Evaluation of traffic safety, based on micro-level behavioural data: Theoretical framework and first implementation

Aliaksei Laureshyn*,1, Åse Svensson1, Christer Hydén1

Traffic and Roads, Department of Technology and Society, Faculty of Engineering LTH, Lund University, Box 118, 22100 Lund, Sweden

1. Introduction

Traditionally, road traffic safety analysis has relied mostly on accident statistics as the main data source. Over the years, however, numerous problems associated with accident data have been discussed. To sum up, the following aspects are of importance: (i) compared to other events in traffic, accidents are very exceptional in the sense that they are the results of a series of unhappy realisations of many small probabilities; (ii) accidents are rare events, making it troublesome to base traffic safety analyses at individual sites on accidents only; (iii) not all accidents are reported and the level of underreporting depends on the accident's severity and types of road users involved; (iv) information on the behavioural aspects preceding the accident is seldom available.

There is a need to use some kind of surrogate and complement to accidents, i.e., traffic safety indicators, to increase the possibility of: (i) evaluating traffic safety changes more efficiently and in a shorter time; (ii) elaborating the relation between design elements and risk; (iii) more thoroughly understanding the relationships between behaviour and risk; (iv) a better understanding of the processes characterising the normal traffic and critical situations including accidents.

This paper is concerned with micro-level behaviour indicators. These are primarily the indicators describing the motion of road users, e.g., their trajectories (sequences of positions), speeds and derivatives of speed (acceleration, jerk) and those variables that can be calculated from this type of data. Until recently, the use of such indicators on the individual level for the purpose of traffic safety evaluation has been quite limited. For example, the existing traffic conflict techniques (Hydén, 1987; van der Horst and Kraay, 1986; Asmussen, 1984), even though very appealing from a theoretical point of view, have so far relied greatly on using human observers, a factor that limits the efficiency of data collection and the level of details it is possible to achieve. The relationships between many other behavioural indicators and safety have not been thoroughly validated and are far too often based on assumptions and common sense.

Automated video analysis is a rapidly developing technology that might provide a solution for effective behaviour data collection. Today’s video analysis systems (e.g., Laureshyn et al., 2009a,b; Messelodi et al., 2004) are already capable of detecting and tracking road users of various types, and there is a clear trend of increasing the studied area size, improving processing time and accuracy of the results. An optimistic, but quite reasonable, expectation is that in the relatively near future there will be a tool available to provide a detailed description of movements (i.e., co-ordinates related to time) of all road users within the studied area, for example an entire intersection. Such data has great potential for traffic safety analysis, but the practical methodology for it still needs to be developed. In the first place, this concerns the choice of safety indicators to be extracted from the data, the way they are to be analysed and how the results are to be interpreted.

The aim here is to propose a theoretical framework for the development of a method for traffic safety evaluation that utilises the detailed micro-level behavioural data provided by a video analysis.
system or similar tools. We also make a first attempt to develop a set of safety indicators that describe a continuous process of interaction between individual road users and relate the individual interactions to the general safety situation.

2. Theoretical framework

An encounter (a simultaneous arrival in a certain limited area) between two road users can be seen as an elementary event in the traffic process that has a potential to end up in a collision. Hydén (1987) suggests the existence of some severity dimension common for all the events in traffic and proposes a model describing the relation between the events’ severity and their frequency (Fig. 1a). According to this model, the higher the severity (presented as the vertical position in the pyramid), the lower the frequency (the volume of the pyramid slice at this height) of the events.

The concept of severity requires some clarification. The severity of an accident is determined by the accident’s consequences (e.g., number of deaths and injuries or total loss in monetary units). This definition is a bit problematic for encounters that do not end in a collision, as, strictly speaking, a near-miss with just a few centimetres between the vehicles and a completely controlled passage with sufficient safety margins have the same consequences (except for differences in the adrenaline level in the drivers’ blood). Two aspects are to be considered here—the potential of an encounter to become an accident and severity of the consequences if this happens.

The accident potential may be explained in the following way. Accidents are stochastic events. Even though one particular accident may be explained by a number of factors that led to it, it may also be considered as an unlucky coincidence that all these factors happened to be there at the same time. If some of the contributing factors had not been present, the accident might have been avoided.

To put it another way, each encounter can develop into an accident if some new factors come up. In any case, a near-miss has less of a safety margin to endure an additional unlucky factor compared to a well-controlled passage; thus the accident potential of a near-miss is higher.

