Interdependence of Certain Operational Characteristics Within a Moving Traffic Stream

PATRICK ATHOL
Expressway Surveillance Project, Illinois Division of Highways

A considerable portion of the data reviewed was automatically recorded on the Congress Street Expressway Surveillance Project's data logging system. Traffic measurements within a 5-mi section of freeway from a total of 38 detectors were logged in a remote office. Examples are shown of the time averaging characteristics of the electronic equipment used.

Lane occupancy, a point measurement comparable with concentration, was compared to volume and speed and the interrelationship was recorded. Lane occupancy has distinct practical advantages in representing the degree of concentration existing within a moving traffic stream. Lane occupancy was used to predict vehicle speeds and its accuracy was compared to measured Doppler speeds. The comparison showed reasonable agreement.

Aerial photographs of the study area were used to compare occupancy and aerial density of traffic. The comparison emphasized the contrast between point measurements and section measurements. Travel times throughout the study area were derived directly from aerial density and the results were compared to measured travel times. The aerial photographic estimates appeared reliable and represented a practical method of recording average travel times of a large sample of vehicles.

Congestion of traffic is discussed, with some consideration given to a possible definition of congestion. Microscopic and macroscopic measures of traffic performance were reviewed as to their value in predicting the advent of congestion. These measures included volume, speed, speed differences, headways, and the variations of individual speeds, speed differences and headways. Further, the lane variations of some of these measures were reviewed. Consideration was given to the possible requirements of a traffic control system which has as its objectives the optimalization of traffic operation by the control of congestion. Reference is made to the needs of traffic studies to determine the response of certain control techniques rather than the confirmed observation of uncontrolled flow. The early prediction of congestion may not be possible from point measurements. However, a system of controls such as ramp metering with known responses on the critical measures of congestion may be initiated at an early enough stage to avoid breakdown. Most of the control responses are probablistic in nature and some degree of breakdown may be expected.

•ONE OF THE research groups studying freeway operations in the Chicago metropolitan area is the Expressway Surveillance Project. This project is charged with developing a pilot automatic information and control system which, it is hoped, will prove beneficial to the operations of a selected highway network system. One of the specific tasks of the project is to investigate the suitability of various operating characteristics as the criterion for the automatic control system.

Paper sponsored by Committee on Characteristics of Traffic Flow and presented at the 43rd Annual Meeting.

FREEWAY TRAFFIC OPERATIONS

The selected study area is a $4\frac{1}{2}$-mi section of the westbound roadway of the Congress Street Expressway. Located within the study section are 38 vehicle detectors which continually transmit measurements of traffic operations to a centrally located office. The information reaching this office is automatically recorded on coded punched paper tape. The punching code selected is compatible with locally available digital computers on which most of the data were analyzed.

The individual outputs of the 38 detectors differ, but the system may record volume, speed, occupancy, headways, and vehicle length. These measurements provide most of the data presented in this paper.

To date, many of the previous studies of the operational characteristics of traffic have been concerned primarily with the measurements per se, and some of these studies stand alone as determinations of operational characteristics of performance. There appears to be an increasing need to coordinate these measurements and to apply them in some practical manner to achieve operational improvements in the traffic performance of highway systems. Emphasis will in part be placed on the practical and the economic considerations of using such measures in a traffic control system.

Volume, Speed, Concentration

Models of traffic behavior are often built within the framework of recorded operational characteristics. The interrelationship of speed, volume and concentration is an excellent example of this (1). By dimensional analysis the relationship may be expressed as volume = speed × concentration or:

$$\frac{N}{T} = \frac{L}{T} \times \frac{N}{L} \qquad (1)$$

The limitation of this relationship is that it applies to a group of vehicles maintaining the same group length and progressing at a uniform speed such that the volume is recorded for the time interval that it takes the group to pass a point. The relationship holds good for the specified group of vehicles and not necessarily for any others. Speed variations may occur provided they do not change the speed of the group. If these conditions are to be satisfied, then the changes of observing such conditions over a 1-mi section of freeway are quite limited.

The speed associated with the group movement is the space mean speed. Most of the manufactured measuring equipment available is only capable of point measurements and the speed measurement is the time mean speed. The measured ratio of volume/(time mean speed) is not truly concentration, but may be called "accumulation" because it has an application in the estimate of the number of vehicles accumulated within a section of highway adjacent to the point of measurement. Whenever a measurement is calculated from the interrelationship of other characteristics, it is strongly recommended that a clear-cut differentiation be made between the measured and the calculated characteristics.

Occupancy

Occupancy is the ratio of the time that a vehicle is present at a detector compared to the time of sampling. If the detector relay denoting the presence of vehicles is closed for 15 seconds in a sampling time of 60 seconds, the occupancy is said to be (15/60) 100 = 25%.

Occupancy is a point measurement, not solely to be considered an estimate of concentration, but a parameter in its own right with certain distinctive properties.

Consider occupancy still further: let N = number of vehicles passing a detector in a given time T, s_i = time speed of i th vehicle, and ℓ_i = length of i th vehicle.

The time duration that the i th vehicle is present at the detector is ℓ_i/s_i; therefore, total time for N vehicles under the detector is $\sum_{i=1}^{N} \frac{\ell_i}{s_i}$.

By definition:

$$\text{occupancy} = \frac{\sum_{i=1}^{N} \frac{\ell_i}{s_i}}{T} \tag{2}$$

Occupancy may be seen to reflect both the performance and the composition of traffic in terms of individual speed and individual vehicle length.

For the condition where all vehicles have length $\bar{\ell}$,

$$\text{occupancy} = \frac{\sum_{i=1}^{N} \frac{\bar{\ell}}{s_i}}{T} = \frac{\bar{\ell}}{T} \sum_{i=1}^{N} \frac{1}{s_i} \tag{3a}$$

$$\text{occupancy} = \frac{\bar{\ell}}{T} \left(\frac{\sum_{i=1}^{N} \frac{1}{s_i}}{N} \right) N \tag{3b}$$

$$\text{occupancy} = \left(\frac{\sum_{i=1}^{N} \ell_i}{N} \right) \left(\frac{\sum_{i=1}^{N} \frac{1}{s_i}}{N} \right) \left(\frac{N}{T} \right) \tag{3c}$$

but the space mean speed,

$$\bar{S}_s = \left(\frac{N}{\sum_{i=1}^{N} \frac{1}{s_i}} \right)$$

and volume $V = \frac{N}{T}$; therefore,

$$\text{occupancy} = \bar{\ell} \times \frac{\text{volume}}{\text{space mean speed}} \tag{4}$$

Let $\frac{1}{\ell}$ = K factor; therefore,

$$\text{concentration} = \text{K occupancy} \tag{5}$$

Assumptions are (a) uniform length, and (b) concentration = volume/space mean speed.

When occupancy is expressed as a percentage and concentration in vehicles per mile K = $5280/\ell$ vpm, where ℓ is the average vehicle length in feet. When using an average vehicle length of 17.6 ft:

$$\text{concentration (veh/mile)} = 3 \text{ occupancy (\%)} \tag{6}$$

There are two general groupings of measurements: those recorded at a point and those recorded within a section. Those belonging to the point measurements include occupancy, spot speed and volume; the interrelationship of these depends on vehicle length. For the section measurements, those included are space mean speed and concentration. Volume to be included should be uniform at both the input and output to the section and the sampling time should be greater than the section travel time.

Aerial photographs provide a direct source of recording the number of vehicles present in the section and although this measure appears to be concentration, the practical limitations of varying volumes and nonuniform concentration may make comparisons with other parameters misleading. It is suggested, therefore, that the term aerial density be retained for such measurements.

Sampling Time

The equipment components of the project's automatic data logging system include variable time averaging devices for each detector output. The mathematics of the averaging follows an exponential pattern which is effected by an electrical circuit with a variable time constant. The average measure is provided by an analogue voltage output. With a time constant of 45 seconds (rate #3), approximately two-thirds of the instantaneous reading is based on the traffic recorded in the last 45 seconds, two-thirds of the remaining is based on the preceding 45 seconds, etc. Thus, the reading at any time has as its component a steadily diminishing effect due to all preceding vehicles. The effect which is due to individual vehicles rapidly diminishes and the sampling time for the average may be considered as approximately twice the time constant.

To consider the effects of these averaging features, comparisons were made on ramp volumes with differing response times. The comparison was made between the analogue reading recorded at each minute of time and the digital count of volume within the minute preceding the analogue reading. The examples cited here record ramp traffic, which is the most variable and accentuates the differences between the analogue and digital measures. Figure 1 shows the fluctuations of the analogue volumes by 10-sec increments at two response times (#3—45 seconds and #5—140 seconds). Table 1

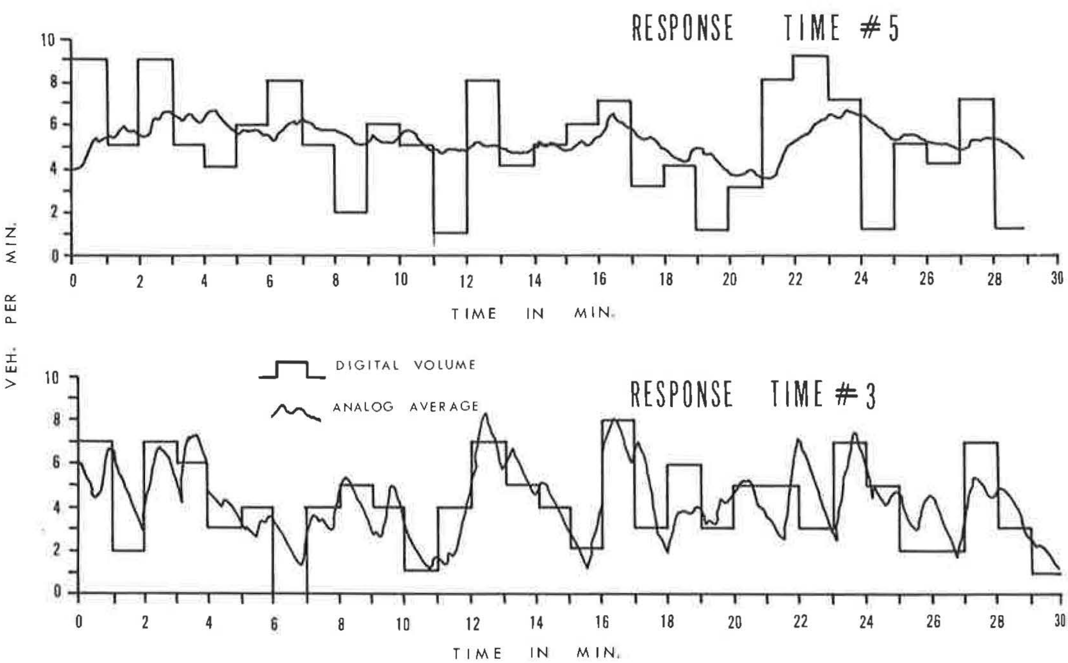


Figure 1. Analogue averaging of ramp volumes.

TABLE 1

ANALOGUE AVERAGING COMPARED TO DIGITAL COUNTS FOR RAMP VOLUMES

Ramp	Response Time (sec)	Analogue Totals (30 min)	Digital Count (30 min)	Diff. (anal.-dig.)	Percent Diff.
Entrance	45	119.9	125	5.1	4.08
	90	79.7	79	- 0.7	0.88
	140	148.2	148	- 0.2	0.14
Exit	45	286.8	298	11.2	3.76
	90	227.0	217	-10.0	4.61
	140	329.1	327	- 2.1	0.64

gives a 30-min comparison of the accuracy of both techniques. In general, the longer the response time, the greater the overall accuracy, but for short periods such as 1 minute, the accuracy may be improved with shorter response times which more closely correspond to the 1-min sampling time.

