what is happening in these situations.

In Panel D, *head movement* (looks behind and looks at crosswalk) is interpreted as a pending turn when in fact the cyclist continues straight. There is also an indication in speed change that the cyclist will turn because participants report that the cyclist slows down. The cues belonging to the categories of *position*, *speed* and *pedaling* indicate turning behavior. These results further suggest that a combination of cues is likely contributing to correct prediction as well as erroneous prediction.

III. MULTIPLE LOGISTIC REGRESSION

In order to more precisely determine the contribution of the different cues to the ability to predict cyclist behavior we let three independent observers judge the salience of cues in each sequence. The observers were instructed to judge the following cues (and salience): *head turn* (clear-slight-missing), *perceived speed* (low-medium-high), *perceived speed change* (slow down-remain constant-increase), *cyclist lean* (backwards-straight-forward), *cyclist position* (prepare to turn-no preparation), and *pedaling near the crossing* (pedaling-no pedaling).

This analysis complements the results from the experiment in the following way. In order to validate the self-reported cues from the experiment, the judges in this analysis rated the salience of the cues according to a 5-point graded scale. By using the averaged score for each cue in each sequence as independent variables and the actual cyclist behavior as the dependent variable, the material was analyzed using multiple logistic regression.

We also created two new durations to test the predictive power of the cues. The observers viewed video sequences that were "cut" at 3.5 meters and 9 meters, which allowed us to see the extent to which cue salience varied as a function of cyclist distance to the crossing.

After registering the judgments of the observers, we carried out a multiple logistic regression analysis on the potential contribution of the different cues, or combination of cues in relation to the actual behavior of the cyclists. The question is: To what extent do the cues (individually or in combination) predict the behavior of the cyclists at two different distances from the crossing?

For the sequences that were cut at 9 meters from the crossing, the logistic regression model correctly classified 74% of the sequences when a cyclist turned and 74% when the cyclist continued straight. For the sequences that were cut at 3.5 meters from the crossing, the logistic regression model correctly classified 87% of the sequences when the cyclist turned and 91% when the cyclist continued straight. The Nagelkerke R-squared for fit of the model was 0.83, which indicates a very good fit.

The contribution to the model for each of the variables is presented in Table 1. *Head turn*, *Speed* and *Position* each significantly contributed to the predictive power of the model. Interestingly, *pedaling* and *speed change* did not reach significance. This indicates a discrepancy between the results from the reported data from the participants in the experiment. In the experiment, *speed change* seemed to be a clear indicator of turning behavior as well as a confusing cue that could lead one to believe that a cyclist would turn but who then continues straight.

Table 1. The contribution of variables in a multiple logistic regression model for judgment of cue salience.

	B	S.E.	Wald	df	Sig.	Exp(B)
Headturn	-1.907	0.770	6.133	1	0.013	0.148
Speed	1.876	0.835	5.053	1	0.025	6.528
Speedchange	1.000	0.659	2.305	1	0.129	2.719
Leaning	-0.057	0.537	0.011	1	0.916	0.945
Position	2.551	0.799	10.193	1	0.001	12.822
Pedaling	0.662	0.823	0.648	1	0.421	1.939
Constant	1.204	0.946	1.620	1	0.203	3.333

IV. DISCUSSION

The results indicate that position, head turn and speed were the most critical cues. Results from the experiment showed that humans are reasonably good at this type of prediction for a majority of the situations studied. However, some situations seem to contain conflicting information. Furthermore, participant reports suggested that the actual process of predicting cyclist behavior relies on a single "strong" indicator, e.g., that the cyclist makes a clear "head movement". As a comparison, predictions based on several indicators, using logistic regression, were significantly more accurate compared to human subjects, suggesting a clear benefit if a human would be able to combine more indicators in their intention prediction.

A first step towards a safety system for intention prediction of cyclists is to find out which cues, or combinations of cues, are important and when those cues occur. We have presented data towards this first step. Naturally, due to the real-time nature of the decision process in traffic situations and cognitive limitations of humans, such potential combinations cannot solely be performed by humans. Future work will emphasize the role of a safety system as a real-time decision support to aid the humans in behavior prediction and will play a better role for improving prediction accuracy. Such a support is reliant on methods that automatically identify indicators, based on sensor data, and combines them into a single strong indicator for the intention.

ACKNOWLEDGMENT

This research was funded by Länsförsäkringsbolagens forskningsfond AB, the research division of a Swedish insurance company.

REFERENCES

[1] Blake, R., Shiffrar, M. (2007). Perception of human motion. *Annual Review of Psychology* **58**, 47-73.

[2] Decety, J., Grèzes, J. (1999). Neural mechanisms subserving the perception of human actions. *Trends in Cognitive Sciences* **3**(5).

[3] Giese, M. A., Poggio, T. (2003). Neural mechanisms for the recognition of biological movements. *Nature Reviews Neuroscience* 4(3), 179-192.

[4] Iacoboni, M., Molnar-Szakacs, I., Gallese, V., Buccino, G., Mazziotta, J. C., Rizzolatti, G. (2005). Grasping the intentions of others with one's own mirror neuron system. *PloS Biology*, **3**(3), 0529-0535.

[5] Saxe, R., Xiao, D.-K., Kovacs, G., Perrett, D. I., Kanwisher, N. (2004). A region of right posterior superior temporal sulcus responds to observed intentional actions. *Neuropsychologia* **42**, 1435-1446.

[6] Satori, L., Becchio, C. & Castiello, U. (2011). Cues to intention: The role of movement *information. Cognition* **119**, 242-252.

[7] Vinciarelli, A., Pantic, M., Bourlard, H. (2009). Social Signal Processing: Survey of an Emerging Domain. *Image and Vision Computing Journal* **27**(12).

[8] Walker, I., Brosnan, M. (2007). Drivers' fixations during judgements about a bicyclist's intentions. *Transportation Research Part F* **10**, 90-98.

[9] Johnson, M. H. (2006). Biological motion: A perceptual life detector?. *Current Biology*, **16**(10), 376-377.

[10] Troje, N. F., Westhoff, C. (2006). The inversion effect in biological motion perception: Evidence for a "life detector"? *Current Biology* **16**, 821-824.

[11] Schmidt, S., Färber, B. (2009). Pedestrians at the kerb – Recognising the action intentions of humans. *Transportation Research Part F: Traffic Psychology and Behaviour* **12**, 300–310.