The severity of the accident’s outcome is influenced by several factors:

- **Type of road users.** If all other variables are equal then: (i) unprotected road users (pedestrians, cyclists, moped drivers, motorcyclists) are likely to suffer more severe injuries than protected road users (people travelling in a car, bus, lorry); (ii) a person travelling in a vehicle of small mass is likely to suffer more severe injuries than a person in a vehicle of large mass; (iii) an elderly person is likely to suffer more severe injuries than a younger person (Englund et al., 1998).

- **Collision angle.** The road users’ angle of approach before the collision may have many different patterns from head-on to rear-end. For collisions involving protected road users this implies different probabilities regarding the collision impact. Head-on collisions are less likely to produce severe injuries than perpendicular collisions (because of less protection provided by the vehicles sides), while rear-end collisions are less likely to produce severe injuries than the other collision types (SIIA, 2009). Presumably, the angle of approach does not have the same effect regarding collision impact when vulnerable road users are involved.

- **Collision speed.** Collisions at higher speeds produce more severe injuries than collisions at lower speeds due to a larger amount of kinetic energy released (Carlsson, 2004). There are indications that the relative speed of the involved road users is a more important variable compared to the absolute speed values.

We leave, for the moment, the question of whether the probability of a collision and the severity of the consequences are to be kept apart or not, and assume that the severity dimension integrates both of them. When severity is assigned to the encounters, they can be placed in some kind of distribution similar to Hydén’s pyramid (such distributions are called severity hierarchies in Svensson, 1998). The way the severity is defined determines the actual shape of the hierarchy. It is reasonable to assume that there is a “true” hierarchy which reflects the objective severity. Introducing various operational measures to describe the severity of an encounter may create many quite different hierarchies in which the same event will probably not be placed exactly on the same level.

If a severity hierarchy represents events at a particular site (e.g., an intersection), a more correct illustration will not be a pyramid but a diamond (Fig. 1b). The least severe events in traffic are quite rare when a road user is completely undisturbed by other road users. The majority of the encounters are of “medium severity”, i.e., road users have to adjust their actions to the other road users, but in a well-controlled manner that characterises the “normal” traffic process. Svensson (1998) also argues for the doubled peak shape of the distribution, but she limits the events included in the hierarchy to only those with a collision course (i.e., at some point the road users will collide if they continue with unchanged speed and path).

Severity hierarchy gives a much better understanding of the situation from a safety point of view compared to accidents that only represent the very top of the distribution. The important question is how the frequency of events in different severity levels is to be interpreted. A robust relation between the frequency of serious conflicts and the actual number of police-reported accidents has been found (Hydén, 1987). Findings in Svensson (1998) suggest that the non-serious conflicts bear different information depending on how close to the serious conflicts they are located in the severity hierarchy. Events located just beneath the serious conflicts, i.e., events with fairly high severities, are characterised by closeness in time and space, thus still having a strong relation to safety. Studies at a non-signalised intersection showed an accumulation of interactions at these fairly high severities while there were no accidents or serious conflicts. Comparative studies at a signalised intersection revealed that interactions with fairly high severities did not
3. Operational definition of severity

The way the severity of an encounter is to be defined in the best operational way is still an open issue. The following aspects are to be considered.

3.1. What measures are relevant?

The literature proposes indicators to reflect the two aspects of severity, the risk of collision and the collision consequences. As a measure of a collision risk, indicators describing proximity in space, proximity in time and intensity of a necessary evasive action can be mentioned (Gettman and Head, 2003; Hydén, 1987, 1996; van der Horst, 1990; Asmussen, 1984; Allen et al., 1977; Hayward, 1971). The general problem is that if only one variable is used, just one side of the truth is reflected (for example, short distance between road users does not say much without information about their speeds).

In this respect, the time-proximity indicators are a bit special since they integrate both proximity and speed. It is probably for this reason that many of the traffic conflict techniques base the conflict severity gradation on some kind of time-based severity measure (Asmussen, 1984). Still, it seems reasonable to complement time-proximity measures with a speed indicator, as, for example, is done in the Swedish traffic conflict technique (Hydén, 1987).

The speed of the road users is also a relevant measure for estimation of the consequences. Other indicators that can reflect the severity consequences are road user type (or some alternative measure of “vulnerability), approaching angle, collision type, etc.

3.2. Relation to a collision course

The collision course at the end of an encounter is a pre-condition for a collision; without it a collision is not possible. Notwithstanding, even encounters without a collision course might have an accident potential, but some changes in spatial or temporal relation between the road users have to occur in order to reach collision course. Generally, these relations at a particular moment may be classified into three types (Fig. 2):

Type A (collision course). Road users are on a collision course—they will collide if no evasive action is taken.

