Control strategies for a highway network

A joint research project of SWOV, the Technical University Delft and the Institute for Perception TNO sponsored by the Dutch Ministry of Transport and Watermanagement

PART II

Dr. P.H. Polak & T. Heijer. Control Strategies for a Highway Network.

E. Wiersma & T. Heijer. Safety of the Traffic Process on Highways.

R-94-34 II Leidschendam, 1994 SWOV Institute for Road Safety Research, The Netherlands

SWOV Institute for Road Safety Research P.O. Box 170 2260 AD Leidschendam The Netherlands Telephone 31703209323 Telefax 31703201261

Contents

PART I

1. Preface

1.1. Introduction

1.2. Synopsis of the results

2. The reports

E. de Romph; H.J.M. van Grol & R. Hamerslag. 3DAS (Three Dimensional Assignment): A Dynamic Assignment Model for Advanced Traffic Management and Driver Information Systems.

E. de Romph; H.J.M. van Grol & R. Hamerslag. Application of 3DAS (Three Dimensional Assignment) at the Washington Metropolitan Area.

PART II

Dr. P.H. Polak & T. Heijer. Control Strategies for a Highway Network.

E. Wiersma & T. Heijer. Safety of the Traffic Process on Highways.

PART III

W.B. Verwey & T. Heijer. Linking Driver Visual Workload on Near-congested Highways to Inductive Loop Measurement.

C.M. Gundy & T. Heijer. A Traffic Model based upon Neural Networks.



Control Strategies for a Highway Network

The development and implementation of a theoretical model for local traffic control

Dr. P.H. Polak & T. Heijer Leidschendam, 1993 SWOV Institute for Road Safety Research, The Netherlands

SWOV Institute for Road Safety Research P.O. Box 170 2260 AD Leidschendam The Netherlands Telephone 31703209323 Telefax 31703201261

1. Introduction

In the previous phases of this project (Heyer, 1991), a control scheme for a network of motorways has been proposed. This scheme features a hierarchical distribution of control actions and a modular structure. One of these control levels has been called sub-systems: a stretch of motorway with at least one on-ramp. It was decided to develop a local controller based upon a specific concept of adaptive control: model reference adaptive control (MRAC). Simply stated, this type of control refers the actual state of traffic to an idealized reference model and decides its control actions on the basis of the difference between actual and idealized state. This reference model is driven on the basis of on-line measured traffic parameters and must be capable of real time operation in order to provide useful reference data.

The aim of the research reported here is to develop such a reference model, together with methods to structure and implement data processing in such a way that real time operation is feasible. The condition of real time operation is not the only condition for the model, however. The problem is yet more complicated as a consequence of the condition that the model must use parameters that are relevant for assessing and controlling the safety of the local traffic stream. Especially this condition necessitated an extension and adaptation of current traffic flow theory. The steps taken in this investigation: the definition of the necessary parameters, the construction of relatively simple empirical models and a possible theoretical explanation are reported here. In the next stage of this project, these models will have to be integrated into a reference model and tested.

2. Basic considerations

Choice of parameters

In existing practice, traffic flow on motorways is usually characterized by using three parameters: average speed, density and flow. Also, these parameters are usually aggregated over the entire cross-section of the carriageway, even though measurements are taken per lane.

Since one of the principal ambitions of this project is to evaluate and improve safety we have to consider whether these parameters are suitable and sufficient for this purpose.

These considerations need an operationalization of "safety" in this context. On the scale of the considered traffic processes it seemed best to relate safety to individual behaviour in the sense that a local traffic stream is considered as safe as possible, when conditions are easily controllable for all participants. Considering that human controlling behaviour is usually best in predictable circumstances, ease of control can be seen associated with (simple e.g. linear) predictability. This leads us to the base parameters for individual predictions and, since the principal human input channel in traffic is the visual channel, these parameters are predominantly related to visible phenomena. Apart from all sorts of traffic signs and instructions these parameters are simple to establish: lane number on the road, individual speed, distances to neighbouring vehicles and speed differences with neighbouring vehicles.

It can now easily be concluded that the usual parameters used to characterize traffic streams are insufficient to relate completely to individual behaviour; we need to extend our range to include relative positional and motional variables of individual vehicles.

From the considerations in appendix 1 we find that induction loops as a measuring tool may provide many of these parameters; only the lateral changes of position of individual vehicles are hard to detect directly. Therefore, it seems to be a workable choice to extend our range of parameters with individual speed, speed difference and gap between adjacent vehicles. All these parameters are taken per lane and may be averaged in various ways. Furthermore, the individual length of the vehicles is included. This length can be used in several ways that have turned out to be rather important during this investigation.

In the first place, the vehicle length can be used to distinguish two major types of vehicles: passenger cars and small vans on the one hand and freight vehicles and buses on the other. These two categories differ strongly in a number of aspects: regulations and manoeuvring characteristics, which are relevant for safety considerations.

In the second place, vehicle length is needed to calculate an other set of stream parameters that are more closely related to individual perception of the traffic state. These parameters can be considered alternatives for flow, density and mean speed, in the sense that, instead of reducing vehicles to points, these parameters are related to the actual size of the vehicle and the coverage of the road. The parameters are derived by weighing flow, density and speed with the vehicle length. Thus we find the following coherent set of alternatives, for one of which we had to adopt a new name:

- lane occupancy	(dimensionless)	in stead of	density
- production	(meters/second)	in stead of	flow

- weighed mean speed (meters/second) in stead of mean speed.

Fundamental diagrams expressed in these parameters show the same general shape as those made with the traditional parameters, but, contrary to the traditional diagrams, the shapes are more alike for different lanes.

3. Measurement and data processing

Induction loops can provide values of three characteristic variables for each detected vehicle: the time of measurement (passing), the (approximate) speed and (also approximate) length. Some of the previously mentioned parameters can be measured or derived directly from these data, some other parameters like intervehicle gap can only be established under certain assumptions regarding the changes in speed of two subsequent vehicles during the time between their measurement : we assume that the speed changes during this time will be negligible.

Even if outward conditions are constant, individual traffic behaviour under those conditions always varies significantly. Some of the variation can be attributed to influence of the behaviour of other nearby road users, the causal part; the rest can be considered random variation or noise. In order to be able to exert some measure of traffic control, we must establish a model predicting the causal part of interactions. This brings up the problem of filtering out noise while retaining the useful information. This will be accomplished by a special averaging process.

3.1 Specifications of the averaging process

The averaging process must fulfil the following demands:

- it must satisfy regularity conditions (e.g. be linear in the measurement values)
- its coefficients should follow from some minimum principal (e.g. minimal sum of squares)
- it must simple enough to be processed locally
- it must be flexible
- it must have an adjustable time constant
- it must be possible to combine lane variables to carriageway variables
- it must produce an new averaged value at every car passing
- the averaged value must lie on a specified curve (e.g. horizontal or oblique line) through the past data
- the curve must be fitted with weights that diminish for older data

3.2 The averaging process

To fulfil the above demands the averaging process is constructed in the following way, where we must keep apart the mathematical formulation and the way the formalism is translated in a (micro)computerprogram.

3.2.1 Mathematical formulation

The averaged value x_a is written as a weighed sum over all past data x_i , where x_0 is the most recent measurement, taken at time t_0 and x_i was measured at time t_i :

 $x_{i} = \Sigma (c_{i} * x_{i})$ with i = 0, 1, 2, 3, ...

We now must specify the constraints for the coefficients c_i . These constraints all follow from our wish that the average computed over certain simple patterns of the data-values has a certain value. The most fundamental of these patterns is that when all x_i are equal to some constant value X, x_a also must have the value X. From this it follows directly that the sum of all the coefficients c_i must be equal to 1. With our - later to be described - minimum principle this fixes the values of the c_i uniquely. In this manner we get a regular weighed average, which however has the usual peculiarity of lagging behind when there is a - downward or upward - trend in the data. To overcome this lagging we can specify a second constraint on the c_i . This would result in a manner of averaging where, if the data were all lying on an oblique straight line, the averaged value would also lie on this same line.

The minimum principle

The constraints do not fix the values of the c_i uniquely. For that we employ a minimum principle that is equivalent to a method known as Discounted Least Squares (DLS) (Harvey, 1981). The value we minimize is a weighed sum of squares of the coefficients c_i . It can be demonstrated that this sum is proportional to a modified expected variance of the computed average x_a . For that we have to assume that the variances of the measured values x_i are larger the more they lie in the past. This can also be interpreted as giving older data less weight in the determination of the averaged value x_a , hence the name Discounted Least Squares. In formula:

minimize $\sum (c_i^2/g_i)$ under the constraint(s),

where the weights g_i have the value $exp(-(t_0-t_i)/T)$. in this way the most recent measurement has weight 1 and older ones are progressively derated. The parameter T gives the time after which the weight has diminished to 1/e (apr 0.37). Large values of T give stable averages because many measurements are taken into account, small values follow changes more quickly. In its simplest form this method amounts to exponential smoothing, but with the important extensions that our method works with unequally spaced data, and that we can specify the pattern that is fitted to the data. These extensions are indispensable for our purpose. For details see appendix 2.

3.2.2 Implementation

Although the formulae we derived are all complicated infinite sums the actual implementation is rather simple.