COMPARISON OF OCCUPANCY WITH OTHER MEASUREMENTS

In general terms when discussing the behavior of parameters, the differences between time mean speed and space mean speed are not critical. For general verbal discussion the terms volume and speed are not critical and in a like manner, concentration may be substituted for a constant times occupancy. Figure 2, giving a generalized traffic operation (2) summary, illustrates the manner in which volume, speed and concentration adjust. In comparing other studies and data it is reasonable to interchange occupancy and concentration where trends are being discussed.

Aerial Density

In an effort to compare point measurements with section measurement, aerial photographs were taken of sections adjacent to the field detector stations. The times of aerial photographs and ground measurements were carefully matched and comparisons

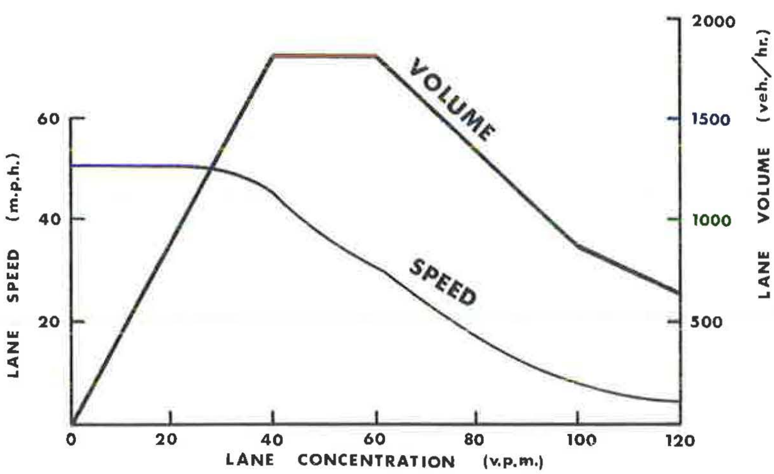


Figure 2. Generalized traffic operations.

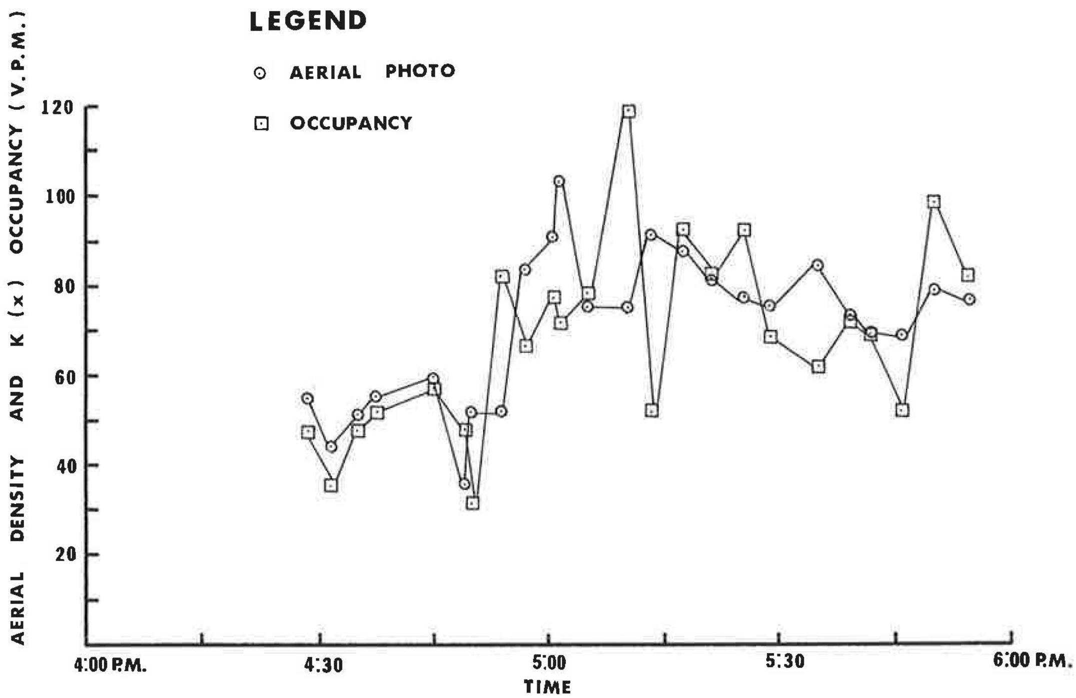


Figure 3. Time variation of aerial density and occupancy.

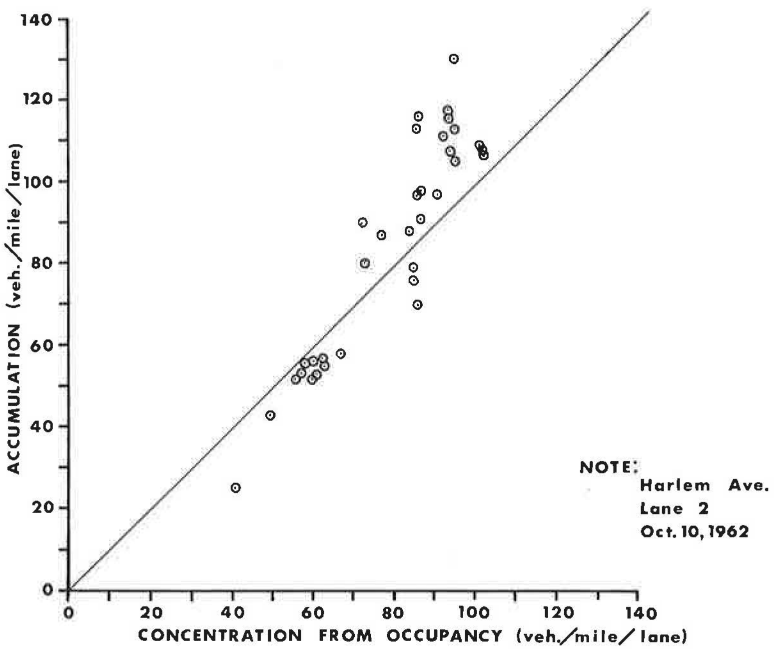


Figure 4. Comparison of accumulation and occupancy.

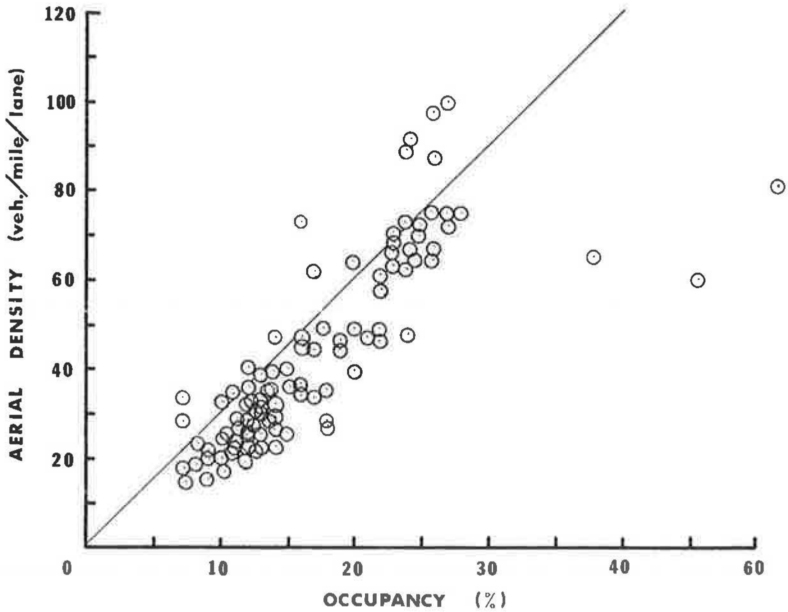


Figure 5. Comparison of aerial density (helicopter) and occupancy.

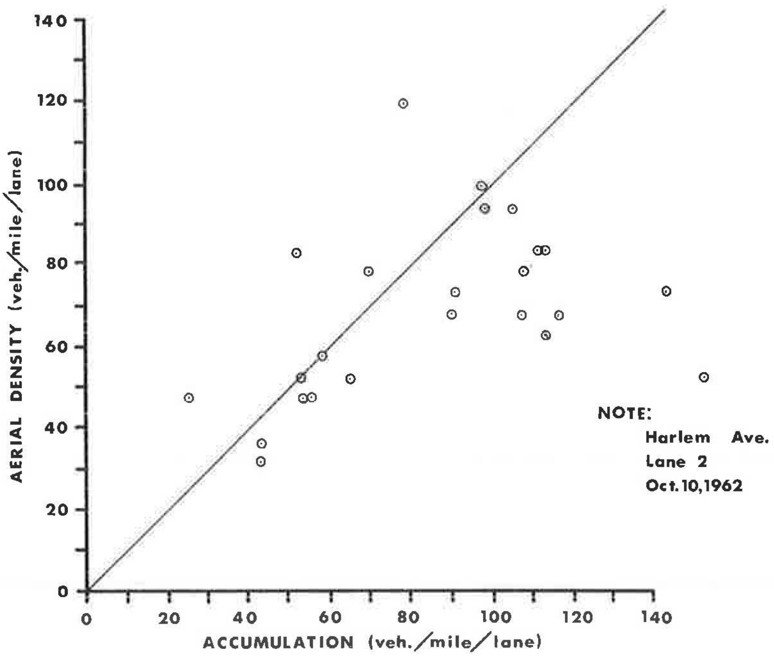


Figure 6. Comparison of accumulation and aerial density (light aircraft).

TABLE 2

COMPARISON OF POINT MEASURES AND SECTION MEASURES OF CONCENTRATION

Measure	Location No.[a]	Mean Diff. (veh/mile/lane)	Std. Deviation of Diff.
Aerial density - Kx occupancy	1	- 0.4	15.3
Aerial density - Kx occupancy	2	- 1.2	13.9
Aerial density - accumulation	1	- 3.1	16.7
Aerial density - accumulation	2	-12.7	20.4
Accumulation - Kx occupancy	1	+ 3.6	3.6
Accumulation - Kx occupancy	2	+11.8	21.3

[a]Location No. 1 Des Plaines, lane 1, Oct. 17, 1963; Location No. 2 Harlem, lane 2, Oct. 10, 1962.

were made of aerial density, occupancy and accumulation. The aerial photographs were collected by two techniques: (a) by flying a light aircraft along the freeway and photographing successive sections, and (b) by taking time lapse photographs from a helicopter circling over the study section (3). The net result was about a 50 percent overlap on successive photographs for the light aircraft and 12-sec intervals with the helicopter. Values of aerial density were only obtained for each pass of the light aircraft, whereas the ground measurements and helicopter photographs were continuously recorded. The light aircraft passed the study section at approximately 3.5-min headways.

Figure 3 illustrates the time variation of aerial density and accumulation while Figures 4, 5 and 6 illustrate numerical comparisons of aerial density, accumulation and concentration derived from occupancy. Occupancy and accumulation show general agreement; however, aerial density shows considerable scatter when compared to accumulation.

These results in the view of the author give weight to the need to differentiate concisely between aerial and ground measurements of the type compared here. Two factors which may contribute to the apparent discrepancies are the high variations of concentration within a section of freeway and the time averaging techniques used in the ground measurements. The variations in concentration may be the result of shock waves or of constrictions within a section, whereas the point measurement may represent the output of such a section and as such may not be subject to the same fluctuations experienced within the section. The differences between aerial density and accumulation appear to increase at concentration levels above 50 vehicles per lane mile. The differences between accumulation and occupancy are probably heavily dependent on the time averaging techniques from which the minute readings are taken.

Table 2 gives the summary statistics on the differences between all the measurements.

Speed

Occupancy may be compared to measured speed in two distinct manners: (a) the generalized traffic operations curve allows speed to be derived directly from occupancy, and (b) the ratio of volume to occupancy can be used to estimate speed. Minute averages of measured time speeds were compared to calculated speed (volume divided by Kx occupancy) and derived speed (S = f (Occ)). The results of this comparison are given in Table 3.