Type B (crossing course). Road users’ planned paths overlap, but collision will be avoided as they pass the common spatial zone at different times. For collision to become possible a correction in time is needed; i.e., one or both road users have to change speed. In other words, the situation has to turn into an A-type situation first.

Type C (non-crossing course). Road users’ paths do not overlap in any way. This occurs when road users move in parallel in two adjacent lanes or have diverging courses. This does not mean, however, that the collision risk is completely zero as some (often very little) adjustment of the path by one or both road users may make their courses overlap and in certain conditions also create a collision course. For example, a pedestrian walking close to the street curb is just a few steps away from cars driving at high speeds, and there is a risk that the situation will develop into a very severe one if he/she suddenly changes the direction of walking. It may be argued, though, that before a collision happens the situation must turn into an A-type situation (collision course), possibly via a B-type situation (crossing course) and then it can be described by methods developed for these situation types.

Even though these three types are theoretically different, they create a continuum in which the transfer between the types occurs smoothly. Since even minor changes in road users’ trajectories (paths) and speed during an encounter may affect whether they appear on a collision course or not, or even if they pass a common spatial zone or not, the behaviour of the road users does not change abruptly at the moment of transfer. For example, Svensson (1998) found that in situations when two vehicle drivers were about to miss each other by a very short time margin, their evasive behaviour was the same as if they were on a collision course. The most obvious explanation is that with short margins even a minor change in speed might put road users on a collision course, and therefore the situation is experienced as being high risk as well. The Dutch traffic conflict technique DOCTOR (Kraay and van der Horst, 1985) includes in its definition of a conflict both situations with a collision course and without a collision course, given that the time margin is small enough. This indicates that the measures used to describe an encounter also have to be flexible enough to allow smooth transfer in the description.

3.3. Encounter as a process

An encounter between two road users is a continuous process and the severity indicators to describe this process should also allow for a continuous description and not only for a certain moment during the process. As the encounter may go through different phases including moving on or not moving on a collision course, the severity indicators must allow for smooth transfers in the description. On the other hand, for an encounter to be placed in a severity hierarchy, it has to be represented by only one value, which has to be derived from a set of indicators all describing the severity of the process with regard to certain aspects. The challenge
is therefore to find a set of indicators that best describes the severity of the process and to unify these indicators into a common severity measure that makes it possible to locate them in one common severity hierarchy. It is, however, very likely that the set of indicators might vary between types of traffic environments, involved road users, accident types, etc.

4. A set of indicators—a first approximation

We propose a set of indicators that continuously describe the process of an encounter and may be used to classify an encounter’s severity. The purpose is to see if it is possible to find a common severity measure for (some of) these indicators that makes it possible to elaborate on a common severity hierarchy. The set includes several time-based indicators and the speed of the road users. Some of the problems discussed in the previous section are addressed, but a solution is not found for all of them. Therefore, this should be seen as a first approximation and further elaboration and validation of the approach are necessary.

4.1. Time-to-Collision

Time-to-Collision (TTC) is defined as the time required for two vehicles to collide if they continue at their present speed and along the same path (Hayward, 1971). Most often TTC is calculated using the simple assumption that the road users’ trajectories cross at a right angle or are parallel (Fig. 3). For example, van der Horst (1990) calculates TTC for the case of a right-angle approach using the following equations:

\[
\begin{align*}
\text{TTC} &= \frac{d_2}{v_2}, \quad \text{if} \quad \frac{d_1}{v_1} < \frac{d_2}{v_2} < \frac{d_1 + l_1 + w_2}{v_1}, \\
\text{TTC} &= \frac{d_1}{v_1}, \quad \text{if} \quad \frac{d_1}{v_1} < \frac{d_2}{v_2} < \frac{d_2 + l_2 + w_1}{v_2}.
\end{align*}
\]  

where \(d_1\) and \(d_2\) are distances from the fronts of vehicles 1 and 2, respectively, to the area of intersection (Fig. 3a); \(l_1, l_2\) and \(w_1, w_2\) are the lengths and widths of vehicles 1 and 2, respectively; \(v_1\) and \(v_2\) are the vehicle speeds.

For the case of rear-end collision (Fig. 3b) Minderhoud and Bovy (2001) calculate TTC as:

\[
\text{TTC} = \frac{X_1 - X_2 - l_1}{v_1 - v_2}, \quad \text{if} \quad v_2 > v_1,
\]

where \(X_1\) and \(X_2\) are the positions of vehicles 1 and 2, respectively.