Each time we record a passing vehicle an update is made of a small number of derived variables. For each to be averaged variable this amounts to one for the horizontal pattern and an additional three for the oblique line fit. Additionally an update is made for the weights. So if the speed, the length, the headway, the passing time and the speed difference with the preceding vehicle are averaged we have to update 4 * 5 + 1 = 21 variables. (If we only use horizontal averaging this number goes down to 6.)

The updates for the horizontal average all go in the following manner-

NEW = measured value + exp(-At / T) * OLD

where At is the time that has passed since the previous update. For the weight update the 'measured value' equals 1. The horizontal average for

some variable v is simply NEW(v)/NEW(weights). For the oblique average and more details see appendix 2.

4. Empirical results

4.1 The data

Empirical data were taken on two different sites :

 the A13 near the town of Delft: a high capacity motorway with three lanes per carriageway

measurements here were taken for 9 consecutive hours starting at 7.45 am in august 1990 and in a single direction

- the A4, around and in the Beneluxtunnel: this is a motorway with two lanes per carriageway.

Measurements were taken during a full week in february 1993, but not completely continuously. Both directions were monitored.

On the A13, 16 cross sections in a row were monitored, spanning a length of apr 8 kilometres.

On the A4, a total of 14 cross sections could be measured and, since the monitors were present in both directions, the total length observed was apr. 2 kilometres.

In both cases, data were taken from every vehicle that passed each induction loop: no aggregation or averaging was performed before storing the data.

Weather conditions during the A13 measurement were clouded but dry with unimpaired visibility. Weather conditions during the week at the A4 site varied: the first 3 days were clear, the remainder of the week the general conditions were more cloudy with occasional showers.

4.2 The fundamental diagram

After deriving the revised stream parameters, one of the first steps we took was to construct the fundamental diagrams from these parameters and compare them with the diagrams based on traditional parameters. This comparison has been made for practically all sites, all showing more or less the same differences and resemblances. These are illustrated in the figures 1 through 4. The diagrams with location nos. like 17023 are taken from the (two times) two lane A4. Those with location nos. like 10001 are from the three lane A13. Always the green line represents the leftmost lane, on the A4 the red line represents the right lane, while on the A13 the red line comes from the middle lane and the blue line from the rightmost lane. The main conclusions that may be drawn from the comparison are, that the diagrams are usually very similar for the two alternate parameter sets, but that the diagrams made with the revised parameters show more consistency between lanes. More specifically: the alternate diagrams of lane occupancy vs speed and production vs speed show curves that are similar save for a vertical (speed) shift. This we deemed a possibly important characteristic, since it implies that a model with a single discriminating parameter, a characteristic speed per lane, may be used to describe the diagrams on all lanes.

The next step, therefore, was to investigate the nature of the differences more closely. In figures 5 through 8 we can see the variation of several

parameters with time at the same location of the A4. As can be seen, all graphs show roughly the same pattern: more traffic on the right lane than on the left during most of the period except during the rush hour, where the left lane scores the highest marks. Also, the differences between the parameters emerge most clearly during the rush hours: where the traffic flow in veh./sec (fig 5) is almost twice as large on the left lane, the production in veh.m/sec (fig 6) does not show nearly so large a difference. The same can be observed by comparing figures 7 and 8 where the density in veh/m also displays a much larger difference between the lanes during the rush hours than the lane occupancy in veh.m/m. So, where the traditional parameters suggest a large difference in use of the lanes during rush hours, the alternate parameters suggest only a small difference. Since the only difference between the two sets of parameters is the weighing with the average length of the vehicles, these differences can only be explained by considering the difference in average length or, alternatively, the different fraction of freight vehicles in the traffic stream in the various lanes. This is illustrated in figure 9.

We see that the left lane is predominantly occupied by passenger cars, while, during traffic build-up in rush hours, practically all longer and heavier vehicles stay in the right lane. Since these heavy vehicles occupy 2 to 5 times more length of road than passenger cars, the vehicle count (flow) on the left lane is considerably higher than on the right lane while the actual percentage of the road length that is occupied by vehicles (the lane occupancy) differs only slightly.

If we consider this from the point of view of individual control by a road user, manoeuvring space on both lanes is practically the same and, since the average speed in both lanes is also very closely the same during the rush hour, lane changing is neither necessary nor attractive. Since the large difference in the flow parameters suggests the opposite, we can conclude that our alternate parameters seem to be more closely related to what road users perceive and hence are the preferable parameters for our purpose.

4.3 The influence of heavy vehicles

Apart from occupying a relatively large amount of space and thus lowering the capacity of the lane, heavy vehicles affect the traffic stream in more and interesting ways. When seen from the point of view of drivers of passenger cars, freight vehicles pose an obstruction to vision, thus frustrating the possibility to observe or predict the traffic situation directly ahead. Moreover, heavy vehicles as a group demonstrate a consistently lower speed than passenger cars in the same lane; this is illustrated in figure 10 where the average speeds for passenger and freight vehicles in each lane are shown in a single graph. In this graph, the lower of the red and green lines pertain to heavy vehicles, the upper line to passenger cars. It is clear that, apart from high density circumstances during the rush hours, the two groups manifest clearly different average speeds in each lane.

Both the visual obstruction and the separate speed regimes, by inducing overtaking, appear to have a significant influence on the behaviour of nearby drivers, especially those of passenger cars. In figures 11 and 12 this effect, specifically on the average speed in adjacent lanes, can be seen for both the 3 lane and the 2 lane motorway. For this purpose the timescale has been stretched and the averaging period has been shortened to apr 1 minute to show the short term effect of heavy vehicles. The upper lines in these figures represent the average speed in each lane, the lower lines with the same colour indicate the fraction of heavy vehicles per lane. As can be seen, the speed in all lanes varies inversely with this fraction. This effect extends into the third lane of the 3 lane road as well, even if there are only heavy vehicles in the rightmost lane.

Apart from this effect on average speed, the influence on overtaking manoeuvres can also be observed to a certain extent. To that end, we have compared the flow in a region next to a heavy vehicle in the adjacent lane to the average flow in that lane, taking into account that during the passage over the induction loop of a heavy vehicle, the flow on the adjacent lane(s) must be higher than average. This effect can indeed be observed, as is illustrated in figs 13 and 14, for the A4 and the A13. Again the exception occurs during the rush hour, when traffic density is so high that overtaking is usually strongly limited.

In conclusion, it can be stated that the influence of heavy vehicles in the traffic stream is considerable. Modelling of the stream must therefore include this influence.

4.4 Parameters related to individual behaviour and safety

The different speed ranges, the visual obstruction and the influence of overtaking probably also affect other parameters of the traffic stream in a lane, parameters that are accessible to individual observation; speed difference and net gap. Therefore, the relation between these parameters and several other stream characteristics were investigated. Eventually from all investigated combinations we have chosen two sets of relations that seemed to provide the clearest insights: the relation between net gap and average speed and the relation between net gap and speed difference, all per lane. Illustration of these relationships can be found in figures 15 and 16.

4.5. Model construction

The empirical relationships suggested to us two relatively simple models, which turned out to fit the data rather nicely.

For the relation between net gap and speed we postulate a horizontal asymptote at some level we interpret as the free speed, and an oblique asymptote through the origin. The intersection of these two straight lines we will call the transition point. The simplest mathematical curve with two asymptotes is the hyperbola. It has as parameters the coordinates of the transition point and a parameter that describes the sharpness of the bend in the curve near the transition point.

The relation between net gap and speed difference resembles a parabola with a horizontal axis. Here there is only one parameter, the quotient of $\star v^2$ and the net gap. This quotient can be interpreted as an acceleration.

To adapt our models to changing traffic characteristics we have made the parameters dependent on the lane number and the fraction of heavy traffic. Fig. 17 and 18 show the fit of the models for some of our data.

It must be borne in mind that these models describe a homogeneous, stationary traffic flow, while our data obviously show some quick changes where one would expect deviations from the models.

4.6 Empirical models and the fundamental diagrams

Because the net gap can be translated to the density or the lane occupancy (with the average length of the vehicles as necessary parameter), we can give the diagrams of our model in this more familiar way. In fig. 19, 20 and 21 this is demonstrated, for a range of fractions of freight vehicles from 0 to 40%. We can see that the whole range, from very sparse to very congested traffic is covered in one simple model. Of course, the most congested part of the curve is absent from our data, so confirmation of that part will still have to be done.

5. Causal interpretation

5.1 Introduction

For this we have developed a new theory for the interaction between vehicles. Instead of - as is usually done - postulating a car following behaviour that is identical for all car/driver combinations which leads to complicated models that still do not fit the observed fundamental diagrams, we explicitly account for a range of car/driver characteristics. We use very simple models which use variables that are observable for drivers.

5.2 Basics

Our aim is to develop a causal theory that describes the flow of traffic on multilane highways, under stable conditions. To be more precise, we set out to model homogeneous, stationary traffic. Homogeneous means in this context that road and traffic are uniform over a stretch of at least several kilometres. Spacial discontinuities like on and off ramps will have to be modelled later. By stationarity we express the condition that the traffic process remains constant during a period of at least several tens of minutes. These are obviously conditions that occur seldom or never in real life. Still a viable theory has to describe these simple conditions first before it can hope to model changing conditions or even break-down conditions. However, a successful stationary model will also describe gradually changing conditions.

For the thus defined stable traffic flow we set out to describe the relation between speed and occupancy, and between speed difference (between neighbouring vehicles) and occupancy, per lane. Also we wanted to describe the relationship between the lanes.

