TRANSPORTATION RESEARCH

RECORD

No. 1394

Highway Operations, Capacity, and Traffic Control

Freeway Operations and High-Occupancy Vehicle Systems

A peer-reviewed publication of the Transportation Research Board

TRANSPORTATION RESEARCH BOARD
NATIONAL RESEARCH COUNCIL

NATIONAL ACADEMY PRESS
WASHINGTON, D.C. 1993

Transportation Research Record 1394
Price: $23.00

Subscriber Category
IVA highway operations, capacity, and traffic control

TRB Publications Staff
Director of Reports and Editorial Services: Nancy A. Ackerman
Senior Editor: Naomi C. Kassabian
Associate Editor: Alison G. Tobias
Assistant Editors: Luanne Crayton, Norman Solomon, Susan E. G. Brown
Graphics Specialist: Terri Wayne
Office Manager: Phyllis D. Barber
Senior Production Coordinator: Sharada Gilkey
Senior Production Assistant: Betty L. Hawkins

Printed in the United States of America

Library of Congress Cataloging-in-Publication Data
National Research Council. Transportation Research Board.

Freeway operations and high-occupancy vehicle systems / Transportation Research Board. National Research Council.
p. cm.—(Transportation research record, ISSN 0361-1981; no. 1394)
"A peer-reviewed publication of the Transportation Research Board."
ISBN 0-309-05463-X
1. High occupancy vehicle lanes—United States—Congresses. 2. Express highways—United States—Congresses. I. National Research Council (U.S.). Transportation Research Board. II. Series.
TE7.H5 no. 1394
[HE336.B8]
388 s—dc20
[388.4′13143] 93-29634
CIP

Sponsorship of Transportation Research Record 1394

GROUP 3—OPERATION, SAFETY, AND MAINTENANCE OF TRANSPORTATION FACILITIES
Chairman: Jerome W. Hall, University of New Mexico

Facilities and Operations Section
Chairman: Jack L. Kay, JHK & Associates

Committee on Freeway Operations
Chairman: Ronald C. Sonntag, Wisconsin Department of Transportation
Secretary: Jeffrey A. Lindley, Federal Highway Administration
B. Beukers, Peter M. Briglia, Jr., Donald G. Capelle, Glen C. Carlson, Robert F. Dale, Conrad L. Dudek, Walter M. Dunn, Jr., Jack L. Kay, Peter R. Korpal, Walter H. Kraft, Steven Z. Levine, Louis E. Lipp, Robert E. Maki, Joseph M. McDermott, Nancy L. Nihan, Colin A. Rayman, James R. Robinson, David H. Roper, James F. Shea, Henry B. Wall III, Thomas C. Werner, Sam Yagar, Michael J. Zezeski, Philip Zove

GROUP 5—INTERGROUP RESOURCES AND ISSUES
Chairman: Patricia F. Waller, University of Michigan

Committee on High-Occupancy Vehicle Systems
Chairman: Donald G. Capelle, Parsons Brinckerhoff
Secretary: Dennis L. Christiansen, Texas A&M University
David E. Barnhart, John W. Billheimer, John Bonsall, Donald J. Emerson, Charles Fuhs, Alan T. Gonseth, Leslie N. Jacobson, William A. Kennedy, Theodore C. Knappen, James R. Lightbody, Timothy J. Lomax, Adolf May, Jr., Jonathan David McDade, C. J. O'Connell, R. L. Pierce, Lew W. Pratsch, Morris J. Rothenberg, Sheldon G. Strickland, Gary K. Trietsch, Katherine F. Turnbull, Carole B. Valentine, Jon Williams

Richard A. Cunard, Transportation Research Board staff

Sponsorship is indicated by a footnote at the end of each paper. The organizational units, officers, and members are as of December 31, 1992.

Transportation Research Record 1394

Contents

Foreword

The papers in this volume were presented at the 1993 Annual Meeting of the Transportation Research Board and are related by their focus on issues related to freeway operations and high-occupancy vehicle (HOV) systems. A wide range of problems reflecting the concerns of both theoreticians and practitioners is addressed.

The specific areas of traffic operations discussed in this volume are receiving considerable attention as a result of the emphasis on intelligent vehicle-highway systems (IVHS), provisions of the Intermodal Surface Transportation Efficiency Act of 1991 (ISTEA), implications of the Clean Air Act Amendments of 1990, and recognition of the importance of incident management as a means of reducing nonrecurring traffic congestion.

Readers with a specific interest in freeway operations will find papers pertaining to algorithms for incident detection using data from freeway traffic management systems, traffic-responsive ramp metering strategies, and public-private partnerships for developing freeway traffic management projects with IVHS applications.

The remaining several papers are focused on HOV systems. Topics include impacts of changes in vehicle occupancy requirement levels, traffic assignment procedures for highways with HOV lanes, HOV lane operation and safety effects, providing HOV priority in a suburban arterial environment, and evaluation of an interim HOV lane in the Seattle area.

The papers presented in this Record were sponsored by the TRB Committees on Freeway Operations and High-Occupancy Vehicle Systems.

On-Line Testing of the McMaster Incident Detection Algorithm Under Recurrent Congestion

Fred L. Hall, Yong Shi, and George Atala

The work reported here represents an elaboration of the logic for incident detection identified in previous work conducted at McMaster University. The improved incident detection logic has gone through three levels of testing; data from the Freeway Traffic Management System on the Queen Elizabeth Way in Ontario were used. An improved logic that could recognize and then ignore recurrent congestion and that could identify incidents that occurred within recurrent congestion was developed and tested off-line. The data used for this stage of the work consisted of 39 days from early summer 1990. The results were sufficiently promising that the algorithm was then installed on-line, and its results were reported to a file instead of to the system operator. Following a period of initial testing and revision to the algorithm and parameters, a major on-line test was conducted during 64 normal weekdays from March 12 to June 18, 1992. The algorithm detected 19 of 28 incidents, a 68 percent success rate. For the 19 incidents, the algorithm time to detection averaged 2.1 min after the time recorded in the operator's log; the median time to detection was 1 min later than for the operator. The false alarm experience was 20 in the 64 days of test, or one in every 6.4 operator shifts.

The purpose of this paper is to present the results of extensive testing of an idea for incident detection developed by Gall (*1*), a feasibility test of which was reported by Gall and Hall (*2*). Gall used a method to identify the cause of congestion by identifying the nature of flows downstream of a traffic queue as suggested by Wattleworth and Berry (*3*). This idea also formed the foundation for the California comparative algorithm (*4*). Gall's contribution was to frame the idea in terms of the congestion-detection logic suggested by Persaud and Hall (*5*). However, Gall was not able to test her idea for distinguishing between recurrent and incident-caused congestion on more than a few days of data. To properly show that it is a feasible incident-detection method, a more extensive test is needed.

There has been considerable previous work on incident detection on freeways, such as that by Payne (*6*), Dudek and Messer (*7*), Cook and Cleveland (*8*), and Levin and Krause (*9*). Stephanedes et al. (*10*) documented some of the difficulties of these existing algorithms in terms of the trade-offs between false alarms and detection rates. It is tempting to use the results from their work for a direct comparison with the results of the study reported in this paper, but because differences in the data bases may be important, caution is needed in the comparison. These differences will be discussed subsequently.

Department of Civil Engineering, McMaster University, Hamilton, Ontario, Canada L8S 4K1.

Three criteria are used to evaluate the algorithm: detection rate, false alarm rate, and time to detection. The detection rate is defined as the percentage of operator-identified incidents with an effect on traffic that were detected by the algorithm. False alarms are defined to be alarms recorded by the algorithm that do not correspond to an incident in the operator's log. The false alarm rate is the number of false alarms divided by the number of decisions made by the algorithm. That rate is calculated as the number of stations involved in the test multiplied by the number of time intervals per day in the test multiplied by the number of days in the test. This definition is consistent with earlier work (*4,9*), and with that used by Stephanedes et al. (*10*).

The first section of this paper describes the logic of the algorithm with reference to Gall's ideas. The second section describes the nature of the data used for the testing. The third section presents the results of off-line testing of the new version, on 39 days of data from the Queen Elizabeth Way (QEW) Freeway Traffic Management System (FTMS). Following successful off-line testing, the algorithm was implemented on-line, in the background. (Results were written to a file instead of being sent to the operators.) The final section of the paper reports the results of 64 days of on-line background testing, following some modifications to the algorithm and parameters.

THE LOGIC FOR INCIDENT DETECTION

Gall's idea was expressed in the form of a template drawn on a flow-occupancy diagram, defining four different states for traffic. Her initial template has been modified to create two templates, depending on the location of the detector station with respect to recurring bottlenecks such as those caused by heavily used entrance ramps (*11*). The template for a normal station, away from ramps, is shown in Figure 1; that for a station affected by recurrent congestion is in Figure 2.

For the stations not affected by recurrent congestion (Figure 1) the template is composed of 4 areas, which are divided by the lower bound of uncongested data (LUD), the critical occupancy (Ocrit), and the critical volume (Vcrit). Area 1, above the LUD, is uncongested data. The area below LUD and Vcrit and to the left of Ocrit is Area 2, one type of congested traffic operation. The area to the right of Ocrit and below Vcrit is Area 3, more heavily-congested traffic operation compared with the data in Area 4, which is below LUD

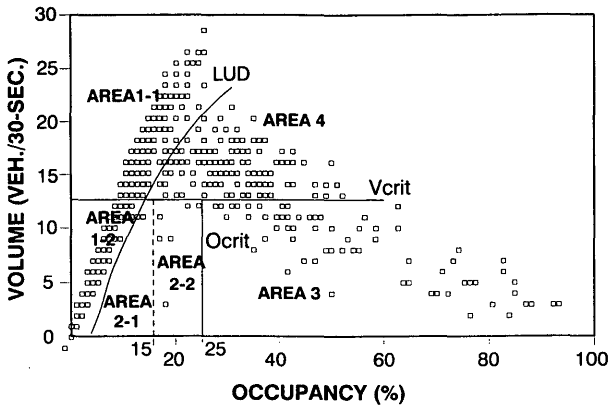

FIGURE 1 Template showing typical parameters for a normal station.

but above Vcrit. The division between Areas 3 and 4 is used for detecting incidents within congestion. Area 1 is further divided into Areas 1-1 and 1-2 by Vcrit, and Area 2 is divided into Areas 2-1 and 2-2 by Ocrit-10. These sub-areas are also used for detecting incidents within congestion.

For the template at the stations affected by recurrent congestion (Figure 2), the only difference from the template just described is that Area 4 represents queue discharge flow (QDF), the flow generated by recurrent bottlenecks. QDF is divided from other congested data by the lower bound of queue discharge flow (LQDF) and by a constant volume (labeled Qconst). Although Qconst is shown in Figure 2 as equal to Vcrit, they may take on different values.