For the case of a head-on collision (Fig. 3c), the previous equation can be easily modified to:

\[
\text{TTC} = \frac{X_1 - X_2}{v_1 + v_2}.
\]

In the general case, two vehicles can approach each other at any angle and, moreover, different collision types are possible for the same angle (Fig. 4). After analysing all the possible collision types, it can be concluded that it is always a corner of one of the vehicles that meets a side of the other one. Since, in the general case, it is not known which corner meets which side, all possible combinations have to be analysed (i.e., 32 combinations assuming that road users have rectangular forms). The procedure for calculating TTC for a moving line section and a point (i.e., a side and a corner of road users) is given in Appendix A. The lowest TTC value found among all the corner-side combinations is to be used, since it is the side and corner that will come into contact first in a collision.

TTC is a continuous parameter and may be calculated for any moment as long as the road users are on a collision course. It is quite widely used and some variations of this parameter have also been proposed, for example, TTCA (taking into account the acceleration of road users, van der Horst, 1990), inverse TTC (1/TTC, Kiefer et al., 2005), Time Exposed TTC and Time Integrated TTC (complex parameters taking into account the time road users spend on a col-
4.2. Time Advantage

Time Advantage (further abbreviated as TAdv) is an indicator used to describe situations where two road users pass a common spatial zone, but at different times and thus avoid a collision course and thereby collision (Hansson, 1975). Proposed initially as a measure describing “normal” traffic conditions, Time Advantage may be seen as an extension of a safety indicator called Post-Encroachment Time (PET). The conventional definition of PET is the time between the first road user leaving the common spatial zone and the second arriving at it (Fig. 5, Allen et al., 1977). Thus, PET for an encounter has a single value and may be observed and measured directly. TAdv broadens the concept of PET, saying for each moment what the PET value is expected to be if the road users continue with the same speeds and paths.

The conventional geometry-based definitions of PET and TAdv are difficult to apply when the vehicle trajectories do not cross at a right angle (which is not unusual in real life). The entrance and exit from the “common zone” are no longer time moments but periods, and it is even possible that both road users appear in the “common zone” but still avoid collision (Fig. 6).

To overcome this problem other non-geometrical terms have to be used. We propose defining PET as the minimal delay of the first road user which, if applied, will result in a collision course and a collision (assuming, similar to TTC, that, apart from the delay, the road users otherwise continue with the same speeds and paths). Fig. 7 helps to explain this definition. Lines I and II describe the movements of two road users over the time (for simplicity we consider only one dimension and neglect the physical size of the road users). The “delay” of road user I means that its travel line has to be shifted along the time axis until it touches line II. The length of the time shift here is the Post-Encroachment Time. Time Advantage is defined in the same manner, but using the predicted travel lines instead. Obviously, the position of the contact point between lines I and II depends on their shapes, which in turn might differ depending on the moment for which the prediction is made. Further, we will refer to this point as an avoided collision point.

In practical calculations, when the dimensions of the road users are taken into account, the TAdv has to be calculated for each possible side-corner combination. The calculation procedure for this is also provided in Appendix A. For the same reason as in the case of TTC, the lowest found TAdv-value is used.

The specific of Time Advantage is that while its low values may reflect the safety aspects, the higher values (above 2–3 s) describe the normal traffic conditions and may be seen as a measure of one road user’s power (advantage) over the other in a competition over the same spatial zone (Hagring, 2000; Hansson, 1975). A road user having a large Time Advantage is most likely to be the one to pass the common zone first. However, if the Time Advantage is small, the second road user may accelerate with the aim of passing first instead, which occurs primarily when one of the road users is “stronger” than the other, for example, in the case of a private car vs. a pedestrian (Várhelyi, 1998) or a truck vs. a private car. The important point here is that the use of the same indicator to describe both safety and efficiency of the traffic processes has certain advantages and may help to better understand how these two qualities are balanced by the road users and to verify the hypotheses of such a relation (Svensson and Hydén, 2006; Svensson, 1998; Näätänen and Summala, 1976).

4.3. Supplementary parameter T2 to Time Advantage

Time Advantage is by itself not sufficient to describe the collision risk since it is also important to know how soon the encroachment will occur. Even if TAdv is small at a certain moment, the road users might have plenty of time to adjust their speeds and trajectories and increase it. As an indicator describing the nearness of the encroachment, the time of the second road user arriving at the “avoided collision point” is proposed (this parameter is further abbreviated as T2).

To use the second road user appears to be more safety-relevant as his/her arrival at the potential collision point is the very last necessary condition for a collision to occur. Whatever the actions of the first road user, it is the one who arrives last who has the largest margin, i.e., most time to take an evasive action. However, if the moment of the first road user leaving is of interest, it can be easily calculated as (T2 – TAdv).