5.3 The fundamental diagram

Classically the fundamental diagram is the relation between two of the three averaged variables speed, flow and density. The three are connected by the algebraic relation flow = speed * density, where the definitions of the three variables have to be consistent. As said before, these classic variables treat vehicles as points. We have found that it is essential to treat vehicles as objects with length. In our analysis it turned out that the net gap was the more fundamental variable. Its (algebraic) relation with density is density = 1/(net gap + length), where - as always in this discussion all variables are averaged in the above described manner. The empirical relation between net gap and speed as shown in figure 15, and the hyperbolic model as shown in figure 19 can be derived by assuming for every individual vehicle a wish speed or free speed which the vehicle will have when it is not influenced by other vehicles. When traffic gets denser the vehicle will keep its wish speed as long as possible, from time to time changing to an adjacent lane if necessary. When this is no longer possible it will change from constant speed to constant net headway in relation to the obstructing vehicle in front of it. Constant headway manifests itself in a speed v net gap diagram as a straight line through the origin with net gap divided by speed in m/s equal to the net headway in seconds. We found 1 second a typical value. Our model for individual behaviour thus consists of the two asymptotes of our empirical relationship, the latter

curving gradually from one asymptote to the other. We now have to explain this form of individual behaviour and how a collective of sharply broken lines can make a gradually curved line.

Individual behaviour

The horizontal part of the individual curve needs no explanation. It is perfectly plausible that a driver tries to maintain his wished for speed as long as possible. The oblique part can be understood by assuming that a driver will follow his predecessor as close as possible, allowing enough distance between them so that when his predecessor starts braking he will have time enough to follow suit. In this manner he has only to take account of his own reaction time, which is of the order of 1 second.

Collective behaviour.

Here we have to acknowledge that the driver population can - for the purpose of our theory - be considered as a joint distribution of free speeds and reaction times. When we take account of this and of the possibility for drivers to change lanes it can be demonstrated that the average behaviour will follow a curve of the form as postulated. For a more detailed treatment see appendix 3.

5.4 The relation between speed difference and net gap

As said before, our model is simply the constant quotient of the square of the speed difference and the net gap. In terms of human behaviour this means that drivers adjust - in the mean - their net gaps to the speed difference with the cars in front in such manner that the average acceleration (or deceleration) needed to stay clear of these other cars stays within a certain small value. We found typical values of Δv^2 / net gap of 0.02 m/s².

6. Conclusions: consequences for local control

As was demonstrated, the models relating net gap, speed and speed difference and fraction of freight vehicles can be used to generate credible, accurate forms of the fundamental diagrams. Thus, by carefully choosing model parameters, the basis for a reference model has been developed. The reference model is an essential part of the proposed adaptive control scheme.

Important in this respect is that the models found do not only describe empirical relations, but also have a plausible causal structure. Thus we may be more confident that, by controlling one of the parameters, other related parameters will change according to the models. Moreover, the models describe traffic situations that generated no incidents. Although this can not be taken as a guarantee that the models describe circumstances that will never lead to incidents (safety monitoring will remain necessary), those circumstances have proven "manageable" to drivers on a succession of days. Of course, since the model is based on these stationary circumstances we can only use them in locations where stationary conditions can be expected. In practice, this means that they can be applied to traffic conditions appr. 500 m downstream of a junction.

General form of a local controller

For the most basic form of local control, using our reference model, we consider a limited stretch of freeway with a single on-ramp. Given the limitations mentioned before, we could then build a local controller using the following "recipe" for a metering control cycle:

- use the measurement data, especially percentage freight vehicles, just before the on-ramp to establish the current form of the reference model,
- establish a reference point in the current model close to capacity (capacity here in terms of *production* instead of *flow*);safety criteria can be used to determine this point more exactly, use the difference between the production pertaining to this reference point and the actual production as input for the ramp metering control: adjust the metering light timing interval in such a way that averaged the on-ramp production (measurable with a pair of induction loops) does not exceed this difference. The latter may be achieved e.g. with a simple proportional algorithm, fuzzy control or by other (simple) methods.

This ramp metering procedure, consisting of the updating of reference model and metered ramp production is of course cyclic with a cycle time in the order of minutes.

Coordination between neighbouring sections

With several access ramps in relatively close proximity, the traffic state at the downstream end is most indicative for the overall performance. In any case, the total on-ramp production should not exceed the maximum production of this downstream reference. The limits for individual ramps can now be calculated as a weighed assignment within this limit e.g. using the length of waiting queues (or other suitable characteristic) as a weighing criterion. The allotted production must be considered a temporary maximum for each ramp.

Using longer term predictions in the local controller

Especially during the onset and decline of rush hours considerable changes in traffic volumes take place. Since controlling actions have a certain time lag which depends on the local situation, these predictions can be used to compensate for this lag. Also, and more importantly, the long term predictions can be used to anticipate the local effects of bottlenecks in the network and to control the buildup of queues adjacent to these bottlenecks by timely limitation of the allowed production of the local controllers. Usually, this limit will be related to all on-ramps in a certain area, which than will have to divide this limited production capacity according to the aforementioned rules. Again, the local situation may cause the local controller to choose a lower production than this limit but not a higher one.

Speed control

The current means for speed control do not allow a really strong control of that important parameter. Normally, speed will be adapted to gradually changing conditions by the "natural" mechanisms of traffic behaviour. Ramp metering, the stronger control mechanism, is therefore also intended to ensure such gradual changes. Only in those conditions that a greater change in speed will be unavoidable, e.g. near bottlenecks, speed control, or at least advance warning of significant changes in speed, will be needed. As long as there are no means for stronger speed control, existing sign systems will suffice for this warning function. The longer term predictions may be used to improve the timing of these messages.

Special circumstances, e.g. caused by extreme weather conditions, often require a different combination of speed and headway to maintain safety than normal traffic behaviour implies. Rampmetering and speed control cannot rely any more on the natural behaviour to generate acceptable traffic conditions. This means the reference model must be changed to represent safe conditions under these circumstances. Speed control must, if possible, be much more forceful and in any case recognizably different from speed control in normal conditions.

Model adaptation to special circumstances

Adaptation of the model reference control can take place by changing the parameters of the reference model. As stated before, dangerous weather conditions for example will often require such changes. The adaptation of model parameters in these conditions can either be based upon empirical data or on theoretical adjustments of speed levels and desired (increased) headway.

Appendix 1

The analysis of induction loop data

1. Introduction

In principal, complete knowledge of the traffic process on a road of a certain length and during a given period could be represented by the trajectories of all vehicles that passed the road during the period. Also, information about the width of the road, the number and layout of lanes etc. would be needed, in addition to vehicle properties as length, width, mass and height. In practice we must be satisfied with only a small fraction of this information. In this investigation we had at our disposal measurements made with a large number of induction loops. On the one hand this type of measurement gives rather precise values of speed and vehicle length for each passing vehicle, on the other hand nothing is known about the acceleration of the vehicle or its lateral position and displacements. Neither do we have any knowledge of what happens between the induction loops. Still it turned out that these data are a good basis for describing the traffic flow for our purposes.

2. Physical layout of the induction loops and associated circuitry.

Induction loops, as used in Dutch freeways, always come in pairs. In this way, besides the usual measurement of speed and time of passing, also the vehicle length can be determined. The presence of electrically conducting material (always present in vehicles) in the vicinity of a loop (of copper wire) changes its selfinduction. Because each loop is part of a resonating circuit the frequency at which it resonates (between 47 kHz and 104 kHz) changes during the passage of a vehicle. This change is easily measured. When the frequency change rises above a certain threshold an electrical signal is generated. The time of change of each loop is measured with a resolution of one millisecond. Each passing of a vehicle gives rise to four times (see Fig. 1). After some consistency checks the three times t_1 , t_2 and t_3 are used for further processing.



3. The primary data.

The speed follows from the times t_1 and t_2 :

$$v = \frac{2.5}{t_2 - t_1} \quad meters/second \tag{1}$$

The electrical - length of the vehicle can be derived from the passing time or covering time $t_p = t_3 - t_1$:

$$l_{a} = v * t_{p} = \frac{2.5 * (t_{3} - t_{1})}{t_{2} - t_{1}} meters \qquad (2)$$

The electrical length l_e is approximately equal to the mechanical length of the vehicle l.

In this way each vehicle passage generates three times out of which its moment of passing, its speed and its length can be inferred. Because the resolution of the time measurements is 1 ms the speed, as calculated from formula (1), is grained, especially at high speeds. E.g. at 120 km/h the resolution is 1.6 km/h! The same goes for the length measurement, but here we have an additional source of possible error: formula (2) is only correct when the average speed during the passage of the two loops equals the speed as computed from formula (1) (and a fortiori for constant speed). When the vehicle accelerates or decelerates while passing the loops the computed length is too short respectively too long. Because only averages over some time and over many vehicles play a role in the sequel these errors are effectively filtered out.

4. Derived data.

The primary data all regard one vehicle. The variables we can derive by manipulating the primary data from one loop or several loops can be distinguished in variables that regard one vehicle, variables that regard two neighbouring vehicles and variables that regard more than two vehicles. These variables usually fluctuate heavily from one measurement to the other and have to be averaged before further processing. The methods of averaging are described in appendix 2.

4.1 Variables that regard one vehicle.

Strictly speaking the speed and the length are derived variables. The only primary variables are the time o

f passing of the front of the vehicle t_1 , the same of the rear of the vehicle t_3 , the passing time t_p and the time the vehicle goes from the first loop of the pair to the second, $t_2 - t_1$.