In general, calculated speed (volume and occupancy) shows good agreement with measured speed over the whole speed range 15-50 mph. This point may prove of greater significance in the consideration of a practical control system; speed is measured utilizing the Doppler effect and it is often more difficult and costly to transmit Doppler

TABLE 3

COMPARISON OF MEASURED, CALCULATED AND DERIVED SPEEDS

Reading	Sample	Veh. Speed (mph)	Measured Speed (avg. diff., mph)	Calculated Speed (std. dev. diff., mph)	Measured Speed (avg. diff., mph)	Derived Speed (std. dev. diff., mph)
1	Sample A[a]:	15-50	+1.5	3.4	+1.5	3.5
2		15-50	+4.3	4.5	+3.6	5.2
3		15-50	-1.6	2.9	-0.4	3.2
4		15-50	+0.9	3.3	+2.0	3.3
5		15-50	+0.9	6.1	+2.0	3.1
6		15-50	-3.1	3.8	-0.2	4.2
7	Sample B[b]:	42	+1.8	2.8	-0.6	2.8
8		29	-1.1	2.5	+1.3	3.8
9		40	+1.1	2.8	+0.6	2.7
10		45	-1.1	2.2	+2.3	3.2
11		40	-1.7	1.9	+2.2	3.6
12		25	-0.2	3.5	-1.2	3.8

[a]Ground photography (K = 3.0 each sample 120 min).
[b]Detector readings (K = 2.8 each sample 60 min).

frequency changes to a central location. Further, as traffic flow breaks down and speed is reduced, the accuracy of speed determinations diminishes, whereas the accuracy of occupancy determinations increases. If stoppages within the traffic stream occur at frequent intervals, there may be occasion when occupancy is affected by the relative position of the stopped vehicle to the detector; however, such occurrences have been very uncommon and have not been found critical in the interpretation of high occupancy readings. Stop and go conditions exist only at the higher occupancy levels and variations at the higher levels are more indicative of the degree of congestion than of the prediction of congestion.

Travel Times

Direct measurements of travel time through the study area were determined from two time lapse cameras which photographed vehicles entering and leaving the study area. By identifying individual vehicles at both limits of the section, their respective travel times were recorded. These travel times are shown in Figure 7 and generally indicate the trend of vehicles to slow down at the peak load. The increased variation in travel time at slower speeds is quite noticeable.

In addition to the direct measurement of travel times, there are two other sources of data from which estimates of travel time may be derived: aerial density and volume and occupancy. Volume and occupancy estimates of travel time were based on the sampling of readings on the center lane only. The aerial density readings are summarized by a contour map such as Figure 8, and the travel time is derived by assuming an average speed for each contour interval of aerial density. The average speed for each contour interval is derived from the generalized traffic operations curve. Aerial density and volume and occupancy have been shown to be estimators of speed and by using these speeds it is possible to estimate travel times. The ground photographic travel times were used as the base for comparison. In addition, the U. S. Bureau of Public Roads impedance vehicles were used as a comparable moving vehicle measurement of travel time. The measures of travel times are compared in Figures 9 and 10 which indicate similarity in trends, but do at times display distinct differences.

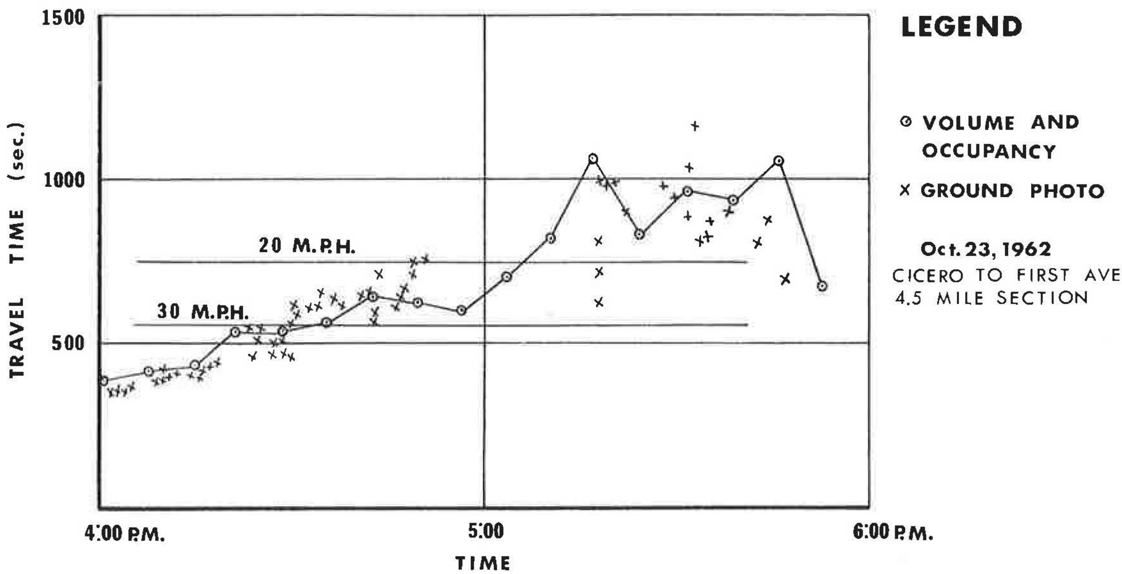


Figure 7. Travel times for ground photography.

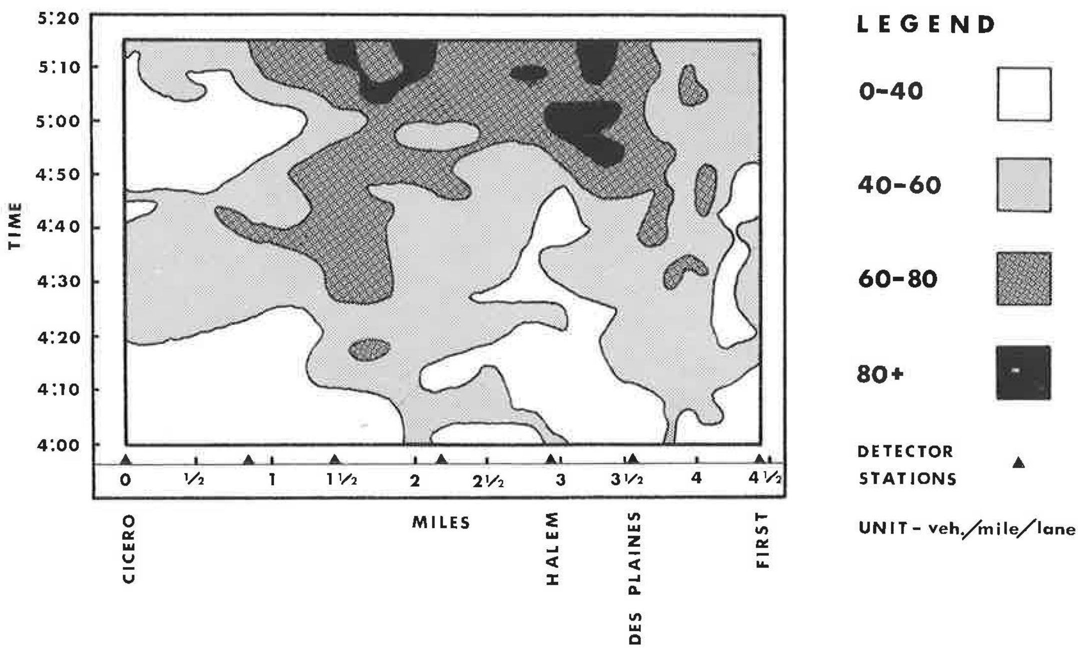


Figure 8. Density contour map.

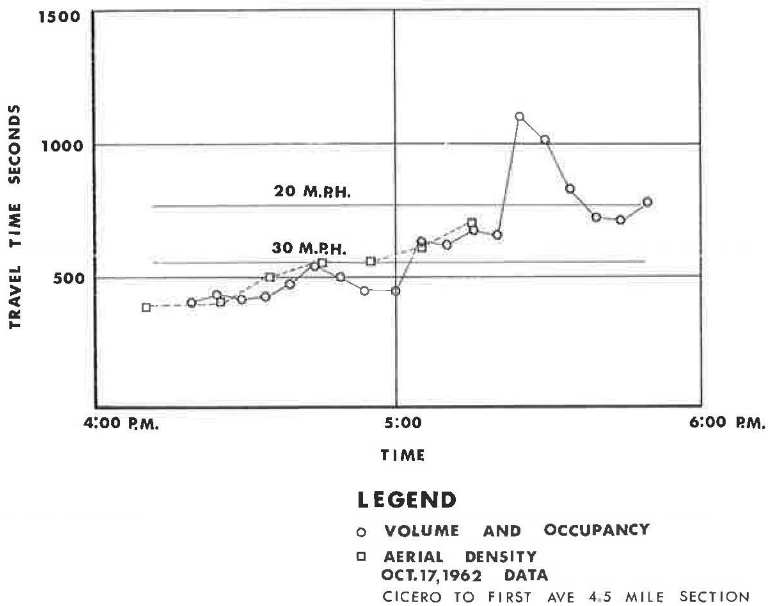


Figure 9. Travel times from aerial density.

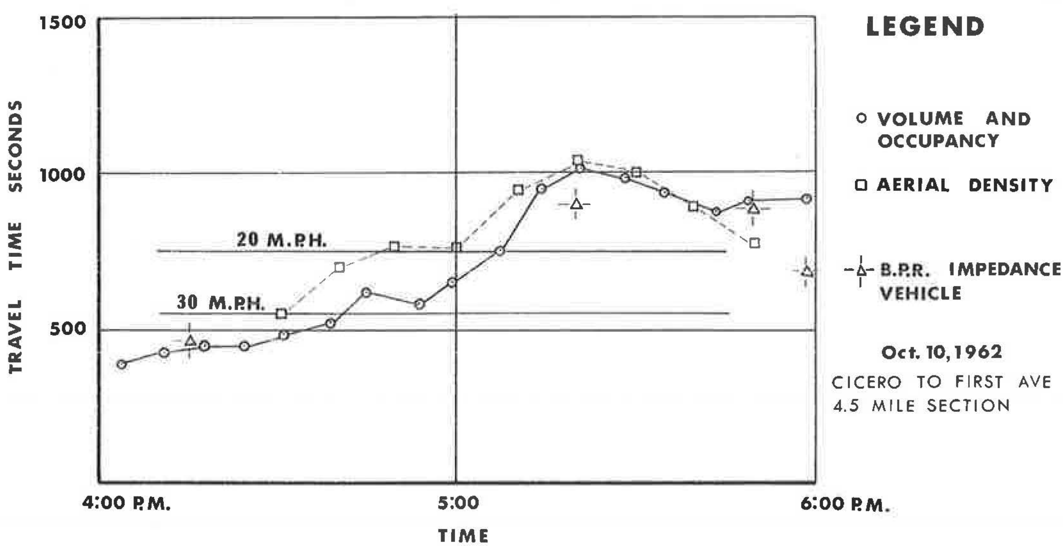


Figure 10. Travel times compared to the U. S. Bureau of Public Roads impedance vehicle.

The large fluctuations in minute determination of travel time at times of low speed suggest that larger samples are needed. From time-lapse photography, sample sizes of 6 or 7 vpm were quite practical at higher speeds; however, once speeds had fallen it was only possible to record 2 or 3 travel times per minute. The individual variations of travel time for a given minute were drastically increased at the lower speeds; changes of the order of 50 percent were common. As an example, at 5:15, October 23, 1963, the recorded times were 869, 505, and 525 seconds, whereas at 4:08 they were 404, 403, 419, 414, 411, 410, 394 and 400 seconds. This fluctuation at lower ranges tends to limit the value of sampling techniques, such as moving vehicles or light aircraft photography; and appears to enhance the value of continuous measurements, such as helicopter aerial photography or electronic detection equipment. The latter two consider all vehicles present in the stream of traffic.

Effect of Trucks

The discussion on occupancy as a parameter noted the influence of vehicle length. Therefore, a first reaction would be to infer that longer vehicles (generally trucks) have a linear effect on occupancy. The K factor appears to be related to average vehicle length. Although this is true from the derivation, it does not imply that it is possible to vary truck percentages totally disregarding the changes which may be induced on speed and volume.