Another important property of T2 is that it provides “smooth” transfer between the “collision course”- and “crossing course”-situations. At the moment of transfer from “collision course” to “crossing course” the TTC ceases to exist, but Time Advantage still equals zero. This makes T2 equal to TTC, and if both TTC and T2 are plotted on the same graph, they will make a continuous curve. Similarly to TTC, T2 “jumps” into infinity if the second road user comes to a complete stop.

4.4. Time Gap

Depending on the relation between road users’ trajectories and speeds, the “collision point” or “avoided collision point”, for which...
TTC or TAdv are calculated, may be far ahead while the actual distance between road users might be not as large. This is especially noticeable when the road users follow each other in the same track or their trajectories cross at a very sharp angle (Fig. 8).

Time Gap (TG) is a parameter that describes the actual distance between road users expressed in time units. In its conventional definition it is applied to vehicles following in a flow, and is measured as the time between the moment of the rear-end of the first vehicle passing a certain point on a road and the front of the following vehicle arriving at that point (Vogel, 2002). This definition implies that TG is a single value measured directly at a certain location. To make it continuous and more in line with other indicators that have been discussed, a “predicted” Time Gap can be used, i.e., “the Time Gap that will be measured if the road users continue with the same speed and path”.

Still, the conventional definition of TG is difficult to apply if the road users do not follow exactly the same tracks. To extend this parameter and preserve its main concept, we propose the following definition. Imagine that the first of the road users is delayed to such an extent that they start moving on a collision course. There are many possible collision points, depending on the size of the delay. Of all the possible combinations of delay and proximity to collision point, the delay that produces the closest (in time) collision point is chosen. The Time Gap here is the time necessary for the second road user to arrive at the collision point. This definition includes the case of following each other along the same course, but can also be applied for any cases of overlapping courses.

As it is not known in the general case which road user is “the first” and what type of collision is the nearest in time, all possible combinations of road users’ sides and corners have to be considered. The calculation procedure for TG between a point and a line section is provided in Appendix A. Again, the minimal TG value found among all the combinations is to be used.

Time Gap, presumably, has a weaker connection to collision risk compared to TTC, since it only considers the spatial proximity between the road users (in time units), but not their relative speeds. Still, it can be used for the detection of potential risks at earlier stages of an encounter. This can be explained by an example of two vehicles following each other on the same course and at the same speed (i.e., no collision course exists). If the first one starts braking, the vehicles suddenly enter a collision course and the pace of the TTC decrease depends highly on the size of the time interval between the vehicles (i.e., TG). Thus, TG reflects the probability of TTC quickly reaching low values if the road users get onto a collision course.

### 4.5. Speed

Even though time-based indicators reflect both the spatial proximity and speed of the road users, one piece of important information is still missing. This can be shown by a simple example. Imagine two pairs of vehicles on a collision course, but in the first case the vehicle speed is 10 m/s and in the second case 20 m/s. When TTC reaches 1.8 s in both cases, the drivers detect the risk and start braking with maximal deceleration of 6 m/s². In the first case they will manage to stop after 1.6 s and avoid collision, in the second case they will crash with a collision speed of 9 m/s.

This example clearly illustrates that time-based indicators are not sufficient to describe the severity of an encounter and need to be complemented with some speed-related indicator. The way a road user adjusts speed during a passage (road user’s speed profile) also provides important behavioural information and describes the encounter as a process (Laureshyn et al., 2009a,b). For these reasons we include the speed of both road users in the indicator set.

### 5. Two examples—crossing and following courses

Fig. 9 illustrates how the proposed indicator set describes the interaction between two road users. The first example illustrates an encounter between a car and a pedestrian on a pedestrian crossing (the pedestrian has priority). First (phase I), the car has a Time Advantage as the pedestrian hesitates and keeps a very low speed. Then, however, the pedestrian decides to go first and increases speed to normal pace. The TAdv of the car goes rapidly down to zero and from moment \( t_2 \) they are on a collision course (phase II). TTC is decreasing as they approach each other. Having noticed the pedestrian’s behaviour, the driver brakes and from moment \( t_2 \) they are no longer on a collision course and TAdv (now the pedestrian’s) starts gradually growing from zero (phase III). From moment \( t_2 \) the pedestrian is no longer in the way of the car and none of the indicators can be calculated. In this example the TG curve follows the TTC and TAdv curve and does not contribute much additional information.

In the second example the two cars follow each other and appear on the same course until their trajectories diverge. The speed of the following car (marked as 2 in Fig. 9) is higher and the “avoided collision point” (for which the TAdv is calculated) lies in the area of trajectory divergence. Here the distance between the two cars is shorter than the distance to the “avoided collision point”, and the TG curve goes lower than \( t_2 \).