4.2 Variables that regard two vehicles.

We shall only consider variables that regard two successive vehicles on the same lane. We can derive some variables without further ado, but other variables will need hypotheses about the behaviour of the vehicles in the period just before or after the measurement. The most common of these hypotheses is the isoveloxic hypothesis (Haight, 1963). This means that the vehicle keeps a constant speed during the relevant period.

4.2.1 Variables that need no hypotheses.

Here we mention the gross headway (the time between the passing time of the fronts of two successive vehicles or the same for the rears), the net headway (the time between the passing of the rear of the first vehicle and the front of the next one), the speed difference Δv and the individual occupancy (the passing time divided by the gross headway).

4.2.2 Isoveloxic variables.

An important variable is the gap, the distance between two vehicles. This distance has only a meaning when the moment is given at which the distance is valid. Because the data the gap is computed from are taken at different times a somewhat arbitrary choice has to be made. We have chosen the moment of passing of the last vehicle of the two. Also we reckon the gap as belonging to the last vehicle. To calculate the gap we extrapolate the movement of the first car as if it continued its movement with constant speed. In this manner we obtain the gross gap (from front to front or back to back) and the net gap g (from the back of the first car to the front of the second).

Other variables that can be defined in this way are the distance to collision, taken from the position of the loop and the time to collision (TTC)- For the latter we have to choose again the starting moment from which we reckon the TTC.

4.2.3 Semi-isoveloxic variables.

Here one of the vehicles is supposed to keep a constant speed (usually the first) and the other reacts in a prescribed way (often a constant deceleration). We can define the minimal deceleration needed to avoid a collision a_m :

$$a_m = \frac{1}{2} \frac{(\Delta \nu)^2}{g} meters/second^2 \qquad (3)$$

4.3 Variables that regard more than two vehicles.

Here we can think of variables that describe the existence and movement of clusters of vehicles. Also the interaction of three or more vehicles in overtaking situations belongs to this class.

Appendix 2

The averaging process for non-equidistant measurements

1. Introduction

There exists a vast literature on smoothing and forecasting methods for time series, most of which is restricted to series with data points that are measured at regular intervals. In our case there are three important sources of deviations from regularity: periods of high flow follow periods of low flow, and on roads with several lanes the flows on different lanes can be quite different. And even in very regular traffic the vehicles pass a given point irregularly. So a method is needed that allows for irregularly spaced data. It turned out that such a forecasting method (based on Holt's method (Holt et alii, 1960)) is available in the literature (Wright, 1986), but - like the regularly spaced methods it is derived from - it is based on ad hoc arguments so its properties are difficult to assess. Also it is not clear how the method can be extended, e.g. to give higher moments of the variables as the variance. This necessitated the development of the method as described in the main text. A complete treatment will be published separately. Here we will give a complete set of the formulae used in our algorithms.

2. The formulae.

First we will give the formulae for the estimation of the current value of one variable p with a horizontal line fit through the data. Then follow the additional formulae for the oblique line fit. For each variable its defining equation is given, a sum over all data from the beginning of the measurement till the most recent; and an updating equation, a linear combination of the last update and the most recent value. The data are numbered zero for the most recent and higher integers for earlier ones. The series thus go from 0 to infinity, with infinity representing the oldest measurement. The measurements are thus represented by the series p_i , with i ranging from 0 to infinity. The measurement times are given by t_i . We define the weights $g_i = \exp(-(t_0-t_i)/T)$ with T the time constant. We need two weighed sums, the sum of the t_i and the g_i :

	Defining equation	Updating equation
total weight:	$G = \Sigma g_i$	$G_{New} = 1 + g * G_{Old}$
weighed sum of p:	$\mathbf{P} = \boldsymbol{\Sigma} \left(\mathbf{g}_i * \mathbf{p}_i \right)$	$P_{New} = p + g * P_{Old}$

where $g = \exp(-(t_{New}-t_{Old})/T)$ and p the most recent measurement of the p_i . The desired horizontal average of the p_i now follows from

$$p_{hor} = P / G = P_{New} / G_{New}$$

For the oblique line fit we need three more series and updates:

defining equation updating equation

1st moment of the time:

 $M^{(1)} = \Sigma (g_i * t_i) \qquad M^{(1)}_{New} = g * \{M^{(1)}_{Old} + At * G_{Old}\}$

2nd moment of the time:

$$\begin{split} M^{(2)} &= \sum (g_i * t_i^2) & M^{(2)}_{New} = g * \{ M^{(2)}_{Old} + 2 \texttt{A} t * M^{(1)}_{Old} + (\texttt{A} t)^2 * G_{Old} \} \\ \text{weighed sum of } p^*t: \ Pt &= \sum (g_i * p_i * t_i) \quad Pt_{New} = g * (Pt_{Old} + \texttt{A} t * P_{Old}) \\ \text{where } \texttt{A} t &= t_{New} \text{-} t_{Old}. \end{split}$$

The oblique line fit now follows from:

$$\mathbf{p}_{\text{line}} = \mathbf{p}_{\text{hor}} + \mathbf{M}^{(1)} * \mathbf{A} / \mathbf{G}$$

with A the slope of the fitted oblique line:

$$A = \frac{P * M^{(1)} - Pt * G}{M^{(2)} * G - (M^{(1)})^2}$$

Appendix 3

Speed-density relations in the isoveloxic regime.

1. Introduction

It is a well known fact that the average speed on lanes of a multilane highway already start to diminish when the gaps between the vehicles reach values of the order of 300 meter. This seems hard to reconcile with our model of individual behaviour, which states that drivers keep their wish speed till they approach a vehicle in front to a distance of the order of 50 m. It turns out that it is possible to derive the lowering of the average speed on a lane at rising intensity without any car actually lowering its individual speed. This would have as a consequence that in this region of densities the average speed of the whole carriageway should remain constant. This has indeed been reported (Van Toorenburg, 1980). The solution of this apparent paradox - speeds per lane going down while the speed of the carriageway remains constant - lies in the fact that quicker vehicles tend to spend a greater part of their time (or distance covered) on the left lane the denser the traffic gets. This means that the speed on the right lane is averaged over a subset of all vehicles which gradually is reduced to the slower vehicles. In this appendix we will derive the speeddensity relation for a very simple model where only the right lane shows a lower speed as the density rises. A complete treatment will be published separately.

2. The model

We will consider a two-lane highway in one direction. The vehicle population will consist of two equal streams of identical vehicles with speeds v + s and v - s. The average speed of the whole carriageway is thus v. All vehicles keep to the right lane except when a quick car must overtake a slow one. We postulate that - as measured from the slow vehicle - the quick car goes to the left at a certain distance, overtakes it and goes to the right lane after moving in front of it over another distance. The total distance the overtaking vehicle remains on the left lane is thus a constant, independent of the density. We will call this distance - as seen from the moving system of coordinates of a slow vehicle - the overtaking distance.

3. The density on the left lane

We will first derive the density on the left lane. In our simple case only the quick vehicles will populate the left lane. The speed there will stay at a constant v + s. The density on the left lane

can be seen as a product of two processes: each quick car stays left for a fraction of its time which is proportional to the density of slow cars, because this fraction equals the overtaking distance divided by the intervehicle gap of the slow cars. The density of quick cars on the left lane is thus proportional to the total density of slow cars multiplied by the total density of quick cars. Because we gave both types of cars constant proportions of the total density of all cars the density on the left lane turns out to be proportional to the square of the total density. If we call the total density D and the overtaking distance L the density left will be:

$$D_1 = L * D^2 / 4$$

4. The speed-density relation on the right lane

The average speed on the right lane can as a consequence be calculated as:

$$v_{r} = v - s * L * D / (4 - L * D)$$

We see that for low densities D the speed goes linearly down with the density, in accordance with our empirical model.

References

Harvey, A.C., *Time Series Models*, Philip Allan Publishers Ltd., 1981 (1988)

Heijer, T, A.R.Hale, J.W.F.Wiersma and Wu, Bai-fan, Electronica in het Wegverkeer: Beheersingsstrategieen II SWOV, Leidschendam, 1991

Holt, C.C., F. Modigliani, J.F.Muth and H.A.Simon, *Planning Production Inventories and Workforce*, Prentice-Hall, Englewood Cliffs, N.J., 1960

Toorenburg, J. van, De Stationaire Verkeersstroom op de 2x2-strook Autosnelweg, Rijkswaterstaat, Dienst Verkeerskunde, Nota 80-3, 1980

Wright, D.J., Forecasting Data Published at Irregular Time Intervals Using an Extension of Holt's Method, Management Science, Vol. 32, No 4, April 1986



Figure nr: 2















Figure nr: 16

net Gap (m)



Figure nr: 18







Safety of the Traffic Process on Highways

Erik Wiersma Safety Science Group Delft University of Technology Kanaalweg 2b, 2628 EB Delft, The Netherlands & Tom Heijer SWOV Institute for Road Safety Research P.O. Box 170, 2260 AD Leidschendam, The Netherlands



Summary

Foreword

In this research project we are developing an adaptive control strategy for the motorway system. The purpose of the system is to optimize traffic flow under safe conditions. This report describes how we try to define safety characteristics of the traffic stream in our project.