Consider the following as an alternate expression of

$$\text{occupancy} = \frac{(\text{vehicle length}/\text{vehicle speed})}{\text{headway}} \qquad (7)$$

Vehicle length is definitely a factor, but so are headway and speed. Therefore, to consider the effect of trucks it is necessary to review the performance of trucks in terms of their speeds and headway. The project detection equipment (General Railway Signal Company—ES 100—VL vehicle length computer) is only able to classify vehicles by length, and to do this it records the speed of a vehicle and then notes its individual time under the detector. It was not possible to record automatically vehicle length and speed at the same time; the speed is lost in the vehicle length determination.

Headway distributions were recorded at varying speeds, and the data were combined by dividing the individual headways by the mean headway for each minute of data. In this manner some compensation is made for varying volume rates (volume is the reciprocal of mean headway).

The table of headway distribution shows a general trend to more uniform headways at lower speeds. The headway distributions are illustrated in Figure 11 for speeds of 45, 35 and 25 mph. This is a graphical representation of the data included in Table 4 and it should be noted that the first class interval of headway/average headway is twice the normal class interval; consequently the percentage frequency is halved. There appears to be a relatively high frequency of short headways; for the speed range 40-45 the average volume would be approximately 30 vpm, at which level there would be 4 percent of the headways less than 0.8 seconds. At the other limit there would be 11 percent of the headways greater than 3.2 seconds.

Figure 12 shows a sampling of headways for vehicles of differing length. In view of the small sample of vehicles over 25 feet in length, the headway distribution is plotted for groupings of vehicles over 25 feet and for those under 25 feet. Vehicles over 25 feet may be broadly classified as truck traffic. The speed-headway distribution for the same operational speed as that of the vehicle length traffic is compared to the vehicle length-headway distribution. At the location studied, the vehicle length data do not indicate great headway differences between vehicles over and under 25 feet in length. All the headway data collected were for the center lane of a 3-lane section of freeway. The percentage of vehicles longer than 25 feet was approximately 15 percent.

In summary, there does not appear to be a clear-cut quantitative "truck effect," either beneficial or adverse. At any location under review, the individual vehicle

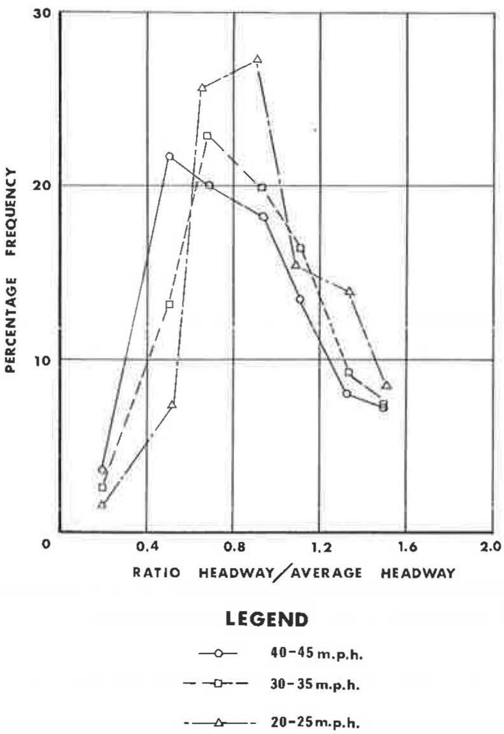


Figure 11. Headway distributions.

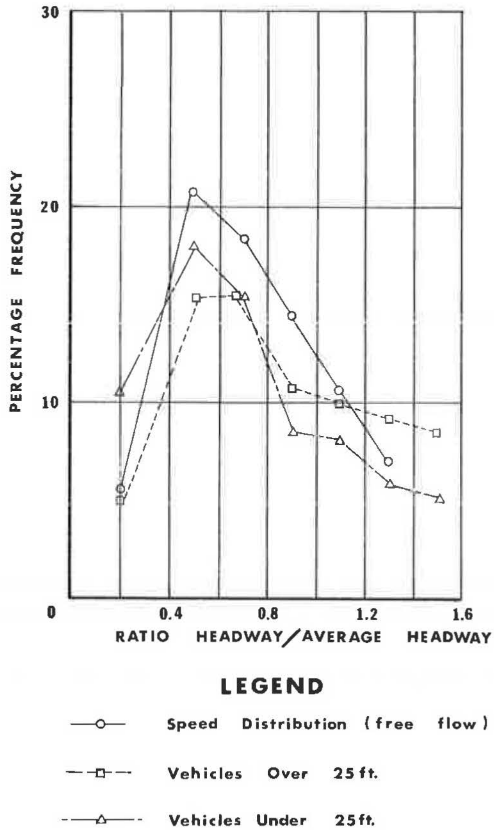


Figure 12. Vehicle length and headway distributions.

TABLE 4

HEADWAY DISTRIBUTION-PERCENTAGE FREQUENCY TABLE

Speed (mph)	Ratio Headway Divided by Average Headway							
	0-0.4	0.4-0.6	0.6-0.8	0.8-1.0	1.0-1.2	1.2-1.4	1.4-1.6	1.6-5.0
55-60	9	16	21	15	7	10	4	18
50-55	10	17	16	14	10	7	5	20
45-50	9	21	21	15	10	8	5	11
40-45	4	21	20	17	13	7	6	11
35-40	3	19	23	20	12	9	7	7
30-35	2	13	23	20	12	9	6	9
25-30	1	9	23	23	21	10	6	6
20-25	1	7	25	27	14	13	7	6

speeds, vehicle lengths and headways are all important factors. It does not appear that K may be modified by a truck percentage figure, but would more likely be modified as follows:

$$K^{1}_{truck} = K \times C_{\ell} \times C_s \times C_{hd} \tag{8}$$

in which

C_{ℓ} = length correction,
C_s = speed correction, and
C_{hd} = headway correction.

This idea of a varying truck effect appears to be far more in keeping with freeway operational studies than a simple truck percentage.

One example might be a long upgrade; the presence of trucks at such a location is likely to be far more critical than on a level tangent section of freeway roadway. The difference between speeds and headways for truck and passenger vehicles would probably be more noticeable where grades are present.

Variations in truck lengths alone may not be too critical when compared to the larger number of passenger cars and their respective length variations. Considering a sample of 30 vehicles in a minute's data, if 27 are 18-ft passenger cars and 3 are 50-ft trucks, the sample mean length would only be about 21 feet.

Detectors

In the discussion of occupancy it was inferred that the vehicle under detection was passing a point or line across the freeway. In practice this is not the case. Each detector type has a differing zone of influence. With the overhead ultrasonic detector, as used in the study, the zone of influence in the line of travel of the vehicle is of the order of 1.5 feet and the pulsing rate of the sound beam is 30-40 times per second. Generally, the zone of influence is effectively added to the vehicle length.

General Railway Signal Company in their equipment compensated, in part, for the length factor by using a combination of pavement and vehicle reflected signals. This length compensation may only be noticeable at lower speeds, such as 20 mph and less. At the higher speeds the pulsing rate is the more critical factor. The pulsing speed may act to increase or decrease the time of detection. A vehicle may enter the zone of detection immediately following one pulse and will not be detected until the second pulse. Similarly, in leaving the zone of influence the time of detection may be extended by one pulse. On the average these time differences probably cancel each other out. For illustrative purposes consideration will be given to a time difference of one pulse and a zone of influence of 1.5 feet.

$$\text{Measured time under the detector per vehicle} = \frac{\text{vehicle length} + 1.5 \text{ feet}}{\text{vehicle speed}} \pm \frac{1}{30} \text{ sec} \tag{9}$$

Total difference between measured and actual time at point locations is (1.5 feet/ vehicle speed) ± 1/30 sec.

Table 5 gives measures of these differences.

In the case of loop detectors only the length factor appears critical. The length of the loop will be a known factor and the difference will be predictable and allowances made. Loop occupancy should be considered on its own merits rather than as a less accurate measure of occupancy. Overhead detectors may be more uniform in the detection of vehicles in that they pick up any vehicle interrupting the sound beam, whereas a loop may vary somewhat in detecting the bumper or leading edge of a vehicle; nonetheless, the loop may more than adequately compensate for this by virtue of its continuous detection features as compared to the discrete detection features of the ultrasonic detectors.

At an operating speed of 40 mph, an ultrasonic detector reading of 15 percent occupancy would correspond to 19.5 percent occupancy with a 6-ft loop detector.

TABLE 5

TIME AND LENGTH DIFFERENCES IN INDIVIDUAL OCCUPANCY MEASURES

Vehicle Speed (mph)	Vehicle Length (ft)	Time at Point (sec)	Diff. due to	
			Length 1.5 Ft	Pulse Time
60	20	0.227	+0.017	±0.033
50	20	0.273	+0.020	±0.033
40	20	0.341	+0.026	±0.033
30	20	0.454	+0.034	±0.033
20	20	0.681	+0.051	±0.033
10	20	1.362	+0.102	±0.033

It may be reasonably assumed that with time, loop detectors will be modified to yield occupancy based on a more narrow zone of influence.

CONGESTION

Congestion is a description generally understood by the motoring public to mean discomfort and inconvenience. In defining congestion the criterion used will attempt to combine the considerations of individual driver comfort and convenience with the efficiency of operations of the highway system.

Congestion may be thought of as overcrowdedness; in other words, crowded to a degree that some detrimental side effects occur. Individual drivers may feel some degree of discomfort at even the presence of other vehicles either within their field of view or within their operational area. Obviously there must be a balance between the extreme individual reaction and the norm for the group, although on occasion the individual may be the critical performer. As the presence of other vehicles increases around an individual his possible actions become more constrained, but as long as these constraints are not overly restrictive it is something that the individual can cope with. It is not until the individual finds his desires frustrated and his actions completely controlled by others that the degree of restraint becomes unbearable. There appears to be a threshold of constraint that the group is willing to endure, but after that threshold is reached a reaction may set in and affect the operation of the highway facility.

Consider then, the buildup of crowdedness on a freeway section. A lone vehicle passing through the section would probably maintain a speed related to the speed limit or physical restraints of the roadway itself. For the time duration that it takes to pass through the section, the concentration would be one vehicle and the volume would be one. Additional vehicles then enter the section; volume increases, concentration increases and the speed remains dependent on the entering vehicle. Subject to minor variations in speed, the buildup of volume necessarily results in increased concentration. The individual driver is directly concerned with concentration rather than volume.

This process of increased demand and increased concentration reaches a critical stage. With higher concentration the degree of constraint on the driver is increased and the threshold level may be reached and difficulties develop. There appears to be a limit to the time duration and speed at which concentration levels are maintained. Thus, instead of input volumes increasing with corresponding increases in concentration, a change in pattern is detected and traffic is unable or unwilling to maintain speed at critical concentration levels. The operational speed diminishes and spacings decrease, resulting in higher concentrations. After the threshold level is attained, increases in concentration are not necessarily accompanied by increase in volume. This situation may be referred to as speed inversion; however, this operation does not necessarily diminish volume.

Capacity Constraint

Capacity in this expression is defined as the volume measured at a bottleneck and is the maximum volume for the sampling time which may pass through the bottleneck (4). The development of congestion under these conditions is simply a traffic demand greater than the capacity. Vehicles arrive at the bottleneck at a volume in excess of the capacity, and storage occurs; concentration may continue to increase to a level at which the output volume from the stored vehicles is less than the capacity. It is a necessary condition, therefore, that arrival volume exceeds the bottleneck capacity, and that the volume from the greater concentration is equal to or lower than the maximum or capacity. The reasoning for the reduction in output volume compared to input volume is that it is a necessary condition to cause storage. (The capacity constraint approach is often reasoned on volume alone.) It is the contention of the author that the capacity restraint is not a necessary condition for congestion to exist. It is hypothesized that congestion may be caused by speed inversion.

Speed Inversion

Speed Inversion refers to the condition at which the freeway traffic rapidly reduces its operating speed at high volumes and is usually in the speed range of 45-25 mph. This speed reduction condition, or inversion, is often referred to as a transitional condition. It will be pointed out that this inversion function may increase concentration to the level of congestion without any necessary capacity constraint.