### 6. Discussion

To find a universal indicator that is applicable to any type of situation during the encounter process and reflects all the relevant aspects is not simple; most probably, a set of indicators is necessary. Furthermore, it is likely that the set of indicators might vary depending on, e.g., type of traffic environment and type of road users involved. On the other hand, it is important to keep the number of indicators as low as possible (at the risk of losing some information), otherwise it will be difficult to make the method operational.

As a first approximation, we propose a set of indicators to describe the process of an encounter between two road users. The suggested indicators address some of the problems outlined in the theoretical framework. Many issues, however, have to be further elaborated on.

It is assumed here that an elementary event in traffic, which might result in an accident, is an encounter between two road users. This assumption excludes single accidents, i.e., situations with only one road user involved, even though such accidents do happen. The single accidents can be divided into several types. The first type is the accident where a second road user is involved. The first, in trying to avoid a collision with the second, drives off the road and
collides with an object such as a tree. In this case the encounter does actually take place and the proposed indicators may be applied to describe it. If no other road user is present, the situation with an “active” driver and external factors that lead to an accident (e.g., an animal suddenly jumping onto the road, unexpected ice on asphalt or vehicle malfunctioning) and a “passive” driver and internal factors (alcohol influence or fatigue, resulting in loss of attention or falling asleep) may be distinguished. An approach similar to what is proposed may still be used here, i.e., the existence of a severity hierarchy for such events may be assumed, e.g., accidents, near-accidents when the driver manages to regain the control of a vehicle on the ice or “wakes up” at the last second to avoid a crash and so on. However, the problem of integrating single road user events and encounters into one common severity hierarchy needs to be elaborated.

Another problem is the accidents that involve several road users. The most feasible way, probably, is to treat such situations as sets of multiple pair-wise interactions, since the extension of the elementary event in traffic from an encounter to interaction of several road users implies enormous complications.

There is a great variety of encounter types, with and without collision course, different approaching angles and types of road users, etc. We argue, for instance, that “non-crossing course” situations also have to be included in the severity hierarchy and some indicators covering these types of situations have to be developed. For a collision to become possible, road users on a non-crossing course need to change the relation in space (to get the paths to overlap) and, possibly, in time (to create a collision course). The proposed set of indicators allows for a continuous description and a “smooth” transfer between “collision course” and “crossing course” situations, but the “non-crossing course” must also be included. The new indicators have to reflect all these aspects. The challenge, however, is to find the optimal set of indicators and unify these indicators into some general severity measure, i.e., to place them into one common severity hierarchy.

There is an advantage in using indicators that can be calculated for any moment during an encounter since the development of an encounter as a process may then be studied. On the other hand, a decision has to be taken on what moment, or combination of moments, characterises the severity of an encounter in the best way. Several options are possible. Hayward (1971) and van der Horst (1990) use the minimal TTC value during an encounter (TTCmin) to show how close to a collision road users actually come. The Swedish Conflict Technique (Hydén, 1987) utilises the TTC value at one point in time. It rates the severity based on the TTC value at the moment the first evasive action is started by one of the road users (called Time-to-Accident, TA). Hydén argues that TA reflects the critical moment when the road user has just detected the risk and has to take the first decision on how to act depending on the perception of severity of the event. The problem with...
values taken at a certain time is that they do not incorporate any information before or after the chosen moment, creating a risk that even very different encounters might be classified in the same category. Alternative non-momentarily-based measures may well be the time spent on a collision course or even more complex parameters, for example time-based indicators as proposed by Minderhoud and Bovy (2001).

Another approach that may be used to classify the encounters with regard to severity is to analyse the shape of the continuous indicator profiles, i.e., TTC or TAdv curves, etc. A detailed analysis of these processes will perhaps reveal the “typical” shapes characterising the critical situations. Similarly, shapes reflecting “normal” (non-critical) processes can be found. These types of analyses can use, for example, pattern recognition methods as discussed in Laureshyn et al. (2009a,b).

Many of the proposed indicators are based on predictions of a collision point in terms of the planned path and current speed of the road users. A human observer can easily project the “planned” trajectory, but it is quite difficult to explain exactly how this projection is done. A possible approximation is to assume that a road user actually follows the planned path, i.e., to use the known trajectory. This may be misleading in case the road user avoids a conflict by changing the planned path, for example taking a larger radius in a turn or changing lanes. Another alternative is to use an “average” path, calculated from the trajectories of many road users making the same manoeuvre. The problem, however, is that in critical situations the paths might not follow the average pattern. A detailed analysis of critical situations might reveal when the deviation from the “average” pattern starts to develop during an encounter, and if the high severity of the situation can be detected before that moment, i.e., when the “average” assumptions are still valid. It would also be interesting to compare these objective descriptions with human observers’ perceptions of the situation in order to pinpoint the relevance of suddenness and closeness to collision point when diverging from the “average” path.

van der Horst (1990) tests different variations of TTC definitions based on assumptions of constant angular velocity and constant acceleration of a vehicle (this is supposed to represent a situation when a driver is no longer controlling the vehicle and the steering wheel and the gas pedal positions are kept unchanged). The paths calculated with constant angular velocity easily take very peculiar shapes and lead outside the road. As for constant accelerations, the TTC values are still reasonable, but there is no clear evidence to show that the predictive power of TTC improves.