Calibration procedures to establish the parameters displayed in these two figures have undergone considerable development since the methods described by Persaud et al. (*12*). The most important change is in the procedure for identifying the LUD line. That paper described a procedure based on subtracting a constant value from a quadratic function fit through the uncongested data by means of regression. Experience with more data has shown that the uncongested data do not display constant variance for volume as a function of occupancy. This means first that one of the fundamental assumptions of regression analysis is not met, and second that subtracting a constant

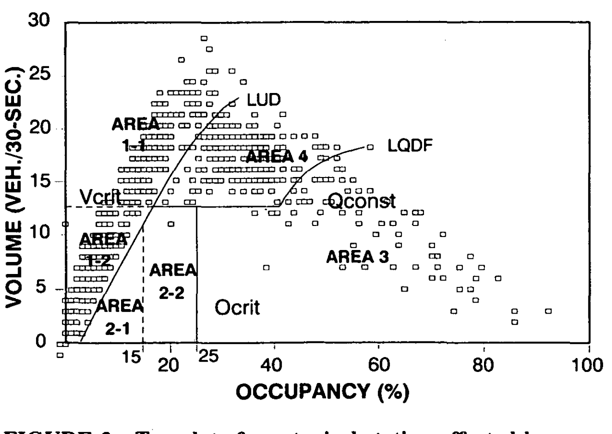

FIGURE 2 Template for a typical station affected by recurrent congestion.

value from the regression function will not reflect the correct location of the boundary of the data.

As a result, LUD is now fit directly to the boundary. The procedure is to start with a standard quadratic function, and then to plot the function against the data. Visual inspection shows if any aspect of the curve needs to be modified (intercept, slope, curvature), and the relevant coefficient for the equation is then adjusted. Although this procedure is not easily automated, experience with it can be gained quite quickly, and then complete calibration of a station may be accomplished in less than 2 hr, including acquiring the necessary data from tape, and when necessary separating the uncongested from the congested data.

The other parameters are also manually identified. Ocrit is simply the occupancy at which the highest observed volume occurs. Vcrit and Qconst may be harder to establish. To properly identify them requires some congested data (which not all stations have), and it must be possible to identify the volume that is normal within recurrent congestion, as opposed to that which occurs only during capacity reductions.

The raw data received from the loop detectors are compared with the appropriate template values. Instead of smoothing the data, a persistence check is used, such that the same state needs to be maintained for a certain number of intervals for a change of condition to be identified. For most of the testing, this persistence check has been set at three (30-sec) intervals.

Any data falling below LUD for longer than three intervals are considered to be congested. The algorithm then attempts to identify the cause. If the cause can be identified as one of two categories, then the congestion is not considered to be from an incident. If the cause cannot be placed in one of these categories, then the congestion is deemed to be caused by an incident. The simplest of the two nonincident categories is secondary congestion, representing the extension of primary congestion to the next station in sequence, either further upstream as a consequence of queue growth, or further downstream when queue discharge effects carry on further than expected.

The other main type of nonincident congestion is recurrent congestion in the vicinity of an entrance ramp. On the basis of experience with the QEW system and data, three stations in the vicinity of each entrance ramp have been defined as stations where recurrent congestion is a possibility: the first station upstream of an entrance ramp and the first two stations downstream of the ramp in the bottleneck. The downstream stations are included because at the time that a queue forms upstream of the entrance ramp, speeds in the queue discharge downstream from the ramp decrease considerably, with the result that the data fall below LUD (although they remain above the volume Qconst shown in Figure 1). Table 1 presents the combination of template states that may occur at the station being checked and at the downstream station and the resulting decision by the algorithm about incident presence. This table also identifies the method for distinguishing incidents from recurrent congestion at stations where recurrent congestion might first be seen, immediately upstream or downstream of an entrance ramp. For the first station immediately upstream of a bottleneck section, if the volume-occupancy data are in Area 2-2, Area 3, or Area 4 of the template (Figure 1), and the data at the downstream station

TABLE 1 Assessment Procedure for Stations Where Recurrent Congestion May Occur

		STATION BEING CHECKED					
	VOL-OCC AREA*	1-1	1-2	2-1	2-2	3	4
DOWNSTREAM STATION	1-1	NO CONGESTION	NO CONGESTION	CONGESTION	CONGESTION	CONGESTION	CONGESTION
	1-2	NO CONGESTION	NO CONGESTION	CONGESTION	INCIDENT	INCIDENT	INCIDENT
	2-1	NO CONGESTION	NO CONGESTION	CONGESTION	INCIDENT	INCIDENT	INCIDENT
	2-2	NO CONGESTION	NO CONGESTION	CONGESTION	INCIDENT	INCIDENT	INCIDENT
	3	NO CONGESTION	NO CONGESTION	CONGESTION	CONGESTION	CONGESTION	CONGESTION
	4	NO CONGESTION	NO CONGESTION	CONGESTION	CONGESTION	CONGESTION	CONGESTION

* See Figures 1 and 2.

are in Area 1-2 or Area 2 (Figure 2), it is likely that an incident happened between the two stations. When there is no incident, the downstream data can be expected to be above QDF. If entrance ramp volume is high enough, however, the downstream data may be above QDF even after an incident has occurred, in which case the incident may not be detected.

It was also expected that the algorithm would be able to detect incidents that occurred within congestion. The logic for detecting an incident within congestion is that if the volume-occupancy data at Station i are in Area 2-2 or 3 while the data at Station $i+1$ are in Area 1-2 or 2, the algorithm will declare an incident at Station i. There are two main differences between detecting incidents from recurrent congestion and incidents that occurred within congestion. The first is that Area 3 in the template may be quite different, as seen in a comparison of Figures 1 and 2. The second is that data in Area 4 at the station being checked will not lead to an incident declaration. Thus Table 1 would be modified for this situation such that all categories under Area 4 would read "congestion," instead of some being "incident."

DESCRIPTION OF THE DATA FOR THE TESTS

Both the off-line and on-line testing of the algorithm were done with data from the FTMS on the QEW west of Toronto. The relevant parts of the FTMS are shown in Figure 3. The road is three lanes in each direction from the western limit until just after Station 25, where it becomes four lanes. Queues regularly form at the entrance ramps from Mississauga Road, Highway 10, and Cawthra Road. Ramp metering is used at these three interchanges, and at the next two west as well, although these other two do not often form congestion on the expressway.

The section from Erin Mills Parkway to the east side of Dixie Road was chosen for the off-line testing because of the daily recurrent congestion there. This section covers 15 eastbound detector stations (from 11 to 25) and is 8.8 km long. The vehicle detectors are installed in each lane at roughly 800-m intervals. Traffic volumes, occupancies, and speeds are recorded for these stations every 30 sec, 24 hr per day. For the on-line testing, the application was extended to cover Stations 6 through 27.

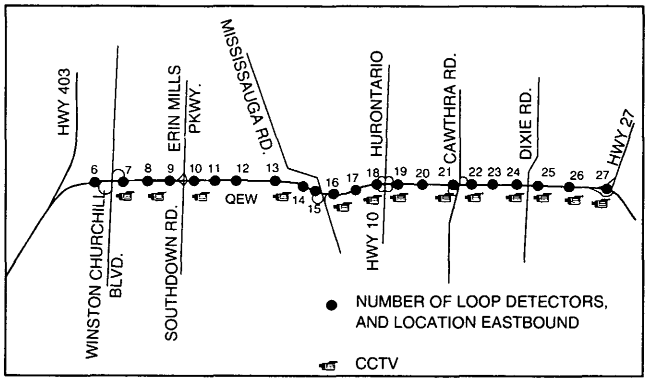

FIGURE 3 Configuration of the QEW FTMS (modified from map by Ministry of Transportation, Ontario).

The results from the algorithm are evaluated against the operator's log for both the off-line and on-line tests. During both test periods the FTMS was operated 16 hr per day, from 6:00 a.m. to 10:00 p.m. Hence the algorithm output can be evaluated for those hours only because there is no record against which to compare the algorithm between 10:00 p.m. and 6:00 a.m.

During these tests no automatic incident detection occurred at the QEW. The on-line test was carried on in the background (i.e., the results were written to a file, to be matched subsequently against the operator's log) instead of being reported directly to the operator. Operators relied on closed-circuit television (CCTV) and routine patrols of the Ontario Provincial Police for incident detection. As on most FTMSs that use CCTV, the remote-controlled cameras swivel almost 360 degrees, and tilt and zoom as well. With these features, and the placement of the 18 cameras (as shown in Figure 3), there is complete coverage of this section of the QEW by CCTV.

Many of the incidents noted on the operator's log are vehicles on the shoulder of the roadway. A critical question in the evaluation of the algorithm results is whether vehicles on the shoulder should be included in the data base of "incidents." For example, for the 39 days of the off-line test, there were 152 events noted in the operator's log, of which only 28 occurred on the traveled lanes of the QEW. If the instances of vehicles on the shoulder affected traffic, then they should be included in the base for calculating the detection rate. If the vehicles on the shoulder did not affect traffic, they should be excluded because any algorithm can operate only on the basis of "observable changes in the traffic flow" [as Stephanedes et al. (*10*) wrote in excluding two incidents that caused no impact on traffic].

The operators report that it is unusual for a vehicle on the shoulder to cause sufficient disruption to traffic flow that there would be any noticeable effect on the data. In early testing, the authors examined the raw data for all missed events, including the vehicles on the shoulder. This inspection supported the operators' impressions: few of the vehicles on the shoulder affected traffic operations. Consequently almost all of the "vehicle on shoulder" events have been excluded from the evaluation of the algorithm.

OFF-LINE TESTING

Thirty-nine days of data from the Mississauga FTMS were used to test the algorithm off-line, from the period May 15 to July 15, 1990. The 39 days cover almost all weekdays during this 2-month period. Five weekdays were not included because the operator's log was not available at the time when the off-line testing was conducted. Some inclement weather conditions, such as a heavy storm, are included in the test. For the 39 days of off-line testing, the operator's log shows 28 incidents on the traveled lanes of the QEW. Twenty-nine incidents were declared by the algorithm during the times the operators were on duty. Fifteen incidents were detected by both the operator and the algorithm, 14 detected only by the algorithm, and 13 on the operator's log that were missed by the algorithm. Complete details of this testing appear in work by Shi (*11*).

Of the 15 incidents that appear in both the operator's log and the output of the algorithm, 2 were detected at the same time by both the operator and the algorithm; 3 incidents were detected 1 or 2 min earlier by the algorithm; and the remaining 10 incidents were detected 1 to 12 min later by the algorithm. For the 15 matched incidents, the mean time to detection was 2.2 min later for the algorithm (including the 12-min delay, or 1.5 min without it). The off-line tests were run with a persistence check of three (30-sec) intervals. Because the first interval is always needed for the congestion to appear, this three-interval persistence check has the potential to add 1 min of delay to the detection time. Any remaining delay in detection must be systemic (i.e., it takes that amount of time for the congestion to move upstream from the point of the incident to the next closest detector station).