Traffic data derived from detection loops in the road surface, form the basis for a quantitative description of the traffic situation. These data will be compared to a traffic reference model which interprets the safety of the situation on the basis of a set of safety criteria. On the basis of this interpretation, action may be taken to change traffic flow conditions, for instance by reducing speed or by increasing following distances between vehicles.

If we want to describe safety on motorways, we have to take perception of drivers into account. What they see and do determines safety on the road. Existing models do not describe a traffic stream in terms that relate to perception of road users.

We therefore have to define these safety characteristics of the traffic stream in our project.

We use a mesoscopic approach, in which we focus on the traffic process. The safety criteria are a result of extensive research, combining several methods, including video analysis, expert opinion research and measurement of mental load of drivers. This report briefly describes last years' projects.

Results

Experiments show that percieved safety of a traffic stream is a combination of traffic stream characteristics and behaviour of individuals in this traffic stream. There is a connection: some traffic stream characteristics invoke dangerous behaviour.

Classic traffic stream characteristics need to be redefined for describing safety of a traffic stream. A description of safety characteristics uses elements that relate to the perception of road users: Manoeuvring space, presence of freight, distances between cars, etc.

Different research approaches -video analysis, measurement of mental load of drivers, etc-., lead to similar traffic characteristics for describing the safety of the traffic stream.

These characteristics can be determined within the traffic stream with currently existing equipment and can be quantified with software developed by Tom Heijer and Peter Polak.

Currently we are trying to define criteria to distinguish between situations that are controllable for individual road users and uncontrollable situations.

Conclusions

It is possible to describe a traffic stream in quantifiable characteristics that relate to safety and to perception of road users.

We have not yet determined criteria to describe a percieved break point between safe and dangerous traffic situations. Experiments do not indicate such break point yet. The combination of research methods we use may provide these criteria.

The approach of this research, study of the traffic process on a mesoscopic level, is useful. It creates a new methodology for traffic research, which combines latest developments in traffic stream theory and traffic behaviour theory.

Recommendations

Research in this area cannot be limited to one approach. The combination of different research approaches -quantifying traffic charateristics, analysis of video fragments of the traffic stream and measurement of mental load of drivers within the stream- is necessary to model the complex interaction between road users.

In the next year we will pay more attention to driver information systems. We not only need to know what changes in traffic flow are required, we also need to know how these changes can be accomplished.

1 Introduction

1.1 Background

Traffic is growing continuously. A traffic growth is expected of at least forty percent in the next twenty years. Present road capacity will not be sufficient to comply with the demand of traffic. This will lead to an increasing number of traffic jams. At the same time developments like industrial Just-In-Time management systems increasingly demand a traffic system in which reliable predictions about travel time can be made. The cost traffic accidents cause to society are very high (in the Netherlands approximately 1300 deaths a year, total economic costs of traffic accidents are estimated at eight billion guilders a year). Government targets have been set to try to reduce traffic casualties with 25% in the year 2000, compared to 1985.

The situation mentioned above has led to an emphasis on the development of sustainable safe traffic systems. On motorways there has been a change in the approach to traffic flows: The traffic will have to be controlled instead of being merely monitored. Electronic systems will be needed to help with measuring traffic and informing the driver on desired behaviour. A current research program at the Delft University of Technology and the Institute for Road Safety Research SWOV is developing an adaptive control strategy for the motorway system. The purpose of this hierarchical system is to provide a maximum traffic flow on a motorway network under safe conditions. The overall system is described in other papers [Wiersma 1993, Heijer 1993]. This report describes how we define and use safety characteristics of the traffic stream in our project.

1.2 Traffic reference model

In our model, the actual traffic situation is compared to a reference model of the traffic which imposes conditions on other layers of the system for dealing safely with the traffic. The system uses traffic information that is primarily generated by detection loops in the road surface. These detection loops keep track of every passing vehicle and provide an estimate of the speed and length of each vehicle. The data must be interpreted and evaluated in terms of safety. This evaluation is then used to imposes conditions on other layers of the system for dealing safely with the traffic and to provide information to the road user, to induce safe behaviour. This information must be attuned to the road users' capabilities through the use of proper semantics, placing and timing of the messages.

The traffic reference model we are developing at the moment is a mesoscopic model. The central focus is on the traffic process itself, i.e. the interaction between road users. To be able to make predictions about safety we need to model the dynamic characteristics of the traffic stream, and at the same time account for the behaviour of the drivers that make up the traffic stream.

Our traffic reference model also needs to be quantitative. The detection loops that provide the input for the model generate the speed and length of each passing vehicle. With these basic parameters we try to make a quantitative description of the current traffic situation. When the traffic situation up- and downstream is taken into consideration as well, it is possible to make a prediction about what will happen in the near future. This description still does not provide us with a measurement of safety, or a guideline for interference in the traffic flow. There has to be a translation of the actual traffic state to a judgement about safety. In a number of experiments we are trying to obtain a qualitative interpretation of the safety of the traffic. The quantification of these qualitative judgements will enable us to create our traffic reference model.

1.3 Previous research

For years detection loops in the road surface have been used to describe traffic flow charactristics. Most researchers report in terms of standard traffic stream theory characteristics: flow, density speed. These macroscopic traffic flow descriptions cannot account for behaviour of individual drivers within the traffic stream. Safety is not modelled directly, but is a derived aspect of overall traffic characteristics.

Microscopic traffic research, on the other hand, focusses on behaviour of individual drivers. In these research tradition, traffic is considered the context in which to study people and their behaviour. Modelling of the traffic process itself, the most significant aspect in the model we are developing, is considered not relevant or impossible to model.

We have proposed a mesoscopic approach of the traffic process [Wiersma 1991]. We use the same detection loops that have been used in traffic research for years. We are interested in the possibilities of quantifying individual passages. These quantitative data open a new field for traffic description. This description combines aspects of the macroscopic traffic flow characteristics with microscopic thinking.

In our previous reports we have pointed out how we want to use expert opinions and video fragments of motorway traffic to relate judgements of traffic safety to quantitative characteristics of a traffic stream. In this report we will extend on that line of thought

In our project we also build on previous research on driver mental load, as has been carried out for years, for instance by TNO Institute for Perception [Verwey 1990, 1991]. In these projects mental load is studied in different task environments, for instance traffic situations. We want to extend this to the motorway traffic in rush hour conditions.

1.4 Report layout

This reports describes the results of last years projects. In a number of experiments we have used video fragments of motorway traffic. Experts in the field of traffic safety had to judge safety in these fragments and these judgements were related to traffic characteristics. Some of these experiments were part of the paractical term and graduation project of a student. The line of research is described in chapter 2. We had planned to do a pilot experiment on driver mental load in september 1992 on the A13 between the Hague and Rotterdam. This experiment would be carried out together with TNO-IP, using a double-task method. This pilot has taken place in the summer of 1993 and is described in a separate report by Verwey and Heijer

In the meantime we have done another pilot experiment instead. The project was carried out as a graduation project. In this experiment we have measured driver mental load on another road, using another method, with another research institute as partner. Results indicate that this line of research should be continued. This line of research and the pilot experiment are described in chapter 3.

In chapter 4 we will state the results of the research: what did we find and how do we connect to others in the project.

2 Safety related traffic stream characteristics

2.1 Introduction

We have used video observation to obtain a qualitative interpretation of the safety of the traffic on the motorway. In a series of experiments we have used the opinions of experts to find a set of criteria that would enable us to define whether a specific traffic situation is safe or not. The persons who participated in this project were considered experts in the field of traffic safety. They were chosen from different backgrounds: Traffic police, scientists with different backgrounds and experienced drivers. They had to judge the safety of the traffic in a number of frames. In the first experiments we were interested in how an evaluation of a specific traffic situation is made. What are the criteria experts use in forming an opinion about the safety of the traffic? These experiments resulted in traffic characteristics experts use in determining the safety of a situation. However, they did not result in strict criteria to distinguish between safe and dangerous situations.

Later experiments tried to quantify the criteria found earlier. In these experiments we tried to define the moments experts wanted to interfere in the traffic stream to improve on safety.

2.2 Qualitative analysis of traffic safety

Research questions:

- 1. What are the criteria experts use in forming an opinion about the safety of the traffic?
- 2. Do experts agree?
- 3. Do different experts use the same criteria?
- 4. If video fragments are used, what is the influence of the fragment itself on judgement (camera point of view, length of fragments etc.)?

2.1.1 Methodology

A video of motorway traffic was recorded from a fly over at a height of five meters on the A13 between the Hague and Rotterdam (see fig 1). Both up-stream and down-stream video tapes were made, around four o'clock an accident happened under the camera, that cannot be seen on the video tapes. The circumstances that led to the accident, however, are shown on the tape. The driver of a van spotted the camera and braked, thinking it was a speed control camera. The cars driving behind this van were at such a close distance that they could not reduce their speed in time to avoid an accident. Several times during the day cars had used their brakes in front of the camera, but only at about four o'clock did this result in an accident. Our experiments were designed to try to find out if the experts were able to determine this critical situation.

In a serie of experiments, fragments of these video tapes often to thirty seconds were shown to experts in the field of traffic safety, scientists, traffic police, experienced drivers etc. They had to judge the safety of the traffic in a number of frames.

Apart from the video experiments we also used interviews held at our institute on on location points where traffic could be observed. The experiments and interviews have been described in greater detail in separate reports [Bangert 1992, Bangert 1993, Wiersma in prep].