Consider the analogy of a conveyor belt traveling at a speed corresponding to the speed of traffic (Fig. 13). Vehicles arrive at P_1, at specified time headways and the speed and headways of these vehicles may oscillate in magnitude between P_1 and P_2,

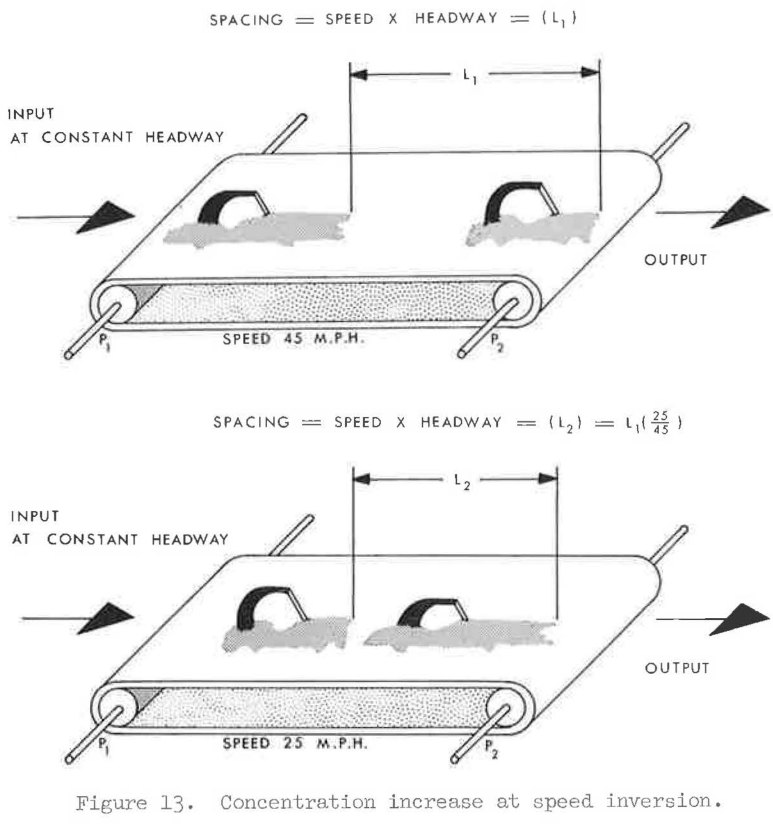


Figure 13. Concentration increase at speed inversion.

but their relative positions at P_1 and P_2 remain fixed. The volume depends on the input headways. It cannot be overstressed that volume measurements should be related to identifiable vehicles. The input and output should be for the same vehicles and not simply a numerical value. Concentration on the belt will be the relationship between the belt speed and the time headway of arrival: concentration = volume/belt speed.

Thus, the conditions are set that with diminished headways at the input there will be a corresponding increase in concentration. The comparable input and output vehicles will be separated in time by the travel time P_1 to P_2. This condition reasonably approximates observed traffic performance.

Consider then, the performance as volume approaches a maximum. The threshold levels are reached and the operator or speed inversion function is effected and congestion is then said to exist. The analogy to this situation is the transfer of vehicles to a slower moving conveyor belt paralleling the initial belt. Vehicles will arrive at P_1 at the same headways and will be transferred with no time difference to the slower belt. For all vehicles preceding the transfer, travel times correspond to the freeflow speeds, while successive vehicles will travel at the slower speed. The last fast vehicle will pass P_2 at a headway compared to the first slow vehicle far in excess of the initial headway condition at P_1. The difference in headway is equal to the difference in travel time between P_1 and P_2 at the two speeds. All sucessive slow vehicles will follow at the same initial headways; the output at P_2 might appear somewhat as in Figure 14.

At congestion the input rate may be sustained and the output rate will equal the input rate except for the period of the time lag occurring as congestion sets in.

This concept is basically quite different from the notion of a restricted capacity and an overloading of the section; it is suggested quite strongly that there need not be a restrictive output of limiting capacity for congestion to set in, but rather that the increase in the concentration is effected simply by the speed change within the section. After the time lag in travel times, the input volumes and output volumes may remain equal to each other and also equal to the maximum. Notice the need to identify vehicles completely in comparing input and output volumes. The volume at P_2 will decrease for a period equal to the lag time, but it hardly appears meaningful to state that the volume has been reduced. There has been a time shift in conditions. Congestion can set in at a very rapid rate, which is compatible with the conveyor belt analogy. In reality it is more probable that vehicles within the section, starting with a critical group of vehicles, reduce speed as a result of concentration levels and that this speed change is reflected back to the input.

If the concentration change at congestion is of the order of 20 vehicles/mile/lane, the time to effect this change would be related to the change in travel time at the differing speed levels. After this time, however, the volumes may revert to the input levels and possibly equal the maximum volume. After the initial change and with sustained input, there need not be any tendency for an increase in concentration, as would appear necessary from the restricted output hypothesis, or a reflection impeding the input to match the output capacity.

The analogy was used to consider a section of freeway free of ramp effects. Witness to the fact that a volume restriction is not necessary to cause congestion is the increased concentration that can be seen on a section of freeway. Concentration on a section may represent congestion and appear as a standing wave such that arriving vehicles joining the queue, and the departing vehicles leaving the queue, have no effect on queue length. The input and output volumes are equal; there simply happens to be a speed change condition resulting in an output time lag.

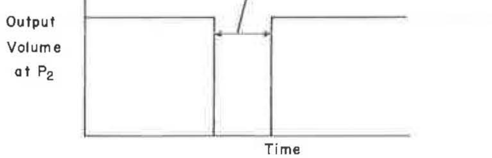


Figure 14.

The analogy used to discuss the setting in of congestion is of further value in the review of the recovery from congestion. In the case of congestion setting in it is easy for a following vehicle to reduce

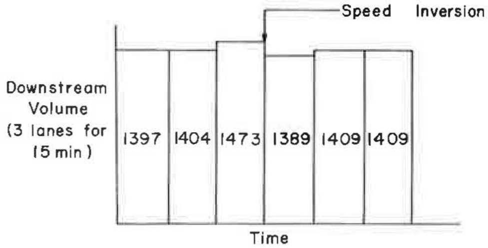


Figure 15.

speed considerably and thereby effect the change in the systems operating speed. When it comes to the reverse, however, a following vehicle cannot accelerate beyond the lead vehicles and this potential conflict is critical in limiting the rate of speed change out of congestion. In this regard it is of concern with a control system that breakdown does not occur, because recovery will be slower than breakdown. It is doubtful that speed inversion is a reversible reaction at maximum volume. In this case the input needs to be reduced.

In summary, the points of emphasis are (a) there need not be a volume (capacity) restriction for congestion to set in; (b) congestion need not impede the input volume, but delays the output volume (input and output volumes should be related for identifiable vehicles); and (c) output from congestion may equal the maximum volume (excluding the time for congestion to set in).

If the speed inversion concept is to be accepted, there are certain characteristics which may be looked for. At the time of inversion there will be a time lag in the output which may be detected as a short time volume reduction at the output boundary. Another condition will be the comparison of the increased number of vehicles within the section attributable to the speed change alone. Another condition to review will be the volume before and after congestion allowing for the time lag between the input and output.

Figure 15 is a sample of recorded data immediately preceding and following speed inversion and shows the apparent volume decrease following speed inversion.

Figure 16 shows the total vehicle count within the study section. The change between 50 and 70 vpm may be attributed to speed changes alone and as can be seen the only storage in excess of the speed change is 115 vehicles or approximately 1.5 minutes volume. This additional storage may be the effect of ramp volumes impeding the mainline traffic. It does not appear conclusive as to which effect this may be attributed to.

Figure 17 compares the volume before and after congestion.

It should be made very clear that speed inversion and capacity restraint are not mutually exclusive. Most operational conditions are probably described individually by capacity restraint or speed inversion or most often, as the two effects in combination. It is hoped that the speed inversion concept may bring together the diverse opinions on the conditions for maximum volume. Maximum volumes may be sustained before and after congestion and the increase in concentration simply results from speed changes. At the greater concentration levels, ramp traffic may become critical in the maximum

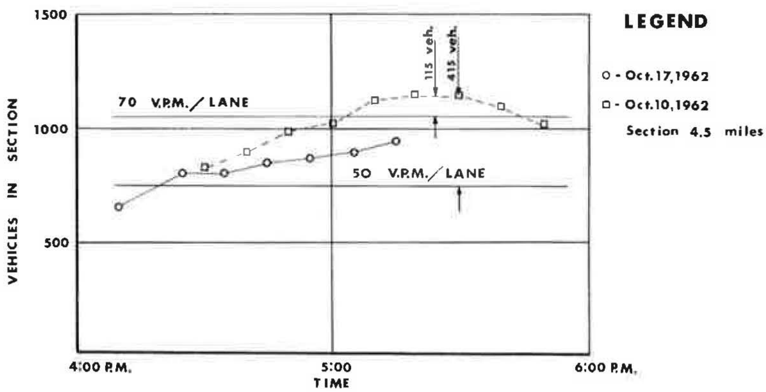


Figure 16. Storage of vehicles within the study section.

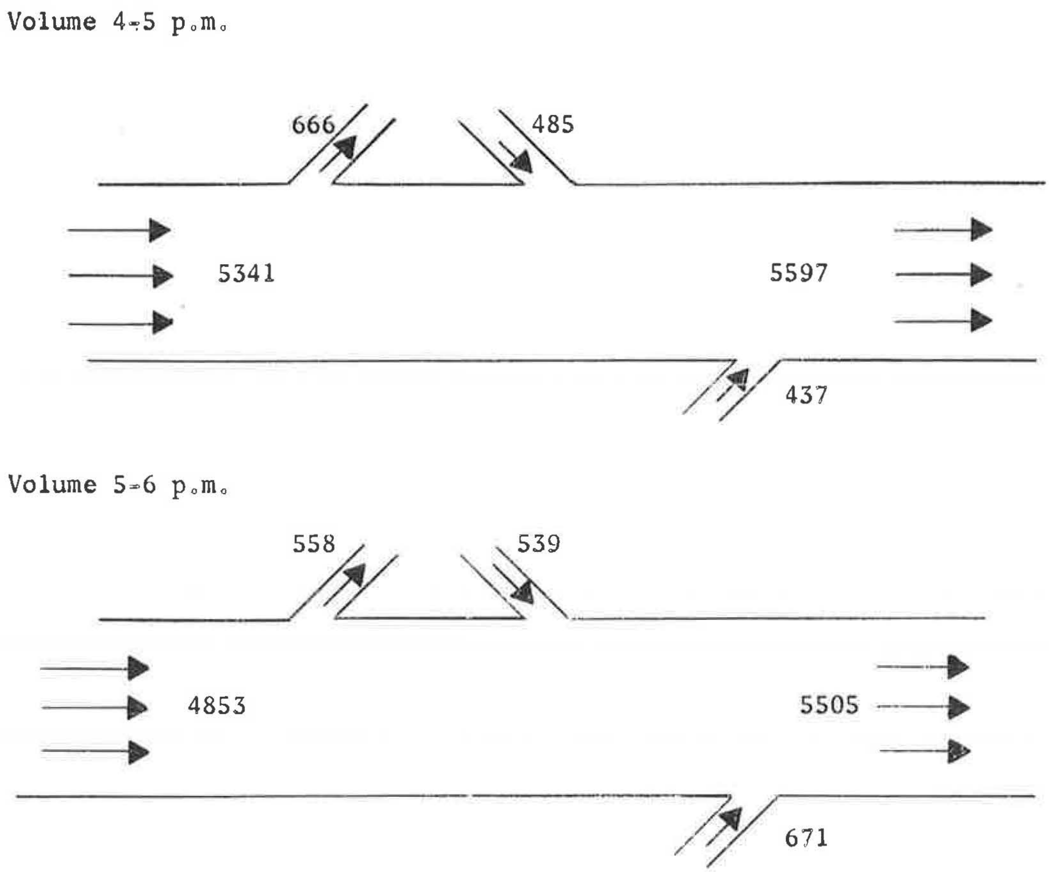


Figure 17. Volume flow map.

volume possible through a section. Figure 17 illustrates the greater proportion that the ramp traffic contributes to the output at the time of congestion. The mainline input may be seen to be reduced at times of congestion simply because the ramp volume impedes the mainline volume input; however, the system concentration may not be radically increased. The ramp volume appears to replace mainline volume without necessarily increasing the concentration on the mainline.