The 14 incidents detected only by the algorithm are deemed to be false alarms. Dividing 14 by the 1,123,200 decisions made by the algorithm (16 hr per day, 60 min/hr, 2 intervals per min, or 1,920 intervals per day, per station, for 39 days at 15 stations) gives a false alarm rate of 0.0012 percent. There were problems with bad or missing data in the off-line test set, so one could estimate more conservatively that only 70 percent of those potential decisions actually were made by the algorithm, which would yield a false alarm rate of 0.0018 percent.

Of the 13 incidents recorded only by the operators, 3 were missed because the data were not collected around the incident location during the time the incidents happened. One incident identified by the operators as occurring east of Dixie Road probably happened beyond the test limits because there is no evidence for it in the data at Station 25 (which is the only station in this test east of Dixie Road). These four incidents should not be considered incidents for which the algorithm made an error because the data were not present for the test. For the remaining nine incidents recorded by the operator but missed by the algorithm, seven have relatively high volumes, from 8 to 21 vehicles per 30 sec and speeds from 75 to 120 km/hr at those stations with speed data. Hence these seven incidents had only a slight impact on traffic, if any. The remaining two incidents happened during the peak period. Congested data existed at the stations both upstream and downstream of the incidents before and during the incident times. These congested data do not show any difference compared with the data at the same time and location on previous incident-free days. Of these nine incidents that appear to have had no effect on traffic, the lane of occurrence is specified for only one (which occurred in the shoulder lane). It is possible that a number of these occurred on the shoulder. Conservatively, however, all nine of these incidents are considered to be incidents that should have been detected by the algorithm.

Fifteen incidents were detected successfully. The last nine incidents are considered to be the incidents missed by the algorithm. Hence at worst the detection rate is 15/24, or 62.5 percent. However if the seven incidents that had no effect on traffic are also omitted, then the detection rate is 88 percent (15/17).

The effectiveness at distinguishing between incident-caused and recurrent congestion may be evaluated by considering the number of wrongly classified occurrences of congestion. For the off-line test, each occurrence of recurrent congestion was also printed out; in total there were nearly 800 such occurrences in the 39 days. Three of the nine missed incidents can be matched by recurrent congestion declarations by the algorithm at a similar place and time, although the match is not exact. Five of the 14 false alarms occurred in locations that allow for the possibility that the algorithm identified as an incident something that might have been caused by recurrent congestion. Compared with the nearly 800 correct identifications of recurrent congestion and the 15 correctly identified incidents, these numbers confirm the acceptable performance of the algorithm in distinguishing the two types of congestion.

The ability of the algorithm to detect incidents within congestion is compared with its overall ability, during both congested and uncongested periods, in Table 2. (The congested period was taken to be 6:30 to 9:30 a.m. daily at all stations.) Nine of the 15 successfully detected incidents were detected within the congested period. With regard to false alarms, the percentage obtained by dividing the number of false alarms by the total number of incident declarations is a useful indicator in this instance. This percentage for the time within the congested period is considerably lower than that during the full off-line testing period, indicating that the algorithm is efficient at avoiding false alarms during recurrent

TABLE 2 Comparison of Incident Detection During Congestion and During Entire Period

	within congested period	during the entire period
(1) successful incident detection	9	15
(2) false alarm	5	14
(3) total declaration	14	29
(2)/(3)	35.7%	48.2%
(4) total incidents that might have been detected	14	24
(5) detection rate (1)/(4)	64.3%	62.5%
(6) total incidents that affected traffic	11	17
(7) detection rate based on incidents that affected traffic (1)/(6)	81.8%	88.2%

congestion. The detection rate within congestion is measured as defined earlier: the number of successful detections divided by the number that should have been found. Two values are reported: 64.3 percent, based on counting all nine missed incidents, and 81.8 percent, counting only those two that had a noticeable effect on traffic. Considering all three of these indicators, the ability of the algorithm to find incidents within congestion was judged to be more than satisfactory in the off-line tests.

A sensitivity analysis using four values for the persistence check was conducted with the same 39 days of data following the main off-line testing. The results (Figure 4) show that a value of three intervals is the most effective in terms of the trade-off between detection rate and false alarm rate and between false alarm rate and time to detection. It is interesting to compare this figure with Figures 3 through 7 in the paper by Stephanedes et al. (*10*). Their data cover 14 stations, spanning 5.5 mi (8.8 km); the QEW data come from 15 stations, also spanning 8.8 km. In making such a comparison, however, it is important to keep in mind the differences between the two data sets. Their data were "confined to the afternoon peak period (4:00–6:00 p.m.), since incident detection under moderate-to-heavy traffic conditions is of greatest importance for Advanced Freeway Management," whereas the data in

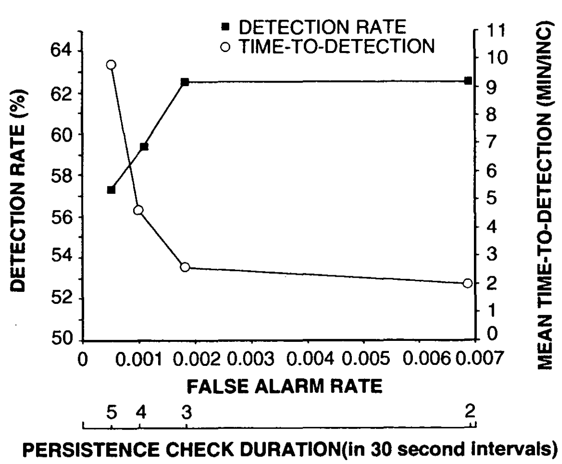

FIGURE 4 Effects of different persistence check durations on false alarm rate, detection rate, and mean time to detection.

the study reported in this paper were collected from 6:00 a.m. to 10:00 p.m.

With regard to mean time to detection, the McMaster algorithm results of roughly 2 min, shown in Figure 4, are two to four times longer than the 0.5 to 1.0 min results at a 60 percent detection rate in Figure 7 of the paper by Stephanedes et al. (*10*). This is perhaps attributable to the different time periods in the two studies; congestion takes longer to reach an upstream detector station in lighter traffic than in moderate-to-heavy traffic. The more striking feature of the comparison is the magnitude of the false alarm rates. At detection rates near 60 percent, Figure 4 shows false alarm rates in the vicinity of 0.001 percent. The results in the paper by Stephanedes et al. (*10*) at comparable detection rates range from 0.2 percent to 0.7 percent for comparative algorithms and are in the vicinity of 1 percent for time series algorithms. At first glance, this difference may also be attributed to the longer daily duration of the McMaster test, in that false alarms may be less likely during light traffic than during moderate-to-heavy traffic. However the rates in the Stephanedes test for the comparative algorithms are similar to those reported in the original literature, so not all of the difference can be due to the daily duration of the test. Hence there is some indication that the McMaster algorithm has an improved false alarm rate and perhaps a slightly increased time to detection (for a 70 percent detection rate).

ON-LINE TEST

Implementation of the algorithm for use on-line had the usual difficulties, but these were overcome in the space of a few months of preliminary on-line testing. One difficulty was that several stations needed to be recalibrated. Another was that two stations in the extended set for the on-line tests should have been defined as recurrent congestion stations and had not been. The final difficulty was that it proved necessary to look further downstream for those cases in which a particular station was missing data at the time congestion had been detected upstream of it. The problem of missing data remained an important one in the testing, with perhaps one-quarter of the data missing or suspect in general.

Testing of version 2.3 began March 12, 1992. Results from that time until June 18, 1992, a total of 67 days, are presented in this paper. Friday and Monday of the Easter holiday weekend are excluded, as is Monday, April 27, because there are no operator's logs for those days. Hence the test consists of 64 days, 16 hr per day, or 1,024 hr. Version 2.3 covers Stations 6 through 27, for a total of 22 stations. Because of treatment of the system end points, only 21 sections are covered. With data coming in every 30 sec, the test includes 2,580,480 decisions (21 sections 1,024 hr 120 data transmissions per hour per station).

During the 64 days of the test, the operators reported 230 incidents, of which 191 were identified as being vehicles on the shoulder. All but four of these have been ignored in evaluating algorithm performance. In one case there was a disturbance visible in the data at approximately the same location, 5 min before the operator reported the vehicle on the shoulder. This has been counted as a matched incident. For another three, the algorithm declared incidents at the correct locations, but the time is later than that on the operator's log. Checking the stored data for these three confirms that there is a clear effect on the data consistent with an incident pattern. Because the only entry in the operator's log to equate these incidents to is the vehicle on the shoulder, these have also been counted as matched incidents.

An additional 10 incidents were identified as "debris on road," or as a truck losing "part" of its load. The operator's practice is to log any occurrence of material on the roadway, whether or not it affects traffic. Consequently the stored data for these 10 incidents were checked to identify the effect on traffic. In 7 of the 10, no visible effect on traffic occurred. In another two, the system was not reporting good data at the relevant stations at the time that the debris was noted. Hence only one of the "debris on road" incidents was of a type that might affect traffic and therefore be visible to an incident-detection algorithm.

For an additional four incidents, the operator's log unfortunately provides insufficient information about the incident to know whether the disabled vehicle is on the shoulder or in one of the traveled lanes. In one of the four, the vehicle might even be on an entrance ramp. The stored data for all four were checked. Three showed no effect on traffic; one had bad data at both stations in the vicinity of the logged location. Hence none of these four logged events would be found by an algorithm.

Thus of the 230 reported events in 64 days, 187 have been removed from the test because they occurred on the shoulder and had no effect on traffic that was noticed by the algorithm, another 9 have been removed as occurrences of debris on the road that did not affect traffic (or for which there was no recorded data), and 4 have been removed because of insufficient information about them (together with no effect on the available data). That leaves a total of 30 incidents on which the algorithm is to be tested. The algorithm identified 19 of these. The stored data were scanned for the 11 missed incidents, and it was found that for an additional 2 incidents, the system was not recording good data at the necessary stations. Those incidents should also be removed from the test set. Hence the algorithm correctly identified 19 of 28 possible incidents, for a detection rate of 68 percent.

With regard to the difference in time to detection between the operator's log and the algorithm, 3 of the 19 matched detections occurred at the same time, 11 were found later by the algorithm, and 5 were found earlier. It is useful to separate the incidents detected on the shoulder from those that occurred on the traveled roadway. Four are identified by the operators as being solely on the shoulder; another two are partially (in time or space) on the shoulder. Of the four completely on the shoulder, the algorithm was 5 min earlier on one, and 1, 2.5, and 11.5 min later on the others. For the two partly on the shoulder, the algorithm was 12 min later on one, and at the same time as the operator on the other. Thirteen incidents occurred solely in the traveled lanes of the roadway, and for these incidents detection time differences were 3, 1.5, 1, and 1 min earlier; two at the same time; and 0.5, 1, 2, 3, 3.5, 4, and 11 min later for the algorithm. The mean time to detection for all incidents is 2.1 min, and for only those on the traveled lanes is 1.4 min. The median difference in detection times for all incidents is 1 min later for the algorithm; for those on the traveled lanes only, it is 0.5 min.