2.2.2 Video experiments I (Pilot)

Experimental design

Materials

Video tapes

From the previously mentioned video tapes fourteen pairs of taped traffic situations were taken from the period between 15:25h and 16:05h. Each pair consisted of two fragments of thirty seconds. Video fragments were picked randomly. The first seven pairs showed upstream traffic, the last seven pairs showed downstream traffic.

The video was shown on a standard VHS-VCR.

Score forms

On a form subjects had to score for each trial: which one of the two fragments was more safe Whether the difference in safety between fragments was large or small Why they scored this particular fragment more safe (short reasoning). There was no list of traffic characteristics

Subjects

Twelf subjects participated in this experiment, five subjects were professionally involved in traffic research, the others were experienced drivers (experienced meaning they had had their driving license for at least three years and were driving at least 15.000 km/year.

Protocol

The experiment took about thirty minutes. In this period pairs of fragments were shown. After each pair the video tape was stopped. Fragments could be replayed on request. In some cases filling in the score form was preceeded by discussion about what was shown.

Results

The aim of this first experiment was not to determine which fragments were scored safe, but to get a grip on the reasoning that was used in determining safety of specific situations. Table 1 shows the traffic characteristics the subjects used. Sometimes subjects with different background referred to the same aspects of traffic, using different terminology. Traffic experts and drivers agreed on which fragments were safe, but used different criteria for supporting their judgements. Also drivers used more criteria. This finding is consistent with literature on this topic, for instance [Rasmussen 1984].

- 1. Flow
- 2. Speed
- 3. Density
- 4. Homogenity
- 5. Gaps
- 6. Speed variance within a lanes
- Speed variance between lanes
- 8. Overtaking behaviour
- Use of lanes
- 10. Limitation of view
- 11. Manoeuvring space
- 12. Lane changing
- 13. Behaviour of trucks
- 14. Proportion heavy traffic
- 15. Illegal manoeuvring and unexpected behaviour
- 16. Aggressive driving behaviour

Table 1. Categorised reasoning for safety scores

Some methodological questions were raised during the analysis, for instance influence of camera point of view: upstream versus downstream. The point of view of the camera influences traffic safety judgements. A fine example of this is trucks overtaking. When the camera is pointed upstream, an overtaking truck is seen as a bottleneck for the traffic upstream. A video fragment of downstream traffic shows the truck blocking the view of the upcoming traffic. In both cases truck overtaking is considered unsafe (see fig 2). Although the result is the same, the underlying reasoning is different.

2.2.3 Video experiment II

Experimental design

Materials

Video tapes

This experiment used the same tapes that were used in the previous experiment. From the period between 12:00h and 16:00h thirty second fragments were taken every 3.3 minutes. The fragments were linked in pairs, the second fragment taken thirteen minutes later than the first. From this set five subsets were created, consisting of twelf of pairs plus two of these pairs in reverse order. Within a subset pairs were randomized. From the total set of fragments ten extra fragments were picked and added at the end of each subset, to enable comparison between subjets. Of the total amount of fourteen pairs, first seven pairs showed upstream traffic, the last seven pairs showed downstream traffic. The video was shown on a standard VHS-VCR.

Score forms

The score forms of the previous experiment were used in this experiment

as well.

Of the last ten fragments safety was scored on a five point scale. Subjects had to justify their scores.

Subjects

Twenty subjects participated in this experiment, five subjects were professionally involved in traffic research, the others were experienced drivers, experienced meaning they had had their driving license for at least three years and were driving at least 15,000 km/year.

Protocol

The experiment took about sixty minutes. In this period pairs of fragments were shown. After each pair the video tape was stopped. Fragments could be replayed on request. In some cases filling in the score form was preceeded by discussion about what was shown. The last ten fragments were shown one by one.

Results

There were no significant differences with the results from the previous experiment. Upstream video fragments were scored slightly more consistent among subjects, but strong conclusions can not be drawn, since the number of subjects was relativly small. No fundamentally different concepts were used in argumenting safety. Safety scores on the downstream video fragments were more consistent between subjects.

Interviews

After these two experiments we can not be sure that we have captured all relevant traffic characteristics for determining safety of a specific traffic situation. Only aspects that were present in the video fragments used have been extracted. We therefore had several interviews in different settings, to enable experts to use different aspects of traffic, not expressed during the experiments. Some of these interviews were held at our institute, with or without use of videos. Others were held at a point where actual traffic could be observed. One of these interviews took place on a point with an overview of the road, others took place within an inconspicuous police car, surveilling the motorway. In these interviews the experts could point to situations in the traffic around them that they considered important for evaluating traffic safety.

2.2.4 Results

The safety of a traffic situation is determined by a combination of factors. Some of these relate to events, some to characteristics of the traffic stream.

The events the experts mentioned in the experiments include trucks overtaking, cars driving in the wrong lane and abrupt manoeuvring. Events are dangerous only in a specific context. The same manoeuvre may be considered dangerous on a crowded road, but harmless on an empty one. This leads to one of the main concepts used in later experiments: Failure toler ance of a traffic stream. This is defined as ability of (other drivers in) the traffic stream to react correctly to an incident or failure of one of the drivers.

Traffic characteristics that were considered important for safety included heavy traffic, close following distances, lane changing, unexpected and complex traffic situations. Safety is related to available space on the road. Not the number of cars, but the space they occupy is important. Another concept was introduced here: Likelyhood of making errors. This refers to characteristics of the traffic stream that induce incidents or failures. Special attention was given to the role of trucks in traffic. They occupy a lot of space on the road, and overtaking trucks work as a bottleneck for upcoming traffic. They also block the view for following cars. Although there is a large consistency between the experts from different backgrounds in their judgement about the safety of the situations, there is a difference in the traffic characteristics the different experts use to form their judgement. Sometimes these different characteristics can be related to identical situations as the truck overtaking example shows. Although experts with different background used different terminology, they did not just refer to traffic characteristics that specifically belonged to their domain:civil engineers, psychologists and police officers alike mentioned traffic stream characteristics and individual behaviour of specific drivers as more or less dangerous aspects of what they were shown.

2.3 Quantification of traffic safety

The experiments in paragraph 2.2 have led to a traffic description based on characteristics experts use when they analyse the safety of a situation. These characteristics have been implemented in computer software for traffic stream description [Polak, 1993].

In the next serie of experiments we tried to find exact criteria, breakpoints between safe and dangerous traffic situations. Again we used selected video fragments and asked experts to judge safety in these fragments.

2.3.1 Video experiment III

In the third experiment we tried to find a lineair relation between a number of traffic characteristics, scores for safety and necessity for interference in the traffic stream [Bangert 1992]. Experts again were shown selected video fragments of motorway traffic. The fragments showed downstream traffic only. Fragments varied in a number of traffic characteristics: density, distances, variance in speed and many more. Subjects had to score safety, failure tolerance, likelyhood of errors and necessity of interference in the traffic stream in each fragment.

Experimental design

Materials

Video tapes

One hundred and twenty video fragments with a thirteen second length were selected from the tapes used in the first and second experiment from the period between 15:00h and 16:00h. Only downstream video tapes were used. Video fragments where the traffic situation was dominated by particular incident, for instance with very aggressive behaviour, were omitted from the set. Each subject was shown 58 fragments. The video was shown on a Umatic-VCR. Before and after the experiment each subject was shown a subset of five fragments (the same set, but randomised), to familiarise the subjects with the experimental conditions. Scores on these fragments were used for comparing results.

Traffic data

Traffic data were scored for each video fragment, derived from induction loops in the surface of the motorway. For left-, middle- and right lane separately, for each fragment we scored:

- Traffic flow
- Coverage
- Length of longest vehicle (middle -, and right lanes only)
- Ratio trucks-passenger cars (middle -, and right lanes only)
- Harmonical averaged speed
- Standard deviation of speed
- Minimum TTC
- Number of lane changes between left and middle lane
- Number of lane changes between middle and right lane

Last two variables were scored directly from video.

Score forms

On a form subjects had to score for each trial: Failure tolerance (ten point scale) Likelyhood of making errors. (ten point scale) Safety (ten point scale) Would you interfere in this traffic situation (Yes/No)

Subjects

Four subjects participated in this experiment, one was professionally involved in traffic research, the others were police officers from the traffic division (AVD).

Protocol

The experiment took about sixty minutes. In this period the fragments were shown. After each fragment the video tape was stopped. Fragments could be replayed on request.

Data analysis

Using a *t*-distribution test a significant difference was shown between test trials before and after the experiment. Safety and failure tolerance were scored lower, while Likelyhood of making errors was scored higher (beyond .95 level two-tailed level).

Regression analysis using all independant variables did explain 33% varianceof safety scores. With only the first four variables (coverage left lane, length of longest vehicle middle lane, minimum TTC right lane, harmonical averaged speed left lane) 32% of all variance was explained. Table 2 shows the traffic characteristics that are most important for safety.