Volume data before and after congestion may be as variable as the tides, and reliance on a few samples should be minimized.

Definition

Congestion is said to exist when speed inversion occurs. Congestion may be considered as the stable condition which exists following a speed reduction of the order of 10 mph occurring at high volume rates.

Congestion does not necessarily reduce volume; it may result in higher volumes. Congestion will be heavily dependent on the drivers existing in the traffic stream. The threshold levels of concentration and their associated speeds will vary with the location and the driver concerned.

Congestion is usually associated with a high rate of change in concentration and a correspondingly small rate of change in volume. As congestion develops, the increase in concentration results in a considerable increase in travel time to the individual motorist, with little or no beneficial effects to the highway system.

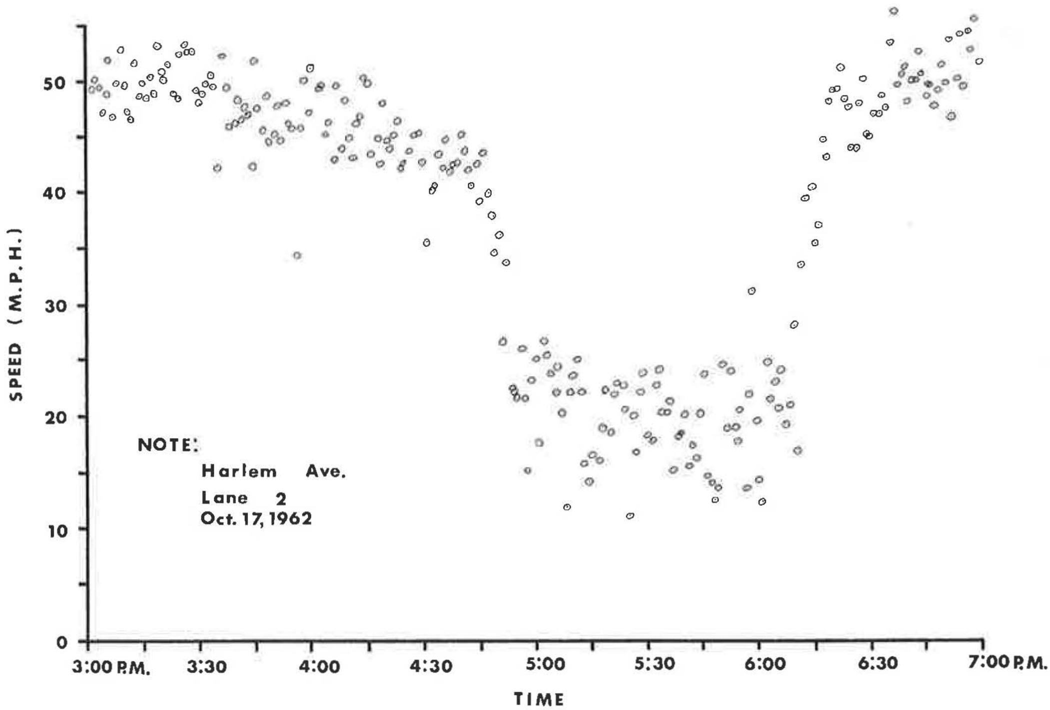


Figure 18. Time variation of speed.

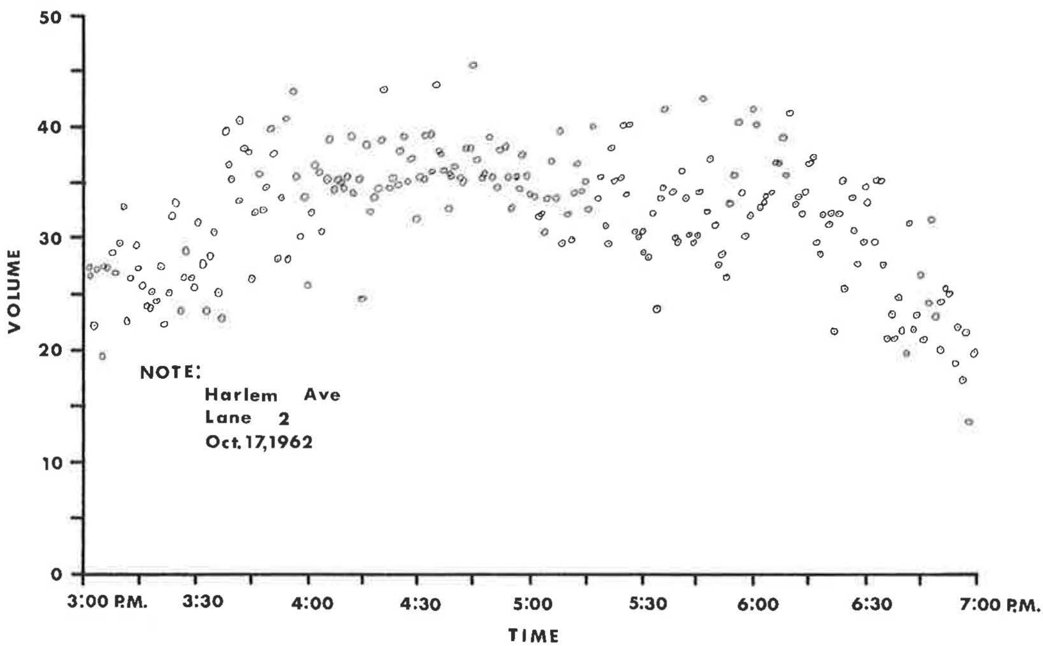


Figure 19. Time variation of volume.

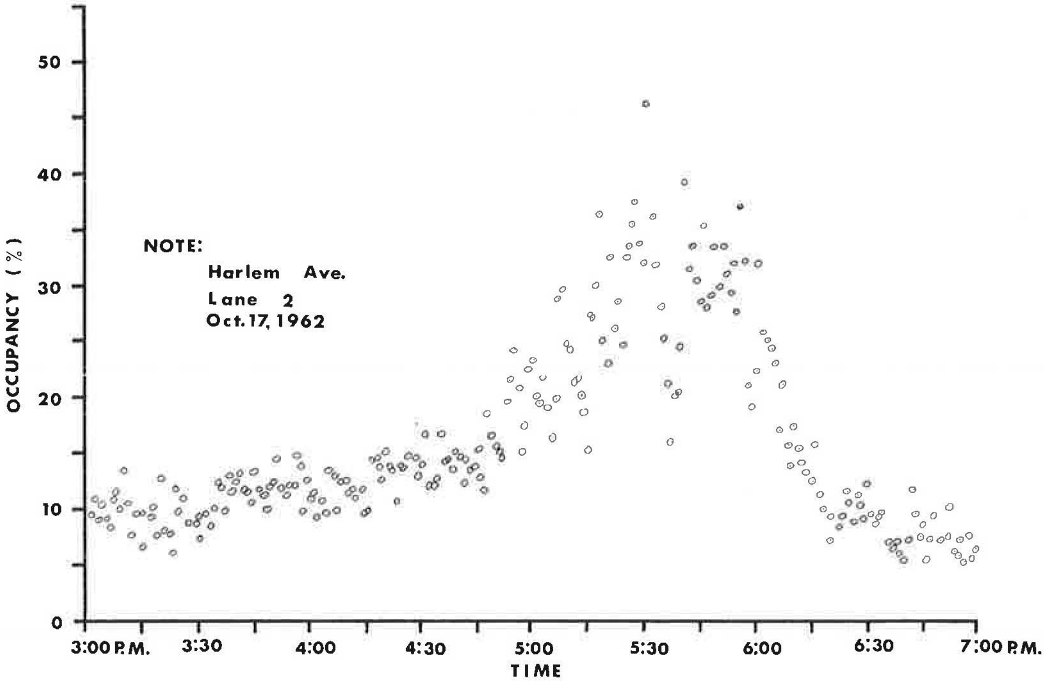


Figure 20. Time variation of occupancy.

Behavior of Characteristics at Breakdown

Figures 18, 19 and 20 are samples of the time variations of speed, volume and occupancy. The most pronounced changes are in speed and occupancy and it may be seen that the apparent change in stable conditions is effected in a relatively short time. Occupancy increases at the time of the speed change (speed inversion) and subsequent to the speed inversion, volume, occupancy, and speed all appear to be in a fairly stable state of oscillation.

The occurrence of speed inversion was noted at a speed (mph) and concentration (vehicle/lane/mile) of almost numerically equal values. This numerical relationship is illustrated in Figure 21 in which the product of volume and speed are plotted against a U factor where U = (Speed/Concentration) - (Concentration/Speed). The product of volume and speed may be considered as a weighting of volume in terms of level of service. The factor U has a value of zero when speed equals concentration and is negative as congestion develops. The data represent the minute averages for 4 hours of traffic. The actual determination of speed over concentration was based on volume and occupancy:

$$\frac{\text{speed}}{\text{concentration}} = \frac{\text{volume}}{(K)^2\,(\text{occupancy})^2} \tag{10}$$

Breakdown occurred at a value of the U factor $\doteq$ -1.0.

Investigation was made of the variation in the lane distribution of volume and occupancy at breakdown. Although considerable data were gathered there has not yet been any determination which has shown predictable trends at breakdown. The lane distribution behavior of characteristics appears to be sensitive to the exact location of detection and as such does not appear to afford a generalization of the variations of lane distribution at the time of breakdown.

TABLE 6

COMPARISON OF SPEED DIFFERENCES AND HEADWAYS

Min No.	Vol.	Average			Standard Deviation		
		Headway	Speed Diff.	Speed	Headway	Speed Diff.	Speed
Sample A:							
1	30	1.9	4.4	45	0.3	3.7	4.1
2	29	2.1	4.3	45	1.4	3.1	3.7
3	32	1.9	5.2	41	0.4	3.7	6.1
4	36	1.7	4.8	45	0.6	3.4	4.1
5	36	1.7	3.1	37	0.8	2.3	3.7
6	31	1.9	4.4	39	0.3	2.5	4.7
7	32	1.9	4.2	46	0.4	3.2	5.0
8	24	2.5	4.6	47	1.9	3.0	3.8
9	23	2.6	5.8	48	1.7	3.4	4.6
10	36	1.6	4.6	42	0.5	3.7	4.2
11	33	1.8	3.1	41	0.5	1.8	2.3
12	36	1.6	3.1	38	0.6	2.0	3.1
13	29	2.1	4.3	42	0.6	3.0	4.7
14	30	1.9	4.1	44	0.2	3.2	4.5
15	33	1.8	4.0	45	0.6	2.3	4.5
16	34	1.7	5.6	44	0.9	5.3	6.8
17	28	2.2	3.5	43	0.8	2.8	4.2
18	27	2.2	5.5	48	0.6	3.4	5.2
19	31	1.8	2.6	37	0.4	2.0	2.6
Sample B:							
1	30	2.0	2.6	30	0.7	2.3	2.8
2	34	1.7	3.1	24	0.8	3.3	5.2
3	25	2.5	2.3	26	0.7	1.7	2.7
4	29	2.0	2.0	27	0.8	0.8	2.5
5	32	1.8	3.6	28	0.7	3.6	3.9
6	30	2.0	1.8	31	0.7	1.0	1.3
7	26	2.2	2.7	24	0.4	3.1	4.3
8	28	2.1	1.9	22	0.7	1.0	3.6
9	30	1.9	2.4	23	0.8	1.2	6.0
10	31	2.0	2.5	30	0.5	1.3	2.1
11	33	1.8	2.2	30	0.7	0.9	3.0
12	33	1.8	2.1	30	0.7	0.9	1.7
13	31	1.9	2.5	29	0.5	1.6	2.5
14	32	1.9	2.9	26	0.6	3.1	3.2
15	33	1.8	1.8	27	0.8	1.1	3.3
16	30	2.0	2.6	31	0.3	1.4	3.1
17	28	2.1	1.9	27	0.6	1.1	2.0
18	31	1.8	1.8	20	0.7	1.5	2.4
19	26	2.3	2.5	23	0.7	2.8	3.2
20	31	2.0	2.6	32	0.8	1.1	2.6