The false alarm rate for the on-line test is similar to the off-line result. There were only 20 false alarms during the 64 days of the test. That is an average of one every 6.4 operator shifts. On the basis of the 2,580,480 decisions identified earlier, this is a false alarm rate of 0.00078 percent. In future work, it might be appropriate to perform sensitivity testing of some other parameters, given the possibility of allowing the false alarm rate to increase slightly if more of the incidents could be captured.

These results may also be compared with those obtained by Stephanedes et al. (*10*), recognizing the differences in the data discussed earlier, as well as with earlier operating characteristic curves such as those in work by Payne and Tignor (*4*). The first point is that the false alarm rate remains two to three orders of magnitude lower in these on-line results. For a detection rate of roughly 68 percent Stephanedes et al. (*10*) reported false alarm rates of 0.1 percent to 0.3 percent. Mean times to detection remain marginally better (by 1 min or less) for most of the algorithms reported by Stephanedes et al. (*10*) than for the McMaster algorithm. Payne and Tignor (*4*) do not report detection times, but the false alarm rates they report are a similar order of magnitude to those in the paper by Stephanedes et al. (*10*). At the closest reported detection rates to that in the McMaster algorithm, the false alarm rates reported by Payne and Tignor (*4*) ranged from 0.13 percent to 0.8 percent.

In summary, the on-line test results suggest that the McMaster algorithm has an excellent false-alarm performance and acceptable detection rates, but that the false alarm performance may have been achieved at the expense of the mean time to detection. For small systems such as that on the QEW and the one used in testing done by Stephanedes et al. (*10*), it may well seem that the false alarm rate for the McMaster algorithm is excessively low because it results in only one false alarm in several operator shifts. However there are larger FTMSs in operation in North America. The system under development on Highway 401 in Toronto, for example, has 136 detector stations, as opposed to the 22 in the QEW on-line test. The system being designed for the Boston Central Artery and Third Harbor Tunnel may have as many as 150, and the Chicago area systems already have 1,800 mainline detectors (*13*), which implies 600 stations, if there are an average of three lanes per station with detectors. Other systems in the design stage with even more detectors (3,000 in Phoenix and 7,000 in Fort Worth) are reported in *Transportation Research Circular 378* (*13*). On a system the size of Chicago's, with 30 times as many stations as on the QEW, the false alarm rate for the McMaster algorithm would mean an average of roughly 4.5 false alarms per shift. Taking the best of the results from the Payne and Tignor report (*4*), a false alarm rate of roughly 0.1 percent, the number of false alarms in a Chicago-size system would be well over 600 per shift, which means more than 1 per minute. False alarm rates need to be as low as those found with the McMaster algorithm for automatic incident detection to be feasible on large systems.

The next step in development of the algorithm for the QEW is to develop version 2.4, which will report to the operators. That will allow for a different on-line test because it will show how many incidents are noticed by the operator before they are detected by the algorithm. However it will no longer be possible in such a test to show the algorithm detecting an incident before the operator because when the algorithm identifies it the operator will know of it, too. It will also allow for a better test of false alarm rates because it may well be that some of the alarms identified as such in this paper were in fact real events that the operators did not notice on the CCTV. In addition to this last step for the algorithm on the QEW system, work has begun on adapting the algorithm for the Highway 401 system, with its more complex geometry and larger number of detector stations.

ACKNOWLEDGMENTS

The work described in this paper has been supported by contracts from the Ministry of Transportation in Ontario. The ideas expressed in the paper represent the opinions of the authors and not of the sponsor. The assistance of Mark Fox, David Tsui, and Phil Masters, of the Ministry's Freeway Traffic Management Section, has been most appreciated. Thanks are also due to Colin Rayman, former head of the section, who had the foresight to take a chance on an idea, and to Peter Korpal, the current section head, who has displayed the confidence to continue that work. The financial support of the Natural Sciences and Engineering Research Council of Canada in developing the original idea should also be acknowledged, as should their support for George Atala through an undergraduate student research stipend. Special mention is due Lisa Aultman-Hall, who worked on several aspects of the algorithm as an undergraduate student before moving to Queen's University for her graduate work.

REFERENCES

1. A. I. Gall. *Distinguishing Between Incident and Recurrent Congestion.* M. Eng. thesis. Department of Civil Engineering, McMaster University, Hamilton, Ontario, Canada, 1990.
2. A. I. Gall and F. L. Hall. Distinguishing Between Incident Congestion and Recurrent Congestion: A Proposed Logic. In *Transportation Research Record 1232,* TRB, National Research Council, Washington, D.C., 1990, pp. 1–8.
3. J. A. Wattleworth and D. S. Berry. Peak Period Control of a Freeway System—Some Theoretical Investigations. In *Highway Research Record 89,* HRB, National Research Council, Washington, D.C., 1965, pp. 1–25.
4. H. J. Payne and S. C. Tignor. Freeway Incident Detection Algorithms Based on Decision Trees with States. In *Transportation Research Record 682,* TRB, National Research Council, Washington, D.C., 1978, pp. 30–37.
5. B. N. Persaud and F. L. Hall. Catastrophe Theory and Patterns in 30-Second Freeway Data—Implications for Incident Detection. *Transportation Research A,* Vol. 23A, 1989, pp. 102–113.
6. H. J. Payne. *Traffic Flow Theory with Application to Freeway Ramp Control and Incident Detection.* Technology Service Corporation, 1974.
7. C. L. Dudek and C. J. Messer. Incident Detection on Urban Freeways. In *Transportation Research Record 495,* TRB, National Research Council, Washington, D.C., 1974, pp. 12–24.
8. A. R. Cook and D. E. Cleveland. Detection of Freeway Capacity-Reducing Incidents by Traffic-Stream Measurements. In *Transportation Research Record 495,* TRB, National Research Council, Washington, D.C., 1974, pp. 1–11.
9. M. Levin and G. M. Krause. Incident Detection Algorithms—Part 1: Off-Line Evaluation, Part 2: On-Line Evaluation. In *Transportation Research Record 722,* TRB, National Research Council, Washington, D.C., 1979, pp. 49–64.
10. Y. J. Stephanedes, A. P. Chassiakos, and P. G. Michalopoulos. Comparative Performance Evaluation of Incident Detection Algorithms. In *Transportation Research Record 1360,* TRB, National Research Council, Washington, D.C., 1992.
11. Y. Shi. *Improvements to the McMaster Incident Detection Algorithm.* M. Eng. thesis. Department of Civil Engineering, McMaster University, Hamilton, Ontario, Canada, 1991.
12. B. N. Persaud, F. L. Hall, and L. M. Hall. Congestion Identification Aspects of the McMaster Incident Detection Algorithm. In *Transportation Research Record 1287,* TRB, National Research Council, Washington, D.C., 1990, pp. 167–175.
13. *Transportation Research Circular 378: Freeway Operations Projects Summary.* TRB, National Research Council, Washington, D.C., Sept. 1991.

Publication of this paper sponsored by Committee on Freeway Operations.

Smoothing Algorithms for Incident Detection

ATHANASIOS P. CHASSIAKOS AND YORGOS J. STEPHANEDES

The majority of automatic incident detection algorithms aim to identify traffic incident patterns but do not adequately investigate possible similarities in patterns observed under incident-free conditions. A classification of major traffic disturbances on freeways is presented. On the basis of this classification, an incident detection logic is developed with the traffic features that result in the best distinction between an incident and other disturbances. The new logic, DELOS (Detection Logic with Smoothing), employs smoothed detector occupancy measurements to signal an incident when significant temporal changes of the smoothed occupancy occur. Three types of smoothers—average, statistical median, and exponential—are considered, leading to corresponding algorithms. The structure of the proposed algorithms is presented and compared with previous algorithms. Comparative evaluation of test results with rush-hour traffic and incident data from I-35W in Minneapolis reveal the improved performance of the proposed method.

Freeway incident detection has traditionally been formulated as a two-hypothesis problem, incident versus nonincident traffic (*1–3*) or incident versus recurrent congestion (*4*). Few researchers have attempted to distinguish incidents from other traffic phenomena that may have a noticeable effect on traffic [e.g., traffic pulses (*5*) and compression waves (*6*)]; further, no single study has considered all major disturbances. These disturbances include random traffic fluctuations that appear frequently and account for a significant portion of false incident alarms.

An incident detection scheme is developed that features simple-occupancy tests and aims to distinguish incidents from other disturbances. The detection scheme is part of IDENTIFY (Incident Detection Enhancements for Traffic In Freeways), a project evolving in two major directions, one employing filtering and a second focusing on neural network applications.

First presented in another article by the authors (*7*), the proposed logic, DELOS (Detection Logic with Smoothing), uses filters that smooth the raw data over sufficiently large time windows to eliminate short-duration traffic disturbances, such as random fluctuations, traffic pulses, and compression waves. Further, it employs comparisons of the smoothed occupancy over time to distinguish between slowly emerging recurrent congestion at bottlenecks and fast-evolving incidents.

The objectives of this work are to (a) introduce a multi-event detection formulation and assess the capabilities of the original algorithm in the multi-event traffic environment and (b) perform sensitivity analysis with several filters that are widely used for smoothing time series data. For smoothing detector occupancy measurements, three types of smoothers are considered—moving average, statistical median, and exponential.

The resulting algorithms have been tested with data from I-35W in Minneapolis and compared with previous algorithms tested with the same data. The wide diversity of the test site in terms of geometric configuration, detector spacing, and location with respect to ramps and the diversity of the incident set with regard to incident type, severity, and location reveals the capability of the proposed detection structure to perform in a wide range of conditions. However, algorithm development and testing have been confined to rush-hour operations.

Test results indicate that, during the peak period, at approximately 60 percent detection rate, DELOS algorithms produce one false alarm per hour in the 8.8-km (5.5-mi) test section, which includes 14 detector stations. This represents a significant false alarm rate reduction in comparison with several previous algorithms.

BACKGROUND

Despite substantial research, algorithm implementation has been hampered by limited performance reliability, substantial implementation needs, and strong data requirements. To investigate the application issue, the authors conducted a survey of transportation departments in the United States and Canada on incident detection strategies currently used in traffic management systems (*8,9*). The survey results indicate that several departments have implemented an incident detection plan. Traffic information is typically collected from loop detectors and includes occupancy and volume averaged at 20 to 60 sec intervals, usually across all lanes. Detector spacing along the freeway is .5 mi on average. Certain systems (e.g., Ontario's Queen Elizabeth Way) also use paired detectors to collect speed data. In a demonstration project in Connecticut, overhead mounted radar detectors will return speed and volume data for incident detection. In Virginia, a switch from loop to video detectors is under way.