- 1. Coverage left lane(pr, 0.48; Sig T 0.0000)
- 2 Length of longest vehicle middle lane(pr, 0.14; Sig T 0.0255)
- Miminum TTC right lane(pr, 0.14; Sig T 0.0444)
- 4. Harmonical averaged speed left lane(pr, 0.10; Sig T 0.1564)

Table 2. Explained variance safety score

12

Discriminant analysis was used to select the most important variables connected with failure tolerance. Table 3 shows the most important variables

- 1. Coverage left lane(pr, 0.46; Sig T 0.0000)
- 2 Length of longest vehicle middle lane(pr, 0.13; Sig T 0.0482)
- 3. Coverage middle lane(pr, 0.11; Sig T 0.1122)
- 4. Harmonical averaged speed left lane(pr, 0.08; Sig T 0.1504)
- 5. Minimum TTC right lane(pr, 0.07; Sig T 0.0122)

Table 3. Explained variance failure tolerance

The same analysis was performed for likely hood of making errors. The most important variables connected with this are shown in table 4

- 1. Coverage left lane(pr, 0.50; Sig T 0.0000)
- 2. Minimum TTC right lane(pr, 0.22; Sig T 0.0011)
- 3 Length of longest vehicle middle lane(pr_i 0.18; Sig T 0.0055)
- 4. Traffic flow right lane(pr, 0.15; Sig T 0.0230)

Table 4. Explained variance likelihood of making errors

Results

The concepts that were scored by the experts (safety, likelihood of making errors and failure tolerance) are closely related.

No clear direct relation between subject scores and traffic characteristics was found. The largest proportion of variance was explained by only one variable: Coverage of the left lane. Other variables, like mimnimum TTC right lane, maximum length vehiclemiddle lane and traffic flow right lane followed at long distances.

Conclusion

There are at least two possible explanations for this last phenomenon. First of all it is possible that there are no dangerous situations in the video fragments used. We doubt this, since an accident happened right below the camera. Secondly it is possible that experts are unable to recognize dangerous situations from video. In a later chapter we will go into this questions further.

2.3.2 Video experiment IV

In previous experiments overtaking trucks were considered dangerous in some situations. The purpose of this experiment is to determine criteria for restriction of truck overtaking. In this experiment we focus on dynamic traffic characteristics, and neglect effects of infrastructural differences and weather conditions. Research questions for this experiment are:

- 1. Is restriction of truck overtaking desirable?
- 2. What are the circumstances under which truck overtaking should be restricted?
- On what criteria is restriction of truck overtaking desirable: safety, traffic flow or both?

In this experiment we use expert opinions to relate traffic situations to safety characteristics. Video fragments of traffic are judged by experts.

Traffic characteristics of the video fragments are quantified. Traffic characteristics are also registered by induction loops in the road surface. Combination of these quantitative methods with the expert opinion experiment enables us to relate quantitative dynamic aspects of motorway traffic to qualitative judgements of traffic safety

Method

In this experiment a number of experts have been shown video fragments of motorway traffic. Each of the fragments shows a truck overtaking other traffic. The experts answered a number of questions for each fragment.

Equipment

Video fragments

Videotapes of motorway traffic have been made on august 30th 1990. These tapes show the A13-west between The Hague and Rotterdam from a flyover near the city of Delft. The tapes show both upstream and downstream traffic. From these tapes 30 fragments were selected showing truck overtaking manoeuvres. These include both upstream traffic (21 fragments) and downstream traffic (9 fragments). The length of the upstream fragments is 20 seconds, downstream fragments have a length of 15 seconds. First upstream fragments were shown, then downstream. Within these sets fragments were presented in random order. One of the overtaking manoeuvres is shown twice, in trial 2 and trial 21, to see if expert scores are consistent.

A number of overtaking manoeuvres is shown from both upstream and downstream angle, for comparison of both conditions (see table 5). Scores on these trials are not expected to be exactly equal, since upstream and downstream shots have been made a few hundred metres apart, but overall traffic characteristics are equal and scores should be comparable.

directionframeframe

up/up 221 up/down 526 up/down 628 up/down 823 up/down 1125 up/down 1224

Table 5 Same overtaking manoeuvres

Video equipment

The fragments were shown on a 67 cm tv set with a VHS video casette recorder with remote control.

Score form

For each video fragment the subjects had to answer four questions.

- 1. Does this trucks' overtaking obstruct traffic flow? (Yes/No)
- 2. According to you, is this trucks' overtaking unwanted? (Yes/No)
- 14

3.

4.

- How much (negative) influence does this overtaking manoeuvre has on safety on a five point scale.
- If this situation is not safe, is it because of:
 - Clustering of cars
 - Obstruction of view
 - Manoeuvring space
 - Other ...

Subjects

Twelve subjects participated in the experiment. Most subjects are working at the Delft University of Technology, ranging from professor to student. Other subjects are working at Institute for Road Safety Research SWOV and TNO-Institute for Perception. All subjects are involved in traffic research, some are working in the field of traffic flow control, while others focus on traffic safety or traffic behaviour research.

Procedure

The experiment took place in one session for all subjects together. They received a short introduction to the experiment and an explanation form of the experiment in Dutch. This explanation form consists of an introduction to the experiment, the research questions and short description of the method used. Each subject filled in a scoreform individually. Video fragments were shown one at the time. The VCR was stopped after each fragment until all subjects had completed their questions. The experiment took about 60 minutes.

Results

Table 6 shows the (negative) influence of the overtaking manoeuvre on safety on a five point scale and standard deviations. Most video fragments were considered safe and without problems for other traffic. Most safe (score = 1) was given more than 75% of the time (see fig #2). Most dangerous (score = 5) was not scored at all. Average score was 1.80.

Nr	Scorestdv	
1	1,730,99	
2	1,090 42	
3	3,180,98	
4	2,091,24	
5	1,821,08	
6	2,640,92	
7	2,091,06	
8	2,640,57	
9	2,090,94	
10	1,180,48	
11	2,090,95	
12	1,450,93	
13	1,550,52	
14	1,450,69	
15	1,910,94	
16	2,361,17	
17	1,550,82	
18	1,090,42	
19	1,090,42	
20	2,091,04	
21	1,180,48	
22	1,440,59	
23	2,221,31	
24	1,220,44	
25	2,221,20	
26	1.440.53	
27	2.001.28	
28	1.781.07	
29	1.560.73	
30	1.891.05	
table 6	Safety scores	

Although scores on a safety scale were relatively low, overtaking manoeuvres did hinder the traffic flow according to the subjects. Table 7 shows the percentages of subjects who considered the overtaking manoeuvre unwanted.

Trial	Scoreflowundes. man.
1	1,732727
2	1,0999
3	3,181873
4	2,099173
5	1,8290
6	2,6410082
7	2,097336
8	2,648264
9	2,095536
10	1,1809
11	2,098264
12	1,4509
13	1,55180
14	1,4599
15	1,914545
16	2,369155
17	1,55189
18	1,0990
19	1,0900
20	2,093618
21	1,1890
22	1,441122
23	2,227867
24	1,2200
25	2,223333
26	1,44011
27	2,006722
28	1,786744
29	1,56011
30	1,893333

table 7. Percentages traffic flow hinder and undesired overtaking manoeuvres

In a number of trials identical obvertaking manoeuvres were shown. Table 8 compares safety scores of these manoeuvres.

scores	on same overtaking manoeuvres
Trial	DirectionScoreStdv
2	upstream 1,090,42
21	upstream1,180,48
5	upstream 1,821,08
26	downstream1,440,53
6	upstream2,640,92
28	downstream1,781,07
8	upstream2,640,57
23	downstream2,221,31
11	upstream2,090,95
25	downstream2,221,20
12	upstream 1,450,93
24	downstream1,220,44
table 8	. Identical overtaking manoeuvres in different trials

Trials were spread throuhout the day, with a focus around four oclock. in thew afternoon. Figures #3 & #4 show scores of safety, traffic hinder and undesired manoeuvres throughout the day.

Conclusion

Although subjects often considered overtaking manoeuvres of trucks unwanted and a hinde to other traffic, situations are scored relatively safe. It is not possible to establish criteria for lilitation of truck overtaking manoeuvres based on this experiment alone, but results strengthen our opinion that truck overtaking is one of the factors that must be carefully examined as one measure in traffic flow control.

3 Mental load research

3.1 Introduction

A methodology we want to apply to determine whether a traffic situation is safe or not, is to investigate mental load of drivers in different traffic situations. Research into mental load is widely used where an operator has to fulfil one or more tasks, which require a lot of mental effort. The task demands on drivers on the motorway are usually low. However, considerable task demands are possible here.

Driving is an interactive process between a driver and his environment. Every driver has his characteristics, (knowledge, temper, fatigue, attention level, etc.) and every situation has its specific circumstances, (traffic, weather-, and road conditions,etc). As a result of this interaction, every driver has a certain behaviour to handle a traffic situation with. Situational characteristics are subject to change. Traffic can become heavier, merging traffic can become complex. Most driver characteristics are considered relatively constant (experience, temper) or gradually changing (fatigue). To deal with a situation, drivers can change their attention level, their mental effort.

3.2 Research questions

In our project two interesting questions arise. The first has to do with the level of mental load, the second with traffic conditions.

- 1. Does mental load rise to a dangerous level? In other words: is traffic (still) controlable for individual drivers.
- 2. What are the traffic circumstances under which mental load changes?

There are theoretical problems to the first question. Mental load researchers do not agree on a standarized score for mental load. In every experiment conditions are compared, but they are not linked to a number of standard tasks or a standard scale. This leads to conclusions about difference in mentqal load between conditions, but not to absolute scores. Therefore the question what level of mental load is dangerous may not be answered.

In the context of the second question there is another phenomenon we would like to investigate. Driver models describing motorway traffic behaviour distinguish between two types of driver control. In light traffic the driver can anticipate the behaviour of other drivers and act accordingly. This is called anticipatory control. In heavy traffic, with short following distances, the driver can only react to the car in front. This is called pursuit control. Determining at what moment drivers change from one mode of control to another may help us understand traffic behaviour and its relation to safety.