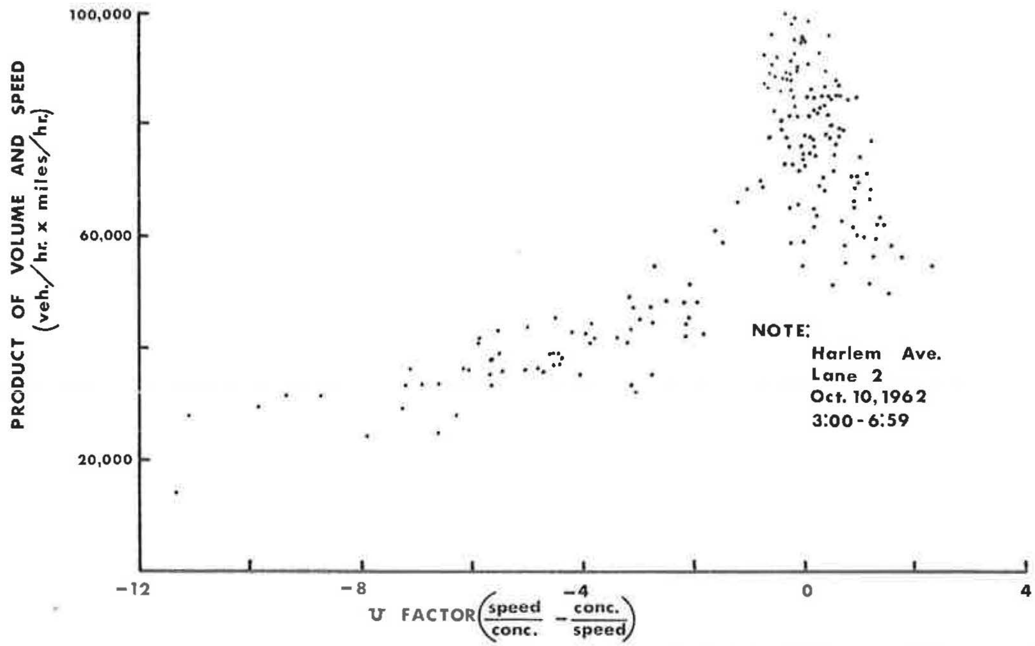


Figure 21. Comparison of U factor with product of speed and volume.

Single-lane operations were reviewed in terms of headways and individual speed data. Upon the determination of individual speeds and headways the data were summarized to include average speed differences, average speeds, and the standard deviations of headways, speeds, and speed differences. A sample summary of these data is given in Table 6.

In view of the relative complexity of the calculations, these measures did not immediately appear to have any distinct advantage over other types such as those of average speed; therefore, further evaluation was limited at this time.

All the previously mentioned measurements were taken at isolated locations. However, as the data were reviewed, it became clear that at all point determinations of traffic operations, consideration should also be given to the location of the sampling point relative to the nearest critical sections. The interpretation of operating characteristics at a detector station requires a knowledge of the upstream and downstream conditions existing at the time of measurement. Inasmuch as the traffic pattern may stabilize itself (for example, by time of day), it may be possible to estimate breakdown, but this disregards variations in surrounding conditions. It becomes a game of chance and may not be suited to control.

Optimal Freeway Operations

There are two general considerations which become apparent when attempting to obtain the optimal level of freeway operations. The system may be considered as a group operation in which the criterion will be set to serve the most vehicles regardless of individual performance. The alternate choice is to provide a high level of service. One of the first tasks is to determine to what extent level of service is compatible with providing service to the largest number of vehicles.

The largest number of vehicles served may be equated with volume or volume weighted in terms of vehicle-miles. This enters directly into the controversy concerning conditions at which maximum volume is attained. The level of service measure may be equated to speed and concentration. Speed is indicative of the time factor and concentration is indicative of the comfort factor.

The attempts at defining congestion in this paper appear to offer a definite choice between performance and level of service. Congestion is said to follow speed inversion which represents a sharp drop in the operational speed of the traffic stream; therefore, from a level of service viewpoint, a control system should strive to avoid congestion. If congestion is to be avoided, what volumes may be expected, and how do these volumes compare to the maximum?

In general, volumes before and after speed inversion do not differ a great deal; however, it is the opinion of the author that maximum volume may occur at differing speed levels depending on the individual vehicles in the group and the locations involved. It is suggested, however, that the relative change in speeds at congestion are far more significant than the corresponding changes in volume before and after congestion. Under such conditions, therefore, it is proposed that the control criterion be the maximization of volume without causing speed inversion and the resulting congestion.

Tables 7 and 8 give speeds and concentrations at which, on a 3-lane section, counts of more than 100 vpm were recorded in a ground photography study. The speed range was 25-55 mph and the concentration (for short section lengths) was 30-80 vpm. These data appear to support the idea that maximum volume does not necessarily occur at a

TABLE 7

FREQUENCY TABLE OF 3-LANE VOLUMES EXCEEDING 100 VEHICLES/MINUTE COMPARED TO AVERAGE SPEED

Average Speed (mph)	Location						Total
	1	2	3	4	5	6	
>50	-	-	2	-	-	-	2
45-50	4	4	4	3	-	9	24
40-45	5	5	11	7	3	6	37
35-40	2	1	18	3	5	1	30
30-35	6	8	2	-	4	-	20
<30	6	6	1	-	1	-	14
Total	23	24	38	13	13	16	

TABLE 8

FREQUENCY TABLE OF 3-LANE VOLUMES EXCEEDING 100 VEHICLES/MINUTE COMPARED TO CONCENTRATION

Concentration (veh/mi)	Location						Total
	1	2	3	4	5	6	
0-20	-	-	-	-	-	-	-
20-30	-	-	-	-	-	-	-
30-40	-	-	3	6	-	4	13
40-50	3	6	10	7	2	12	40
50-60	6	7	24	-	7	-	44
60-70	6	7	1	-	4	-	18
70-80	7	4	-	-	-	-	11
80-90	1	-	-	-	-	-	1
Total	23	24	38	13	13	16	

unique set of conditions. Equally, these values were for 1-min counts and do not generally represent longer time periods of sampling. The range between the 15 and 85 percentiles was 30 to 50 mph, 40 to 70 vpm and 14 to 25 percent where K = 2.8.

It should be made quite clear, however, that the control criteria do not necessarily maximize volume alone; it is recognized that maximum downstream volumes from a ramp-mainline configuration could, under certain conditions, exceed the maximum volume possible by the exclusion of congestion. It may be expected, however, that control will sustain the maximum noncongested flow volumes and will have an advantage over the congested flow condition in that congested flow may tend to deteriorate in volume as the degree of congestion is increased by such effects as ramp turbulence and merging difficulties.

If the input demand on a section of freeway was at a speed greater than congestion, and congestion developed within the section, then the section acted as an operator on the input to incur a tradeoff between speed and concentration. This function of the section in modifying the input has little or no advantage to the system and only increases the individual travel time. Therefore, to optimize the system the objective would be to prevent the operator function of the section on the input. This could be achieved by stipulating uniform concentration throughout the section. The continuity of traffic flow as suggested by Barker (5) may be compared to the detection of the operator function.

The input and output fluctuation should be relatively free from modification; however, this may not necessarily apply to volume. But it does not exclude the variations in input concentration which may be expected to traverse the section.

Thus, it may be proposed that the optimal operating conditions are those of uniform concentration at the threshold level of congestion.

PREDICTION OF CONGESTION

Prediction of congestion becomes a critical determination for the operation of a control system. A balance quickly arises between the predictive aspects of measurements and the response times of the control system. Observations of individual measurements such as speed, volume and occupancy at times preceding breakdown in flow appear to change from one state to another at rapid rates. Such speedy changes may occur from varying preceding levels and achieve the change in state at varying speed. The predictive nature of these measurements appears to be associated with probabilities of failure. Thus, for a given set of conditions there may be an associated probability of failure. Such a system may be set up with certain confidence limits; however, the time to failure generally exceeds the response time of a control system (e.g., ramp metering).

One more practical approach appears to be a control system initiated at an earlier time and exercising control throughout the time period of buildup to breakdown: an approach in which the predictive aspects are covered by varying degrees of control exercised before failure or imminent breakdown. This stage of initial control allows a traffic response pattern to be developed for each control device. If this approach of initial control is accepted, the whole question of prediction becomes rephrased. Study of uncontrolled behavior is less important and the control responses become the critical dimensions in the prevention or lessening of congestion. Individual measurements of speed, volume and occupancy may not provide a workable control system based on critical levels of these characteristics for an "on-off" control system. Nevertheless, these characteristics may prove adequate in initiating the controls and may also serve as the measure of effectiveness of the controls.

Volume Occupancy

The condition set forth as the control criterion was uniformity of concentration at the threshold level of congestion. In practical terms occupancy would be used for concentration. More detailed studies were made of the volume and occupancy interrelationship (Fig. 22). There is a pronounced linear relationship at the lower occupancy levels which agrees with the concept of small speed changes before congestion. Breakdown may occur at less than the maximum volume. This feature makes one critical value of occupancy suspect in a one-parameter control system. It may be noted that deviations from this linear section are perhaps the best indicator of trouble. This corresponds to the rate of speed change. This same measure was developed in the review of the

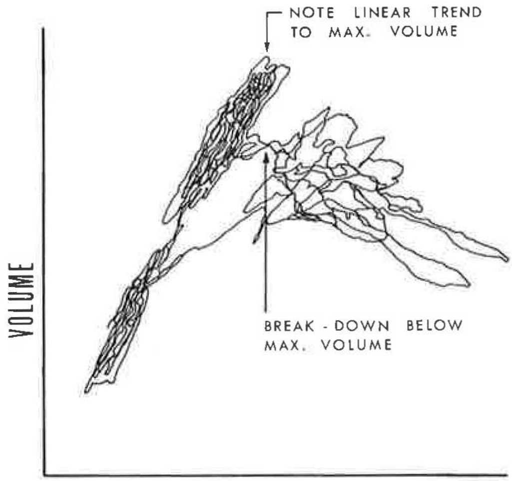


Figure 22. X-Y plot of occupancy and volume.

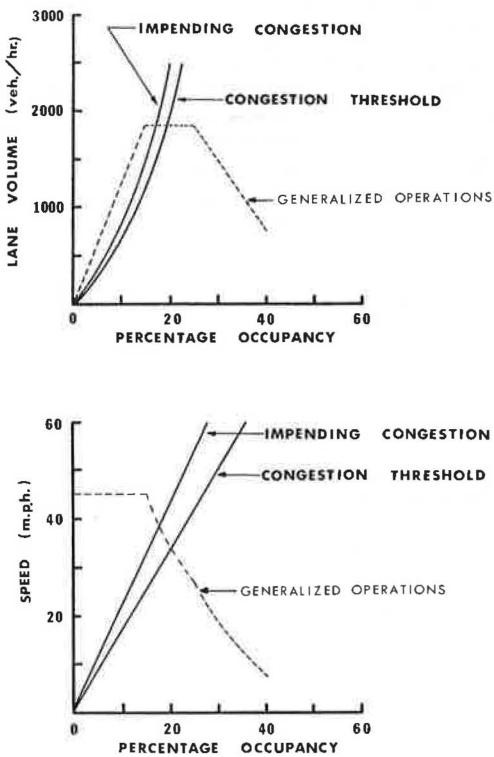
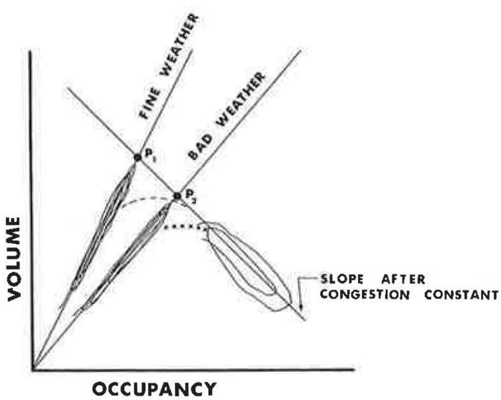


Figure 23. Prediction of congestion derived from the U factor.