Most systems use a California algorithm (*6*) for incident detection. The original California algorithm is used in Minnesota, Ontario, and Virginia. The modified California Algorithm 2, which additionally requires persistence of the incident alarm for two consecutive periods, is used in Los Angeles

Department of Civil and Mineral Engineering, University of Minnesota, 500 Pillsbury Drive, S.E., Minneapolis, Minn. 55455.

and Seattle. Algorithm 7 is used in tunnel locations in Seattle. Different algorithms are often used, depending on traffic conditions.

In other cities, locally developed algorithms have been implemented. In Connecticut, a simple algorithm indicates an incident when speed drops below a threshold. In Illinois, a Bayesian approach (*10*) used the relative spatial occupancy difference as detection parameter. The approach focused on using the probability distributions of the detection parameter under incident and incident-free conditions for determining an optimal detection threshold. Because of its excessive computer time requirements, the Bayesian algorithm was replaced by a simpler one that considers the occupancy difference between the upstream and downstream station; an incident is signaled if this difference continuously exceeds a threshold for 5 min. Although the 5-min persistence test results in long response times, this is preferable to responding to frequent false alarms.

Thresholds for operational algorithms have been typically calibrated by trial and error (in Los Angeles and Seattle), empirical experimentation on historical data (in Illinois), and performance curves obtained from multiple runs of the respective algorithm on the data with incrementally changing thresholds (in Ontario). In Los Angeles, algorithms are frequently recalibrated, especially at locations that produce frequent incident alarms. Algorithm output consists of either textual description (as in Illinois, Virginia, and Seattle) or color computer graphic maps (as in Los Angeles, Minnesota, and Ontario). In the latter case, several congestion levels are indicated with different colors, and incidents are separately indicated (e.g., with flashing red).

Most systems have not quantitatively assessed the operational performance of algorithms in terms of detection and false alarm rates. In Ontario, off-line evaluation of the California algorithm produced a large number of false alarms. Ontario is consequently switching to a promising Canadian algorithm (*4*); this algorithm, in an off-line evaluation, achieved false alarm rates of 1 per station every 64 hr at 75 percent detection. However, In Illinois, where it was also evaluated off-line as a potential substitute of current methods, it resulted in good detection after a difficult calibration period, but the false alarm rate was not satisfactory.

To date, high false alarm rates have prevented implementation of fully automated incident detection. Instead, algorithm alarms typically trigger the operator's attention; the operator verifies the validity of the alarm and decides on the appropriate incident response. In certain cases, operators assume that frequent alarms are false, and they tend to ignore them (as in Illinois). In Los Angeles, incident response is initiated only after an incident has been reported by motorists or a highway patrol officer. Elsewhere, incidents are identified via closed-circuit television cameras (Minnesota) or review of the raw data by the operator (Illinois). Incident detection, especially in the latter case, heavily relies on the expertise of the operator.

Detector failure is an additional concern. Although malfunctioning detector rates have not been systematically assessed (in Los Angeles four to five malfunctioning detectors are identified and repaired weekly), they lead to significantly deteriorated algorithm performance. In certain systems (e.g., Seattle) specific types of detector failure are preempted by measurement validity checks.

TRAFFIC DISTURBANCE PATTERN CLASSIFICATION

Effective incident detection requires consideration of all major sources of false alarms. In particular, traffic flow presents a number of inhomogeneities that are difficult to distinguish from those driven by incidents. Events producing traffic disturbances include incidents, bottlenecks, traffic pulses, compression waves, and random traffic fluctuations. Sensor failure, also treated as an event, is only related to the measurement component of detection systems. The major characteristics of each event are discussed in the following sections.

Incidents

Incidents are unexpected events that block part of the roadway and reduce capacity. Incidents create two traffic regimes, congested flow upstream (high occupancies) and uncongested downstream (low occupancies), as indicated in Figure 1 for a typical accident blocking moving lanes. Two shock waves are generated and propagate upstream and downstream, each accompanying its respective regime. The congested-region boundary propagates upstream at approximately 16 km/hr (10 mph), and its value depends on incident characteristics, freeway geometry, and traffic level. Downstream of the incident, the cleared region boundary propagates downstream at a speed that may reach 80 km/hr (50 mph) (*6*).

The evolution and propagation of each event is governed by several factors, the most important of which are incident type, number of lanes closed, traffic conditions before incident, and incident location relative to entrance or exit ramps, lane drops or additions, sharp turns, grade, and sensor stations. Other, less important factors, which are harder to model, include pavement condition, traffic composition, and driver characteristics.

Incident patterns vary depending on the nature of the incident and prevailing traffic conditions (*6*). The most distinctive pattern occurs when the reduced capacity from incident blockage falls below oncoming traffic volume so that a queue develops upstream. This pattern, which is clearest when traffic is flowing freely before the incident, is typical when one or more moving lanes are blocked following severe accidents (Figure 1). The second pattern type occurs when the prevailing traffic condition is freely moving but the impact of the incident is not severe. This may result, for example, from lane blockage that still yields reduced capacity higher than the volume of incoming traffic. This situation may lead to missed detection, especially if the incident is not located near a detector. The third type characterizes incidents that do not create considerable flow discontinuity, as when a car stalls on the shoulder. These incidents usually do not create observable traffic shock waves and have limited or no noticeable impact on traffic operations. The fourth type of incident occurs in heavy traffic when a freeway segment is already congested. The incident generally leads to clearance downstream but a

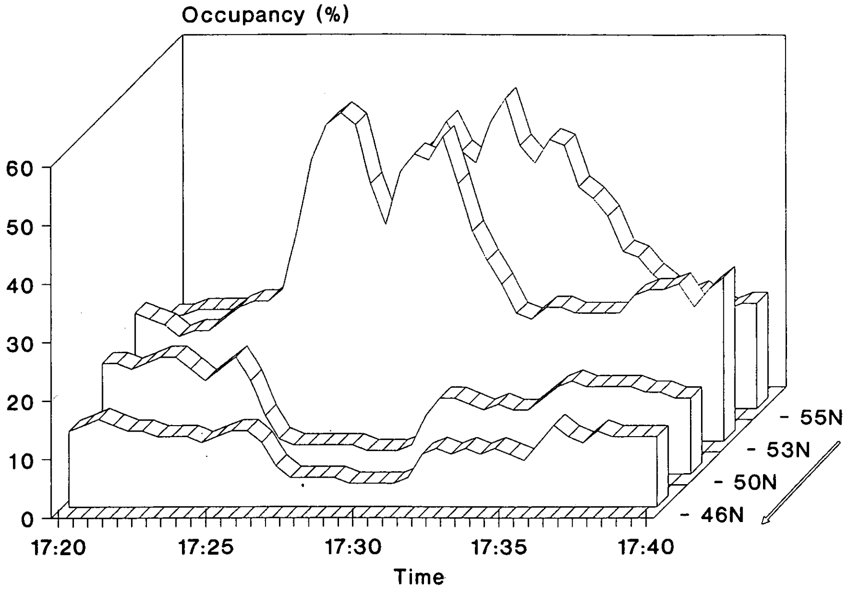

FIGURE 1 Incident pattern (I-35W North, 11/21/89).

distinguishable traffic pattern develops only after several minutes, except in a severe blockage. This type of incident is often observed in secondary accidents at the congested region upstream of an incident in progress.

Bottlenecks

Bottlenecks are formed where the freeway cross-section changes (e.g., in lane drop or addition). While incidents have only temporary effects on occupancies, bottlenecks generally result in longer lasting spatial density or occupancy discrepancies. A typical bottleneck is shown in Figure 2. The figure presents occupancy measurements at three consecutive stations of a freeway segment involving a lane drop between the first two and a lane addition between the second and third stations. Under normal conditions, the three stations operate at different average occupancy levels. This difference is more pronounced between stations 61S and 62S.

Traffic Pulses

Traffic pulses are created by platoons of cars moving downstream. Such disturbances may be caused by a large entrance-ramp volume caused by the exodus from a sporting event, for example. The observed pattern is an increase in occupancy in the upstream station followed by a similar increase in the

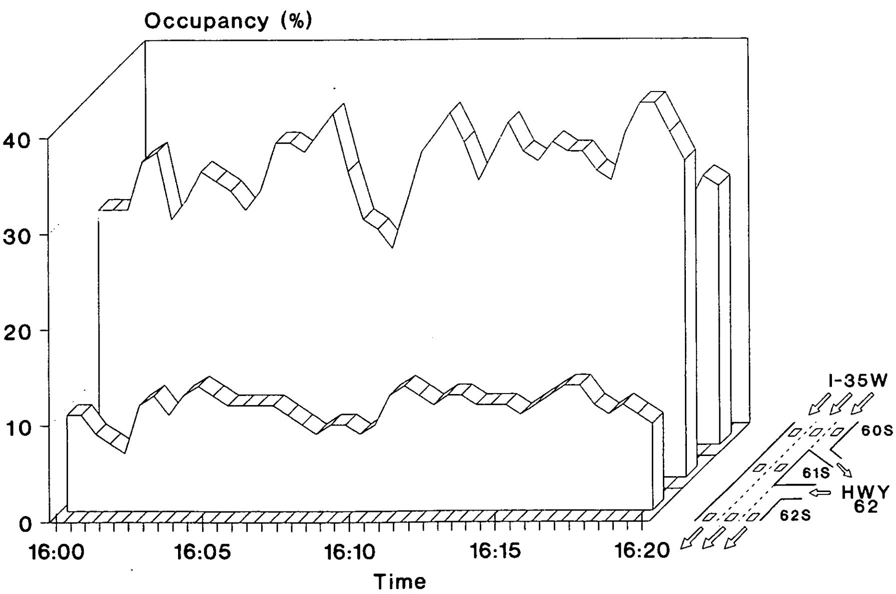

FIGURE 2 Occupancy measurements at bottleneck.

downstream station. Because of ramp metering during the testing period, traffic pulses are rarely observed in this data set.

Compression Waves

Compression waves occur in heavy, congested traffic, usually following a small disturbance and are associated with severe slow-down, speed-up vehicle speed cycles. Waves are typically manifested by a sudden, large increase in occupancy that propagates through the traffic stream in a direction counter to traffic flow (Figure 3). The data reveal that compression waves result in significantly high station occupancies of the same magnitude as those of incident patterns.

Random Traffic Fluctuations

Random traffic fluctuations are often observed in the traffic stream as short-duration peaks of traffic occupancy. These fluctuations, although usually not high in magnitude, may form an incident pattern or obscure real incident patterns.