The methods for combining the accident risk and the severity of consequences into one severity measure are still missing. In the Swedish Conflict Technique (Hydén, 1987) this problem is circumvented by defining the severity of an encounter as a potential of an injury accident, i.e., property-damage accidents are not considered at all. However, there are indications of the existence of a more universal severity measure. During the calibration study of traffic conflict techniques from different countries (Grayson, 1984), the severity rating of the conflicts based on objective measures was compared with subjective ratings of the human observers. A stronger agreement was found among the subjective ratings of different observers than among the objective ratings based on definitions of the techniques tested. One of the explanations offered was that human observers considered both the collision risk and the consequences, while the objective measures often reflect just one of the aspects (in most cases, the collision risk). The challenge, however, is to find an objective and operational measure that corresponds to the subjective severity judgements. This will probably lead to more valid conflict measures.

Our hypothesis is that a feasible way forward regarding the description of the severity of the encounter process is to elaborate further on the shape of the severity hierarchy and the assumption of a “true” hierarchy for a specific set of indicators. Nonetheless, the set of indicators describing the severity of the process has to be translated into one common severity measure since only a single dimension can be strictly ordered in a severity hierarchy. It is important to further elaborate on what the whole shape and the accumulation of events at different levels represent. The severity hierarchies proposed earlier (Svensson and Hydén, 2006; Svensson, 1998) were only based on events with collision courses. It was argued, for instance, that interactions with fairly high severities could be positive from a safety point of view because they were frequent and severe enough to increase awareness, and that these events were predominant at a site without serious conflicts and accidents. It will be interesting to analyse whether the same interpretation is valid for the hierarchies proposed here. With information about the encounter processes and the severity of these processes it will be possible to formulate and test hypotheses on the interrelationships of design of the traffic environment, behaviour and risk. It is also important to point out that the behaviour described by the severity hierarchies could reflect other qualities in traffic besides safety, like mobility and the desire to balance these qualities.

Our expectation is that video analysis is the tool that will provide the necessary micro-level data on the behaviour of road users. With automated video analysis it will be possible to elaborate on the severity hierarchies as long-term recording will provide us with accidents as well. However, some important indicators that have direct implications for the accident risk, and especially the severity of the accidents (e.g., use of helmets, road user age, eye contact and other signals sent by road users) is hard to extract from video data. It might be necessary to complement video analysis with some other data collection method (e.g., human observers who look through a video that has been initially automatically filtered) to get the necessary information.

When introducing a new method and new indicators, the most important aspects are their reliability and validity. The reliability is a property to produce the results of the same accuracy irrespective of where and in what conditions and by whom measurements are made, thus ensuring that the difference in the results is attributable to the difference in the studied quality (safety) and not to a measurement error. The validity guarantees the robust relation of the used indicators with the studied quality. Establishing these two qualities is to be seen as a necessary step in the development of the proposed method. Again, with automated video analysis and a framework around relevant indicators, it will be feasible to elaborate on validity and reliability.

7. Conclusions

An encounter between road users is a process that can be described as a continuous interplay over time and space. For some encounters the road users are on a collision course, while in other encounters there is no collision course. As a further complication, moving on/not moving on a collision course may change during the encounter process. Most prevalent traffic safety indicators do not consider the severity of the whole process, but assign a severity to a certain moment during this process without considering occurrences just before or after this moment. Moreover, safety is far too often treated in a one-dimensional manner as if it is the only motive while moving in traffic. Hence, other motives like efficiency and comfort are not considered. This paper elaborates on how to improve the calculation of the existing indicators like Time-to-Collision, Time Gap and Time Advantage for all types of approach angles and complement them with speed and TAdv in order to estimate the severity. The paper also suggests how to make smooth transfers between the different indicators in order to reveal the
whole encounter process as a continuous interplay between road users. By describing the processes preceding accidents with regard to TTC, Time Gap and Time Advantage, speed and T2, as well as “normal” encounters not resulting in accidents or serious conflicts, it can be possible to analyse what distinguishes a “safe” process from an “unsafe” process. With this information it will be possible to organise all encounters in a severity hierarchy and classify the severity of the encounters with regard to the whole encounter process. This will be a considerable contribution to increasing knowledge of the traffic safety process and understanding road users’ trade-offs between safety and efficiency in traffic.