P_1 INTERCEPT OF FINE WEATHER CONDITIONS AND CONGESTION

P_2 INTERCEPT OF BAD WEATHER CONDITIONS AND CONGESTION

Figure 24. Effects of weather on congestion.

numerical ratios of speed and concentration, described as the U factor. Impending congestion corresponds to negative values of the U factor not smaller than minus one, and congestion corresponds to values of the U factor smaller than minus one. The congestion level varies with location. Figure 23 shows a method of predicting congestion at a point from a volume-occupancy graph. The volume-occupancy graphs shown are for uncontrolled conditions. In the case of a control system of known response characteristics, point measurements of volume and occupancy are of considerable value.

Under varying weather conditions the slope of the volume-occupancy curve may be even more significant. High volume rates may not be reached and the maximum volume may occur at differing values of occupancy. Under such conditions a graph such as Figure 24 might be useful in adjusting the control criteria.

One consideration which suggests that early control might have practical advantages is the apparent stability of the traffic flow stream. The stability of traffic appears to be greatest before and after congestion. If a form of control is implemented within this stable time period, it may smooth out traffic fluctuations and tighten up control measures each time that the flow approaches the more critical levels.

Ramp Control

At a critical congestion point such as an entrance ramp, there would be two sources of traffic which might cause a deterioration in the performance of the traffic stream; namely, the ramp traffic itself and the mainline traffic. Considerations of control on the ramp through the physical nature of the problem would suggest the measurement of volume. Thus, in an early warning control system, the fluctuation in the ramp volume would be smoothed out. (Volume is chosen for its expediency.) There may be considerable variation in performance, inherent in measuring volume and especially for small sample size, depending on both the vehicle type and the manner of merging of the individual driver. In terms of ramp control for small volumes (approximately 10 vpm), the presence of an apprehensive driver within the stream might be significant at a critical location. The measure of volume could not indicate the presence of this driver.

There is then the need to measure the effects of the ramp volume entering the freeway. The technique should be a qualitative measure, such as occupancy or speed. In other words, the control system, allowing a certain ramp volume, must then know to what extent this volume has affected the traffic on the mainline. Furthermore, the mainline may be considered as the result of output of a preceding ramp and mainline situation. Control of the mainline input at a critical location, may depend on ramp control at a preceding location. A situation may even arise in which all ramp traffic at the critical location would be handled smoothly and only at a preceding ramp, where alternate routes or other facilities were available, would restriction of flow be incurred. Such a control concept necessarily involves specific details, and generalized discussion is less fruitful.

In another situation, considering the frequency of control at a certain ramp and the fluctuation in control, it may be advantageous to close either an entrance or an exit ramp on a full-time basis. The resulting revised travel pattern might lend itself to a more stable control system. Again, such discussion needs to be carefully reviewed on the merits of a specific system. A system of individual local controls may necessarily become an overall system of control. The output of the preceding controller becomes the input of the succeeding controller, and the more stable the situation becomes, the more possibility there is of producing a large overall control system satisfied by a series of local controllers.

Modifications to such a system may be carried out by weighting the ability of each ramp to handle excessive traffic demands. Certain ramps will have a far greater ability to handle a small ratio of volume served over volume demand capability ratio than other ramps. Thus, successive local controllers may be modified depending on this capability ratio. If the capability ratio drops drastically, this may be conveyed readily to the preceding ramp in the form of a command to diminish the mainline input. This would, of course, appear to be a system development; however, much of this would be based on physical configuration rather than traffic measurements. There would in fact be a traffic demand for such action; however, the action would be predetermined on pre-established criteria. A local controller, therefore, would attempt to stabilize an existing situation and would modify the mainline input only when the capability ratio diminishes below certain values. Obviously this passing to the preceding ramp may in turn result in a reflection all the way upstream of the system. Such an occurrence need not be frowned upon; it may, in fact, be quite a desirable situation. From a review of operations there may appear to be certain advantages in controls of a more permanent nature (e.g., ramp closure to change the pattern of the system).

The development of prediction from a point removed in both time and space requires a considerable degree of sophistication to know the conditions existing between the two locations so that a realistic prediction may be made. There is need in a practical system to weigh the accuracy of prediction against the effectiveness of local control. The greater the degree of stabilization within the system, the better its chances of success. As in most cases, the economics of the situation will play a most important role in the selection of a control system.

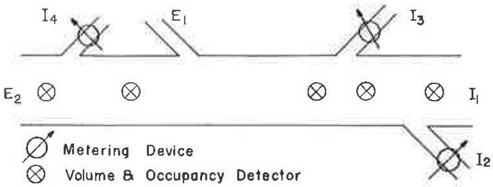


Figure 25.

Local Control Situation

Figure 25 is an example of the control of a system of three entrance ramps and one exit ramp. The possible ramp controls may be assumed to be at I_2, I_3 and I_4. The other measurements may be assumed to be uncontrolled. The criterion for control was uniform concentration at the threshold of congestion. This prevents the section acting as an operator inverting speed and also implies maximizing volume up to congestion.

Considering the volume aspects, the output of the system may be considered as E_2 + E_1 and the output may be related to input provided there is no storage within the section in excess of the initial conditions; i.e., continuity of concentration.

The input may be equated to the output

$$E_2 + E_1 = I_1 + I_2 + I_3 + I_4 \tag{11}$$

To maximize $E_2 + E_1$ is equivalent to maximizing $I_1 + I_2 + I_3 + I_4$. However, I_1 is the uncontrolled input to the system and the control may be approximated by simply maximizing $I_2 + I_3 + I_4$ which corresponds to the metering of ramp volume.

The constraints on maximizing the ramp volumes are (a) that the occupancy within the section should not increase without a corresponding increase in input, and (b) that the level of occupancy would remain below a certain critical level. Another necessary condition is that the input, I_1, should not be impeded by vehicles within the study section and would represent the demand on the system resulting from the preceding section.

The minimal detection system would therefore record ramp volumes and sample occupancy within the study area. The ramp volume would be controlled so as to maximize input without causing levels of occupancy to exceed the critical; however, as a real time control problem it may not be possible to maximize volume, but rather to control ramp volumes to maintain occupancy levels. The record of volumes would be reviewed after the period of control.

The setting of the critical occupancy levels may be subject to varying local conditions, including local driver behavior, geometric design and weather. Further, the setting of a critical level incurs with it a probability of failure. On the occasions at which the system may exceed its critical levels, the control system needs to be sufficiently corrective as to restore satisfactory conditions within a short time period. If the corrective response time is short, there may be an advantage in loading the freeway to the breakpoint at frequent intervals and then allowing the system to recover.

The degree of refinement of control appears limitless; however, from the standpoint of economic benefits, the return for additional refinement may be less impressive.

In considering a traffic control system it is soon surmised that from a simple volume input-output measuring system both volume and concentration may be rapidly determined. The economics of volume detection do not appear at this time to provide a practical system. Although volume detector accuracy is of a high order close to 100 percent for vehicles passing within the zone of influence, the accuracy of a multilane count station is more nearly 98 percent when recorded under all traffic conditions. The 98 percent accuracy is inadequate for a straight forward input-output count measure of storage for periods of several hours.

The development of an economical and accurate detector count station could radically revise the scope of practical freeway control systems.

SUMMARY COMMENTS

In detailed studies of traffic characteristics for short sampling periods there is a definite need to state concisely any assumptions made concerning the interrelationship of the characteristics, and at the same time to explain the time sampling effects on such measures and interrelationships. The analogue averaging technique used for some of the data collected in this paper was not designed for discrete minute sampling, but as a continuous average.

The comparison of occupancy with other measurements generally showed good agreement. The relationship between speed and occupancy was perhaps the most consistent, showing agreement over the 15-50 mph speed range. The relationship between concentration and occupancy showed agreement in the range of 0-60 vpm, but at higher levels of concentration this uniformity of concentration within a section diminishes and the point of detection becomes critical. At higher concentration levels, the measure of occupancy does not diminish in accuracy; rather, the point variation of concentration increases and becomes less representative of the section measurement.

By sampling the center lane volume and occupancy of a 3-lane section of freeway, reasonable estimates may be made of the travel time through the section; this travel time gives a measure of the level of service existing within the freeway section.

Speed-headway distributions were recorded and adjusted to compensate for the effects of the same volumes occurring at differing speed levels. The speed-headway distributions were skewed and the skewness varied with speed; at higher speeds the skewness was greater.

As the interest in freeway control develops, one of the most critical needs is for the collection of data which measures the traffic response to various control devices or techniques. There may be serious limitations in the use of uncontrolled traffic characteristics in the prediction of the effects of control measures.

Congestion, as defined in this paper, included stop and go traffic; in addition, it included traffic operations at speeds as high as 30 mph. The inclusion of speeds at the 30-mph range is suggested by grouping stop and go conditions with any conditions deemed to be inefficient and labeling these as congestion. For a control system it is suggested that the prime function is to avoid the condition described as speed inversion. Speed inversion refers to the condition where freeway traffic undergoes a rapid reduction in speed of the order of 10-15 mph without a corresponding change in volume. This condition rapidly increases concentration and may account for the largest single cause of vehicle storage within a section of freeway.

DEFINITION OF TERMS

Accumulation.—Ratio of measured volume over measured speed—number of vehicles per unit length of roadway.

Aerial Density.—Number of vehicles per unit length of roadway, actual count of vehicles within a section.

Capability Ratio.—Relates to the ramp metering of volume and is equal to the ratio of volume served (metering rate) over volume demand (includes storage).

Concentration.—Number of vehicles per unit length of roadway.

Derived Speed.—Speeds derived from the generalized traffic operations curve of speed and occupancy.

Headways.—Time separation of the leading edges of successive vehicles.

K Factor.—A conversion multiplier which relates occupancy and concentration; concentration (vpm) = K occupancy (%) for a sampling time of one minute, a practical K value = 2.8.

Occupancy.—Ratio of the time a vehicle is at a detector over the time of sampling, expressed as percentage.

Spacing.—Distance separation of the leading edges of successive vehicles.

Space Mean Speed.—The speed corresponding to arithmetic mean of the elasped time of vehicles traversing a measured distance.

Speed Inversion.—Reduction in the speed of traffic which occurs at high volumes without a corresponding change in volume.

Time Mean Speed.—The arithmetic mean of the measured speeds of vehicles passing a speed measuring device.

REFERENCES

1. Wardrop, J. G. Some Theoretical Aspects of Road Traffic Research. Proc. Instn. Div. Engrs., Vol. 1, No. 2, p. 325, 1952.
2. Athol, Patrick. Expressway Surveillance Project Report No. 2: Traffic Flow Analysis. Unpublished, Nov. 1961.
3. Parker, William. Expressway Surveillance Project Report No. 3: Equipment Evaluation. Unpublished, Nov. 1961.
4. Wattleworth, Joseph. Expressway Surveillance Project Report No. 11: Peak Period Control of a Freeway System—Some Theoretical Considerations. Aug. 1963.
5. Barker, J. L. Determination of Discontinuities in Traffic Flow as a Factor in Freeway Operational Control. Traffic Engineering, Vol. 32, No. 2, p. 11, 1961.
6. Haikalis, G., and Hyman, J. Economic Evaluation of Traffic Networks. Highway Research Board Bull. 306, pp. 39-63, 1961.