Detection System Failures

Detection system failures may be observed in several forms, but a particular form has resulted in a specific pattern in the data observations discussed here. This pattern is observed with isolated high-magnitude impulses in the 30-sec volume/occupancy measurements, appearing simultaneously in several stations. These values are considered outliers or impulsive data noise.

PROPOSED ALGORITHM DESCRIPTION

The authors' review of incident detection strategies that are currently used indicates that algorithms that are intuitively appealing, computationally simple, and based on widely available aggregate (20 to 60 sec) traffic data are most likely to be implemented in freeway control systems. Within this specification, the proposed logic aims to develop simple occupancy tests to distinguish incidents from other traffic disturbances. Two major characteristics can be used for this purpose. First, incidents result in rapid temporal changes in traffic conditions. Second, incident duration is longer than that of other disturbances.

The first characteristic distinguishes incident congestion from bottleneck (recurrent) congestion, which evolves more slowly. This is because recurrent congestion results from demand increasing over capacity at bottleneck locations. The demand increase generally does not occur as fast at incident locations. The duration characteristic can differentiate incidents from short-duration disturbances. The basic concepts behind the algorithm development (*7*) are summarized next. The major focus of the current effort is to investigate the capabilities of the method to avoid signaling false alarms across each type of traffic disturbance and to perform sensitivity analysis across different types of smoothing filters.

DELOS algorithms involve smoothing occupancy measurements to distinguish short-duration traffic inhomogeneities from incidents. When an inhomogeneity is present, smoothing eliminates or diminishes its impact; on the other hand, smoothing does not substantially modify the incident pattern if its duration is greater than the number of terms in the smoother. Although smoothing may conceal the patterns of some nonsevere incidents, the large reduction in false alarms compensates for a few possibly missed incidents. Test results indicate a significant reduction in false alarms as compared with similar algorithms [e.g., Standard Deviation (*1*), Double

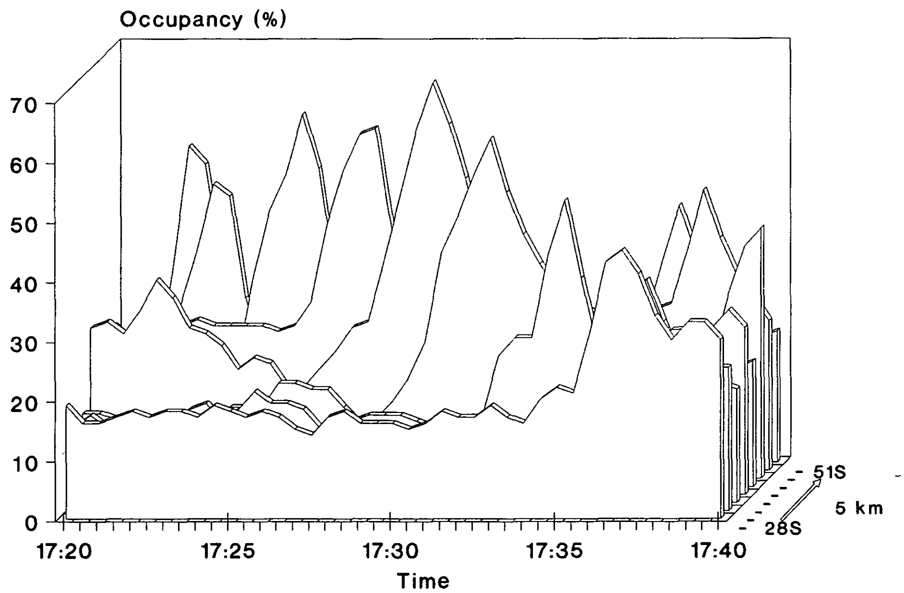

FIGURE 3 Compression wave (I-35W South, 11/16/89).

Exponential (2), and California (6)], which attempt raw data manipulations.

Further, in a manner similar to but more effective than previous algorithms, the proposed structure attempts to distinguish recurrent from incident congestion on the basis of slow or fast evolution of congestion, respectively. In particular, the distinguishing logic is based on temporal comparison of the detection variable, spatial occupancy difference between adjacent stations. For comparison, the incident test of the California algorithm considers occupancy reduction at the downstream station; however, such reduction is not always observed during incidents.

Two smoothed values are considered for the detection variable: one represents current traffic conditions, and the other represents past conditions. For an incident occurring at time t, define $OCC_i(t + k)$, smoothed occupancy at station i from k occupancy values after t, and $OCC_i(t)$, smoothed occupancy at station i from n occupancy values before t, where k and n represent the window size to smooth the data for the current and past periods, respectively. The incident is likely to create congestion in upstream station i and reduce flow in downstream station $i + 1$, leading to a high value of spatial occupancy difference, $\Delta OCC(t + k)$, as described in Equation 1.

$$\Delta OCC(t + k) = OCC_i(t + k) - OCC_{i+1}(t + k) \quad (1)$$

Further, to distinguish from bottleneck congestion, the spatial occupancy difference $\Delta OCC(t + k)$ for the current period is compared with the corresponding value $\Delta OCC(t)$ from the past period (see Equation 2).

$$\Delta OCC(t) = OCC_i(t) - OCC_{i+1}(t) \quad (2)$$

Both tests, congestion and incident, are normalized by the highest value of the two occupancies, upstream and downstream, as defined in Equation 3.

$$maxOCC(t) = max[OCC_i(t), OCC_{i+1}(t)] \quad (3)$$

This reflects changes with respect to existing conditions before an incident. The normalization increases the potential for algorithm transferability across locations. In summary, the proposed detection logic involves two tests, congestion (Equation 4) and incident (Equation 5), where T_C and T_I are the respective thresholds.

$$\frac{\Delta OCC(t + k)}{maxOCC(t)} \geq T_c \quad (4)$$

$$\frac{\Delta OCC(t + k) - \Delta OCC(t)}{maxOCC(t)} \geq T_I \quad (5)$$

The major concerns in selecting a smoothing technique are related to its effectiveness in eliminating undesirable sources of false alarms, the extent to which smoothing distorts the information content of incident patterns, and the detection delay imposed from the need to obtain a number of measurements while an incident is in progress.

Moving average, a linear transformation, is a simple but effective smoothing technique. The occupancy measurement at time t and detector station i, $o_i(t)$, is smoothed via Equation 6.

$$OCC_i(t) = \frac{1}{L} \sum_{l=0}^{L-1} o_i(t - l) \quad (6)$$

Moving averages of a different order, $L = k$ and $L = n$, corresponding to smoothing factors $1/k$ and $1/n$, are used for the current and past periods, respectively. Window sizes k and n are selected to optimize algorithm performance. An additional length constraint is imposed on k because long smoothing windows (e.g., longer than 10 samples) would result in excessive delays in algorithm response. This linear transformation, although effective in removing traffic fluctuations, distorts information-bearing edges (i.e., step-like changes caused by incidents), possibly obscuring their information content. An alternative, nonlinear, transformation employing the statistical median of the data window (see Equation 7) has been considered to address the issue.

$$OCC_i(t) = median\,[o_i(t), o_i(t - 1), \ldots, o_i(t - L)] \quad (7)$$

Exponential smoothing is a third smoothing technique, extensively used in determining data trends. The general form of the smoother is shown in Equation 8, where α is the smoothing factor.

$$OCC_i(t) = \alpha \cdot o_i(t) + (1 - \alpha) \cdot OCC_i(t - 1) \quad (8)$$

A number of algorithms have been developed along the three major types of smoothing. The algorithms are coded as DELOS $x.y$ (z,w), where x and y represent the smoother type used for the past and current periods, respectively, with the values of 1 for average, 2 for median, and 3 for exponential smoother. Further, z and w represent the past and current period window sizes to smooth the data in the average or median smoother. In exponential smoothers, z represents the smoothing factor , and w is the time lag k between the end of the past and the end of the current period. In combinations of exponential smoothing for the past data with other types of smoothing for the current, the window size for the current data period represents the above time lag. For example, DELOS 3.1 (0.05, 6) smooths past data exponentially with $\alpha = 0.05$ and current data with an average 6-sample window.

DATA DESCRIPTION

The proposed algorithms and several algorithms from the literature were tested with actual data. In particular, 140 hr of afternoon peak-period (4:00–6:00 p.m.) traffic data from a 8.8-km (5.5-mi.) segment of southbound I-35W in Minneapolis (Figure 4) were collected through the Minnesota Department of Transportation Traffic Management Center. The freeway segment has 3 lanes along most of its length. It includes 2 major bottlenecks. The first is in the merging area between I-35W southbound and Highway 62 westbound, where 3-lane I-35W drops a lane for a short section. The typical occupancy pattern in these 3 stations is presented in Figure 2. The second bottleneck location is at Minnehaha Creek bridge, north of Diamond Lake Road, where a freeway segment with an uphill grade is followed by elimination of the shoulder at the bridge. The test segment includes four entrance and five exit ramps.

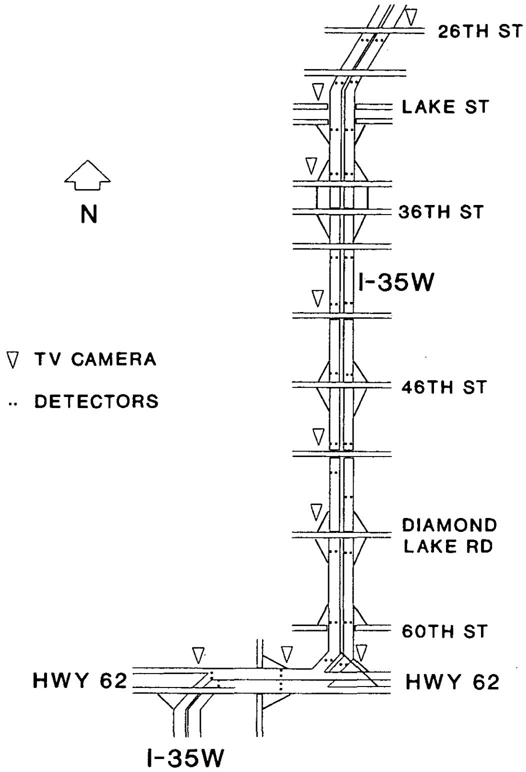

FIGURE 4 Study site in Minneapolis (I-35W).

The traffic data consist of 30-sec volume and occupancy measurements from loop detectors, forming 14 stations imbedded along the road, 0.5 to 1.1 km (0.3 to 0.7 mi) apart. The 30-sec data are averaged across lanes, producing one measurement for each station at every time interval. This traffic data set is a typical one, collected routinely in most U.S. cities. Algorithm development has been aligned with typical longitudinal data availability and, therefore, the algorithm can be implemented across a wide range of systems, including those that do not place detectors in all lanes. Algorithms that depend on additional information (e.g., measurements for each lane, speed, and shorter time measurements), although potentially effective, cannot be implemented across all systems.