Appendix A

A.1. Calculation of Time-to-Collision for a point and a line section

Let \((x_p, y_p)\) be the current co-ordinates of the point and \(v_p\) its speed vector (Fig. 10). Then the position of the point at an instant time \(t\) can be described by equations:

\[
\begin{align*}
    x_p &= x'_p + v_p x t \\
    y_p &= y'_p + v_p y t
\end{align*}
\]

where \((x'_p, y'_p)\) is the initial position of the point and \(v_p x\) and \(v_p y\) are the projections of the speed vector on the X- and Y-axes.

The position of the line section ends \((x_{ln1}, y_{ln1})\) and \((x_{ln2}, y_{ln2})\) is described by equations:

\[
\begin{align*}
    x_{ln1} &= x_{ln1} + v_{ln x} t \\
    y_{ln1} &= y_{ln1} + v_{ln y} t \\
    x_{ln2} &= x_{ln2} + v_{ln x} t \\
    y_{ln2} &= y_{ln2} + v_{ln y} t
\end{align*}
\]

where \((x'_{ln1}, y'_{ln1})\) and \((x'_{ln2}, y'_{ln2})\) are the initial positions of the line section ends; \(v_{ln x}\) and \(v_{ln y}\) are the projections of the line speed vector \(v_{ln}\) on the X- and Y-axes.

The line equation in its canonical form is

\[
x - x_{ln1} = \frac{y - y_{ln1}}{k}.
\]

or, in case where the denominator in Eq. (7) is zero \((k \to \infty)\):

\[
t_{coll} = \frac{x_p - x_{ln1}}{v_p x - v_{ln x}}.
\]

Here only positive \(t_{coll}\)-values are of interest. The condition for the point to cross the line within the section is that at the moment \(t = t_{coll}\):

\[
\begin{align*}
    x_{ln1} &\leq x_p \leq x_{ln2} & \text{if } x_{ln2} \geq x_{ln1} \\
    x_{ln2} &\leq x_p \leq x_{ln1} & \text{if } x_{ln1} \geq x_{ln2}
\end{align*}
\]

\[
\begin{align*}
    y_{ln1} &\leq y_p \leq y_{ln2} & \text{if } y_{ln2} \geq y_{ln1} \\
    y_{ln2} &\leq y_p \leq y_{ln1} & \text{if } y_{ln1} \geq y_{ln2}
\end{align*}
\]

The collision point \((x_{coll}, y_{coll})\) coincides with the point position \((x_p, y_p)\) at the moment \(t = t_{coll}\).

A.2. Calculation of Time Advantage and Time Gap for a line section and a point

To find the time distance between a line section and a point, it is enough to check the time differences with which the point and the line section ends pass the common points 1 and 2 (Fig. 11). For point 1 the time difference \(\Delta t_1\) is:

\[
\Delta t_1 = |t_{p1} - t_{ln1}| = \frac{S_{p1}}{v_p} - \frac{S_{ln1}}{v_{ln}},
\]

where \(t_{p1}\) and \(t_{ln1}\) are times necessary for the point and the section end to reach the common point 1; \(S_{p1}\) and \(S_{ln1}\) are the distances from the point and the line section end to the common point 1.

Calculations of the time difference \(\Delta t_2\) for point 2 are done in the same way. Time Advantage will be the minimal value between \(\Delta t_1\) and \(\Delta t_2\).

A special case when the point and a section end trajectory do not cross is shown in Fig. 11. In this example \(t_{ln1} = 0\) and \(\Delta t_1 = t_{p1}\). Similarly, the cases when \(t_{p2}\) and \(t_{ln2}\) are equal to zero have to be considered.
To find Time Gap, points 1 and 2 have to be checked first. The “second” road user takes a longer time to arrive at a common point. For example, the time necessary for the “second” road user to arrive at point 1 is

$$t_{2-1} = \max(t_{p1}; t_{ln 1}).$$  \hspace{1cm} (12)

The time $t_{2-2}$ necessary for the “second” road user to arrive at point 2 is calculated in the same way. Time Gap will be the minimal value between $t_{2-1}$ and $t_{2-2}$. The special cases such as the one shown in Fig. 11 also have to be considered.

These calculations of TG are performed in cases where the point and the line section are not on a collision course. If TTC can be calculated, the Time Gap is equal to TTC.

References