3.3 Measuring mental load

Roughly speaking, mental load research techniques can be divided in three categories: subjective measurements, physiological measurements and performance measurements.

The most simple method of mental load measurement is asking the driver about it. This gives a fair impression of the mental load the driver experiences. Although the metohd is intuitively very acceptable, there are some drawbacks in using it. Percieved mental load is sensitive to reconstruction errors, when events are scored retrospectively. Another problem is that adaptation to the task environment can lead to over- or underestimation of metnal load. Subjective judgements can be accurate in situations which require low mental effort.

Physiological processes in the body reflect mental activity. Perspiration, brain activity, increased heart beat etc. have been used in mental load research. Different processes work on the same mechanisms, for instance stress, alertness, movement and mental load all effect heart beat ritmh. This makes data sensitive for noise and interpretation of data complicated. Variability of the heart beat is associated with mental effort. Especially the 10^{-1} Hz. component of the heart beat variability is thought to reflect mental effort. If mental effort increases, the heart beat variability decreases.

The third method of measuring mental load uses performance as an in-

19

dication. The idea behind this is that when mental load increases, for instance when more tasks have to be performed, the performance on these tasks decreases. Globally, two types of performance measurement can be distinguished: measuring performance on the primaire task (in our case the driving task) or adding a secundary task.

Primary tasks on which performance can be measured are, for instance, steering rates, lateral control etc. There are indications that sometimes increasing mental load can also increase performance, so performance is not a total inverse reflection of mental load, but since we are interested in performance and not in theoretical concepts this problem may be overlooked.

Secundary tasks are added to the driving task. Subjects are instructed to perform normally on the primairy task. In demanding situations increased effort on the primairy task will lead to decreased performance on the secundary task. This performance is then considered a reflection of mental load. Adding a secundary task is in itself increasing mental load. Therefore researchers are looking for in between tasks: tasks that are a part of the driving task, but not directly related to vehicle control. An example of this is mirror watching.

Differences between- and applicability of different techniques have been discussed to large extend in a number of papers [Verwey 1990, Linde 1992], but researchers do not agree on their final choice for a methodology. They do however agree on the fact that, if possible, more than one method should be used at once.

3.4 Projects

In our research program we had planned a pilot experiment in september 1992 together with TNO-IP. In this experiment performance on a secundary task and steering rates are coupled to a number of traffic characteristics in rush hour traffic. We wanted to do this experiment on the A13, the motorway that had been used earlier. In this period control over the road and detection loops has been transferred from the VCRR-Delft-Zuid to the VCRR-Benelux. We therefore were unable to use the detection loops for data collection. Since we needed these data we had to postpone our pilot experiment until june 1993. This experiment is described in a separate paper [Verwey, Heijer 1993].

Meanwhile, we have done another pilot instead. Together with VSC-Groningen we have done an experiment using physiological measurements of mental load. In this project we have linked heart beat variability to changing traffic flow conditions on the A6 in Groningen.

3.5 Pilot experiment

3.5.1 Introduction

In an experiment we have linked mental load to traffic flow in rush hour traffic. The experiment was carried out as part of a graduation project [Linde 1992]. In this experiment we were supported by researchers of VSC Groningen.

3 5.2 Method

Five subjects participated in this experiment. During rush hour they drove on a motorway in an instrumented car, while their heart beat was measured. On this particular motorway, the A6 near Groningen, in morning rush hour traffic is dense towards the city and light away from the city. This is reversed in the evening rush hour. Each subject drove during one morning and one evening a part of the motorway, which included both directions.

Measurements of driving performance, heart beat variability and scaling of subjective mental load were combined in every trial. Data were processed using equipment of VSC.

3.5.3 Results

Unfortunately, rush hour was not very heavy during the time of the experiment. Subjects did not find the task very demanding in their subjective ratings. Nevertheless, an effect on heart rate variablity of traffic density was found.

Scores seem to depend more on incidents than on differences in traffic density.

As a point of interest a clear influence was found of unexpected upcoming traffic during an overtaking manoeuvre. Although the driver in the particular trial declared not be frightened by the incident, heart rates showed a clear rise in mental load for the remainder of this trial.

3.5.4 Conclusions

The pilot experiment shows the usefulness of physiological measurement in mental load research. Even in low density traffic situations effects can be measured. But we should not be too optimistic. Careful experimentation, with good design, a combination of methods and a lot of subjects are necessary if we want significant answers to our research questions. Although the results of the pilot are hopeful, they must not be used as definite answers.

4 Characteristics of traffic safety

The purpose of this research project is to understand what makes a traffic stream safe or dangerous. This understanding must lead to traffic stream characteristics that can be quantified with detection loops and indicate whether a traffic stream is safe or not.

Our starting point was a traffic description based on characteristics that have been used for decades, flow, density, speed. Although these characteristics may be useful in describing the performance of a road in terms of flow and capacity, they are not sufficient when we want to describe safety. Safety has to do with characteristics that relate to what drivers experience. This is usually the research territory of social sciences. In our projects we have tried to account for this in three separate lines of research.

We use a traffic description method no one else use. the data collected with the detection loops provide almost any traffic description we want. Development of this data processing is absolutely necessary for our research. The work of Heijer and Polak on this field immediately fits in our experiments. We need their quantitative description of the traffic stream for our experiments, although we use the same data in a slightly different way.

We look at traffic from a distance. Using video fragments of actual motorway traffic we try to find safety characteristics of traffic that can be seen from the outside. In this context people are able to define traffic characteristics that can help in describing the safety of a traffic stream. In other parts of the project we have shown that these characteristics can be used in a quantitative description of the traffic [Polak 1993]. This results in a traffic description that does not only incorporates safety aspects of the traffic stream, but also can be used in models that fit the empirical data better than classic traffic descriptions.

Safety does not only depend on traffic stream characteristics. It is always the combination of incidents and a context, that lead to accidents. In our control strategy we want to model this context. We want to maintain a traffic stream that allows incidents, a traffic stream that is **failure toler**ant. At the same time we don't want the stream to induce unsafe behaviour and incidents. Here we need the concept of **likelyhood of making errors**. With these two main concepts we try to model the traffic stream. So far we have only focussed on the context, which is necessary if we want to define an incident at all. Further research will be necessary if we want to complete the picture. Incidents, conflicts and mistakes have not been defined in our context. But we have defined the context that may be used as framework to do just that.

Although our experiments show that experts seem to know what traffic characteristics influence safety, they are unwilling or unable to define criteria that can be used as breakpoints between safe and dangerous situations. They carefully avoid to make their judgements of safety concrete. When asked they refer to the complexity of the process and to uncertain elements involved.

Our third line of research goes one step deeper into the traffic process. The central question here is: Is the driver still able to cope with his situation? The traffic process is made up by drivers who act and react to each other. If the situation gets out of control for them it is a sign for us to interfere in the process. When they can cope with a situation we better not interfere.

A problem here is that mental load researchers do not agree on the methods used, on standarized tests, scales etc. Research in this area is complex and progress is slow. Our pilot experiment shows one method that may prove useful, our experiment with TNO-IP uses a method that has already proved itself in previous projects.

We are not alone in this line of research. A recent report of INRETS [Malaterre 1993] shows they are following the same line of thought. The important distinction between our project and theirs is, that our way of measuring traffic characteristics is much more advanced than anybody elses.

Figures

The Hague





Figure #2





Safety Score



Figure #4



Percentages hinder and undesired overtaking manoeuvre

Figure #5

References

Bangert, W.F. 1992. Kwalitatieve analyse van verkeersveiligheid met behulp van videoregistratie. (in Dutch, Summary in English in preparation). Delft: Safety Science Group, Delft University of Technology.

Bangert, W.F. 1993. Kwantitatieve analyse van verkeersveiligheid met behulp van videoregistratie. (in Dutch, Summary in English in preparation). Delft: Safety Science Group, Delft University of Technology.

Linde, R.J. v.d., 1992 Het bepalen van de invloed van dynamische verkeerskenmerken op de mentale inspanning van de automobilist rijdend op de autosnelweg. (in Dutch). Delft: Safety Science Group, Delft University of Technology.

Malaterre, G., 1992 Evaluation de la charge de travail du conducteur au cours dúne tache de conduite simulee. Convention DSCR-INRETS 91-41024-01-224-75-01, etude no 14.

Polak, Heijer 1993, in this series of reports

Verwey, W.B. 1990. Adaptable driver-car interfacing and mental workload: a review of the literature, Soesterberg, TNO Institute for Perception

Verwey, W.B. 1991. Towards guidelines for in-car information management: situation specific load on visual and cognitive processing resources of the driver. Soesterberg, TNO Institute for Perception

Verwey, W.B. 1993. (....). Soesterberg, TNO Institute for Perception

Wiersma, J.W.F., Hale, A.R., Heijer, T., 1991. Kwantitatieve gedragsreferentiemodellen in het verkeer. (in Dutch). Delft: Safety Science Group, Delft University of Technology.

Wiersma, J.W.F., Heijer, T., 1993. Dynamic aspects of motorway traffic safety. Presentation at the First World Congress on Safety of Transportation, Delft, 1992.

Wiersma, J.W.F. in prep. Limitation of truck overtaking; a video experiment. Delft: Safety Science Group, Delft University of Technology.