During the testing period, 27 incidents were reported by the traffic operator. Detection was accomplished mostly through closed-circuit television cameras along the freeway segment. Of all incidents, 15 were accidents, 3 occurred in the moving lanes, and 12 were moved to the shoulder. According to the operator log, 6 accidents had severe impact on traffic operations, 4 happened in an already congested region, 3 had limited congestion impact on traffic, and the rest were not classified. Besides the accidents, 12 vehicle stalls were observed. All occurred on the shoulder, 1 had a severe impact on traffic, 1 occurred in an already congested region, 7 produced limited congestion, and 3 were not classified.

SENSITIVITY ANALYSIS AND COMPARISON OF ALGORITHM PERFORMANCE

Results from tests evaluating the effectiveness of the new algorithm include the main algorithm performance measures, namely detection rate (ratio of incidents detected out of all incidents), false alarm rate (ratio of false alarms out of all decisions, incident or nonincident, made by the system), and mean time to detect (average time duration needed for detection). (Detection time is measured from the time incidents reported in the operator's log instead of from actual occurrence time).

Algorithm performance is assessed via operating characteristic curves, an evaluation method whose effectiveness lies on its independence from algorithm structure. Operating characteristic curves depict detection and false alarm rates accomplished by an algorithm across threshold values. To construct these curves, the threshold parameters are allowed to vary over a wide range of values. Every threshold set (pair) produces a performance point (P_D, P_F) on the curve.

Three types of smoothing have been considered, linear (average), median, and exponential. Past and current occupancy measurements are smoothed according to one of these types and are included in the corresponding version of the algorithm. For each type of smoothing, several alternative structures were tested by varying the number of terms in the smoothing windows or the value of the smoothing parameter, α, in the exponential version. In particular, window sizes with 5 to 20 terms for the past and 3 to 10 for the current period, and exponential smoothing factors of 0.03 to 0.10 were considered. For each algorithm type, only the structures whose parameters result in optimum performance in terms of detection and false alarms are presented. The characteristics of the selected algorithms are presented in Table 1. Threshold sets and algorithm performance measures are presented in Table 2.

To assess the performance improvement from using smoothed data instead of raw data, the performance of DELOS was compared with that of an older algorithm featuring a structure similar to DELOS. In particular, the Double Exponential algorithm is based on smoothing the surveillance data (e.g., the spatial occupancy difference between adjacent stations) according to Equations 9 and 10. Functions $S_1(t)$ and $S_2(t)$ provide a forecast of spatial occupancy difference, and an incident is signaled when the cumulative error between forecast and current measurement exceeds a threshold. To obtain comparable performance measures, the algorithm was tested on the same data set as the new algorithms.

TABLE 1 Characteristics of Smoothing Algorithms

Algorithm	Past period smoother	Present period smoother
DELOS 1.1 (10,8)	Average, length 10.	Average, length 8.
DELOS 2.2 (9,9)	Median, length 9.	Median, length 9.
DELOS 3.3 (0.05,6)	Exponential, α=0.05.	Exponential, α=0.05, time lag 6.
DELOS 3.1 (0.05,6)	Exponential, α=0.05.	Average, length 6.

TABLE 2 Thresholds and Performance Results

Algorithm	T_C	T_I	Detection Rate (%)	False Alarm Rate (%)	Hourly Number of False Alarms*	Average Detection Time (min)
DELOS1.1	0.30	0.30	85	0.431	6.7	0.8
(10, 8)	0.40	0.40	74	0.224	3.5	1.1
	0.55	0.55	63	0.074	1.2	2.1
	0.60	0.60	48	0.053	0.8	1.6
	0.70	0.70	41	0.033	0.5	2.1
	0.80	0.80	33	0.016	0.2	1.3
DELOS2.2	0.30	0.40	78	0.361	5.6	1.4
(9, 9)	0.40	0.50	74	0.223	3.5	1.8
	0.40	0.60	59	0.137	2.1	1.8
	0.50	0.70	48	0.080	1.2	1.8
	0.60	0.80	41	0.039	0.6	1.6
	0.80	0.80	30	0.031	0.5	0.7
DELOS3.3	0.07	0.07	85	0.411	6.4	1.0
(0.05, 6)	0.15	0.10	78	0.176	2.7	1.1
	0.17	0.12	67	0.113	1.8	1.4
	0.16	0.16	59	0.059	0.9	1.3
	0.15	0.20	48	0.030	0.5	1.5
	0.23	0.23	33	0.014	0.2	3.2
DELOS3.1	0.40	0.30	85	0.476	7.4	0.6
(0.05, 6)	0.50	0.40	78	0.257	4.0	1.1
	0.60	0.40	70	0.173	2.7	1.1
	0.60	0.60	63	0.084	1.3	1.1
	0.70	0.70	48	0.045	0.7	1.2
	0.80	0.80	41	0.029	0.5	1.9
	1.00	0.70	30	0.022	0.3	2.1

* test site length is 8.8 km and includes 14 detector stations

$$S_1(t) = \alpha\,[OCC(i,t) - OCC(i+1,t)] + (1-\alpha)\,S_1(t-1) \tag{9}$$

$$S_2(t) = \alpha\,S_1(t) + (1-\alpha)\,S_2(t-1) \tag{10}$$

The two exponential DELOS algorithms are compared with the Double Exponential algorithm in Figure 5. The comparison indicates that smoothing current measurements leads to substantial reduction in false alarms.

The evaluation results for the four alternative smoothing types are shown in Figure 6. Although the performances of average and exponential smoothers are comparable, they are superior to the median algorithm. This may be attributed to the fact that median smoothers tend to better preserve data fluctuations, and this produces a higher number of false alarms.

To better appreciate the performance of the proposed algorithms, Figure 6 includes performance curves from two approaches tested here, modified California and Algorithm 7. The comparison indicates significant detection improvement of the proposed algorithms, especially at low false alarm rates that are most suitable for operational use. In particular, presenting false alarm performance in the hourly number of alarms indicates algorithm potential as an operator's primary tool for incident detection. For instance, from Table 2, the proposed algorithms, at 60 percent detection rate, yield approximately 1 false alarm per peak hour in an 8.8-km (5.5-mi) heavily traveled freeway segment with 14 detector stations.

Performance may also be seen in terms of detection rate for each type of incident. Accidents, for instance, are more important to detect than stalls because they typically have a strong congestion effect and require prompt emergency assistance. To investigate the algorithm effectiveness in detecting each type of incident, detection performance of DELOS 1.1 (10,8) in detecting accidents and stalls is shown in Figure 7. The figure indicates, for example, that at approximately 1 false alarm per hour (false alarm rate = 0.07 percent) in the whole section, accident detection rate approaches 90 percent.

Regarding average detection time (Table 2), the range for the average algorithm is 1 to 2 min, for the median 1.5 to 2 min, for the exponential 1 to 1.5 min, and for the combined exponential-average 0 to 1 min. The incident occurrence times are from the operator's log. Detection time is the time between occurrence and end of the current period for which an incident alarm is signaled. Therefore, these values do not reflect actual detection times but the delay with regard to detection by the operator. This partly explains why an 8-interval forward seeking method [e.g., DELOS 1.1 (10,8)] has a detection time of 1 to 2 min instead of at least 4 min. A second reason is that, especially in severe incidents, the algorithm may start sensing occupancy changes while only part of the current window overlaps with the incident period. DELOS detection times are slightly longer than those of algorithms employing raw data. For instance, the California algorithms exhibit response times of 0 to 1 min.

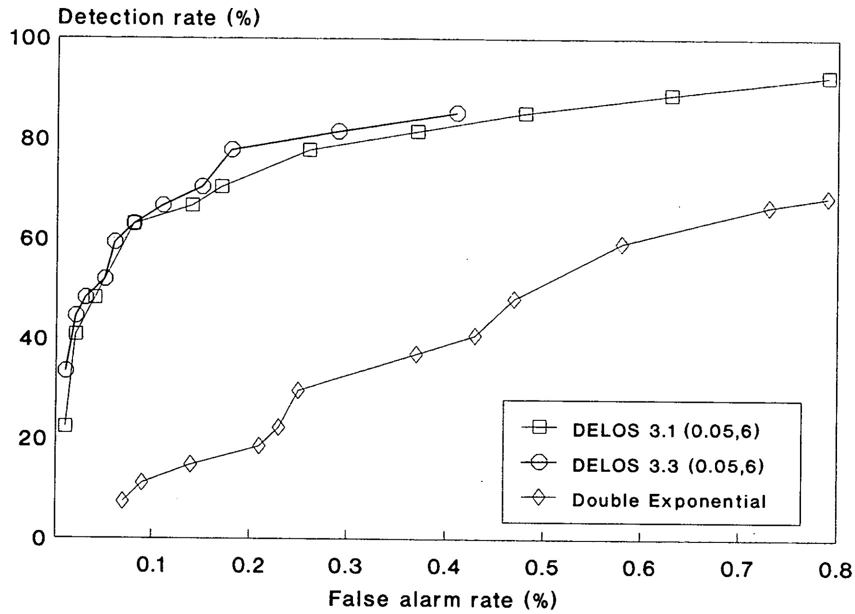

FIGURE 5 Algorithm performance comparison: DELOS versus Double Exponential.

A final comment brings up a positive contribution of computerized algorithms—the reduction in missed detection. In particular, algorithms can identify incidents not detected by operators. To investigate this, the major false alarms produced by the new algorithms were examined; 22 of these exhibit incident-like patterns, including patterns that strongly resemble incidents. However, because no off-line incident identification can independently verify their occurrence, such alarms were treated as false. Further, all incidents missed by the algorithms were stalls on the shoulder and had no impact on traffic. Because they were recorded by the traffic operator, they were classified as missed incidents. If they had been discarded, the detection rate of the new algorithms would increase substantially.

CONCLUSION

The aim of this research is to improve automatic incident detection on freeways by developing strategies to distinguish incidents from other traffic disturbances. Major disturbances identified in this paper include recurrent congestion at bottle-

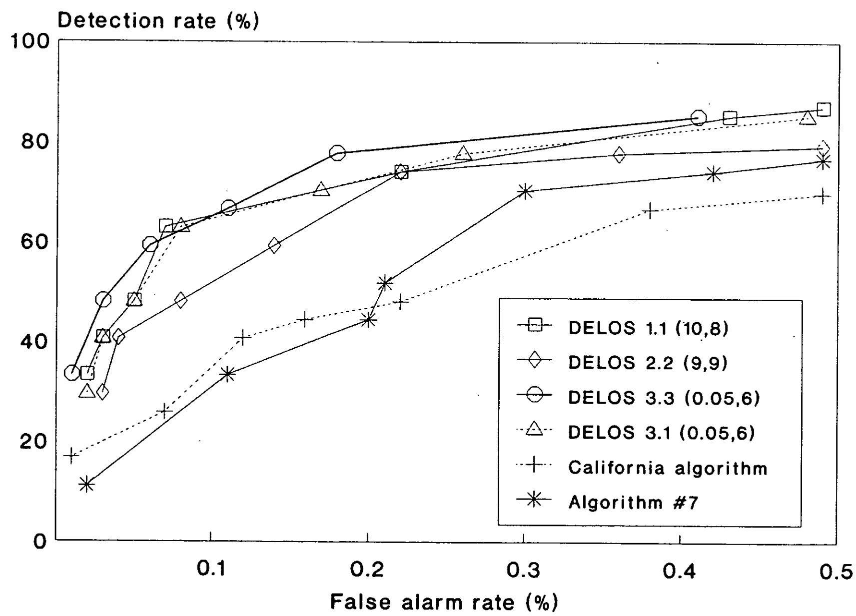

FIGURE 6 Algorithm performance comparison: DELOS versus California algorithms.

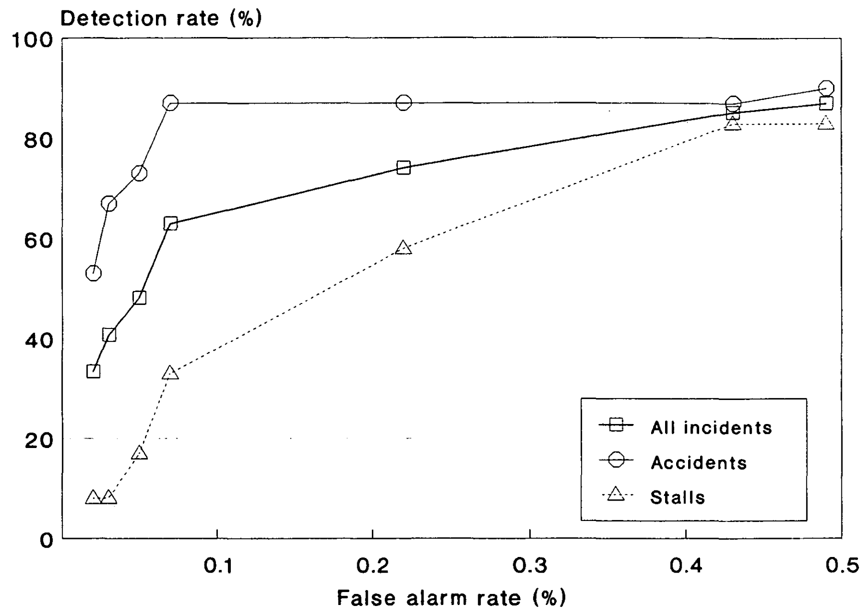

FIGURE 7 DELOS detection performance by incident type.

neck locations, traffic pulses propagating downstream, compression waves propagating upstream, short-duration random fluctuations, and impulsive noise in the traffic measurements.

An incident detection logic was developed within the framework of comparative algorithms (i.e., algorithms that employ simple comparisons of traffic data). The detection logic attempts to distinguish incidents from other disturbances based on two major incident characteristics—fast evolution of congestion following incidents and long duration of incident patterns. To distinguish sharply evolving incident congestion from gradually developing recurrent congestion, temporal comparisons of traffic patterns were performed. Further, to filter out short-duration traffic disturbances, several types of data smoothing were employed. The smoothing algorithms effectively filter short-duration traffic inhomogeneities, such as random traffic fluctuations and traffic pulses, but do not adequately handle compression waves of long duration.

The proposed algorithms were tested with loop detector data from I-35W in Minneapolis with promising results (for instance, 1 false alarm per rush hour per 14 detector stations at approximately 60 percent detection rate). They were also compared with major algorithms used for assisting incident management personnel in urban areas and were found to be superior at all false alarm rates.

ACKNOWLEDGMENTS

This study was supported by the National Science Foundation and in part by the Minnesota Supercomputer Institute. The Center for Transportation Studies of the Department of Civil and Mineral Engineering at the University of Minnesota is acknowledged for its support. The Traffic Management Center of the Minnesota Department of Transportation also cooperated in the study by providing the necessary data.

REFERENCES

1. C. L. Dudek and C. J. Messer. Incident Detection on Urban Freeways. In *Transportation Research Record 495*, TRB, National Research Council, Washington, D.C., 1974, pp. 12–24.
2. A. R. Cook and D. E. Cleveland. Detection of Freeway Capacity-Reducing Incidents by Traffic-Stream Measurements. In *Transportation Research Record 495*, TRB, National Research Council, Washington, D.C., 1974, pp. 1–11.
3. J. F. Collins, C. M. Hopkins, and J. A. Martin. *Automatic Incident Detection—TRRL Algorithms HIOCC and PATREG.* TRRL Supplementary Report 526. Transport and Road Research Laboratory, England, 1979.
4. B. N. Persaud, F. L. Hall, and L. M. Hall. Congestion Identification Aspects of the McMaster Incident Detection Algorithm. In *Transportation Research Record 1287*, TRB, National Research Council, Washington, D.C., 1990, pp. 167–175.
5. A. S. Willsky, E. Y. Chow, S. B. Gershwin, C. S. Greene, P. Houpt, and A. L. Kurkjan. Dynamic Model-Based Techniques for the Detection of Incidents on Freeways. *IEEE Transactions on Automatic Control*, Vol. AC-25, No. 3, June 1980, pp. 347–360.
6. H. J. Payne and S. C. Tignor. Freeway Incident Detection Algorithms Based on Decision Trees with States. In *Transportation Research Record 682*, TRB, National Research Council, Washington, D.C., 1978, pp. 30–37.
7. Y. J. Stephanedes and A. P. Chassiakos. Application of Filtering Techniques for Incident Detection. *ASCE Journal of Transportation Engineering*, Vol 119, No. 1, Jan./Feb. 1993, pp. 13–26.
8. Y. J. Stephanedes and A. P. Chassiakos. *Development and Application of Incident Detection Techniques to Improve Incident Management in Freeway Corridors.* CTS Report 0797-5027. University of Minnesota, Minneapolis, 1992.
9. Y. J. Stephanedes, A. P. Chassiakos, and P. G. Michalopoulos. Comparative Performance Evaluation of Incident Detection Algorithms. In *Transportation Research Record 1360*, TRB, National Research Council, Washington, D.C., 1992, pp. 50–57.
10. M. Levin and G. M. Krause. Incident Detection: A Bayesian Approach. In *Transportation Research Record 682*, TRB, National Research Council, Washington, D.C., 1978, pp. 52–58.

Publication of this paper sponsored by Committee on Freeway Operations.

Effect of Response Limitations on Traffic-Responsive Ramp Metering

JAMES H. BANKS

Simulations of ramp meter responses were used to study the feasibility of replacing locally based moving-average estimates of mainline flow currently used in San Diego with estimates based on upstream data. The choice of estimation methodology made little practical difference in the performance of the system; instead, the major problem in providing precise control of meter outputs was the limited ability of the system to respond. The most important limitation is that the difference between the maximum and minimum metering rates tends to be small relative to normal random variation in mainline input flows at the minimum counting interval of 30 sec. Consequently, meters respond with their maximum or minimum rates most of the time, which leads to biases in average responses. Other response limitations include a comparatively large number of ramps at which demands are less than minimum metering rates and insufficient total metering capacity. For multiramp systems, the most promising way to prevent biases caused by the meters' limited response ranges is to set flow targets for upstream meters to cause the average response of the bottleneck meter to be about halfway between its maximum and minimum rates. This strategy may be employed only where there is sufficient total metering capacity and may conflict with other strategies for setting flow targets in multiramp systems, such as the so-called Wattleworth strategy.

Traffic-responsive ramp metering is an important technique of freeway traffic control. Its most obvious advantage is its potential to hold flows through bottlenecks close to capacity, despite fluctuations in traffic volumes arriving from upstream. In order to do this, it is necessary to calculate meter responses based on information about the state of the traffic stream.

Most past discussions of traffic-responsive metering strategies were focused on the best way to calculate the meter's response, given certain types of information about the system, or on determining the best indicators of the flow state. Little explicit attention is paid to the system's ability to respond and how this might affect the control strategies; at most, certain limitations may be included implicitly as constraints in mathematical models. Important response limitations include maximum and minimum limits to metering rates, the possibility that ramp demand may fall below the minimum metering rate, and constraints on ramp queue lengths.

The work reported here began as a feasibility study of a minor proposed modification to an existing traffic-responsive metering system. In the evaluation of that proposal, it became evident that the most important issue was the way in which the system's output was affected by its response limitations, particularly the maximum and minimum metering rates, and not the way in which the nominal response was calculated. A scheme for analyzing response errors in traffic-responsive metering systems is presented in this paper, and the scheme is applied to the simulated results of proposed modifications to the San Diego system. Some of the ways in which response limitations affect traffic-responsive strategies are discussed.

Civil Engineering Department, San Diego State University, San Diego, Calif. 92182-0189.

TRAFFIC-RESPONSIVE METERING STRATEGIES

Traffic responsive metering strategies may be divided into two types. Local strategies involve control of a single ramp to produce capacity flow through a bottleneck; in addition, considerable attention has been given to so-called "gap-acceptance" strategies (*1*). These strategies, which are not covered here, are intended to smooth flow without necessarily providing capacity operation. Global strategies manage a number of ramps to control flows throughout a more extended section of freeway. Because it is often impossible to hold flows approaching bottlenecks to capacity without metering several ramps, control of flow at the bottleneck is often the ultimate goal of multiramp systems; however, global strategies are usually expressed in terms of target flow states (usually volumes, traffic densities, or lane occupancies) distributed throughout the system.

Several global traffic-responsive metering strategies based on automatic control theory have been proposed. The most common variety is based on minimization of a quadratic performance functional that penalizes deviations from nominal values of flow and traffic density throughout the system and employs some variation of the overlapping decentralization scheme of Isaksen and Payne (*2–4*). Another proposed global strategy is the hierarchical system of Papageorgiou (*5,6*). This consists of three functional layers: (a) an optimization layer based on the steady-state linear programming formulation proposed originally by Wattleworth and Berry (*7*) and later extended by Wattleworth and others (*8–11*); (b) an adaptation layer that reacts to congestion and significant deviations from assumed origin-destination trajectories; and (c) a direct-control layer that implements local feedback controls. Papageorgiou's scheme allows the overall metering problem to be decomposed into separate global and local strategies, with the results of the global strategy entering the local strategy as output flow targets (or average metering rates) at the various meters. Such a decomposition simplifies the design of global traffic-responsive systems but also raises the issue of the compatibility of particular global and local strategies.

Local traffic-responsive strategies include demand-capacity strategies, traditional occupancy-based strategies, and feedback strategies (*12*). Demand-capacity and traditional occupancy-based strategies are similar. In both, metering rates are calculated as the difference between a target output volume and the traffic volume measured (or estimated) just upstream of the ramp. The only difference is that traditional occupancy-based schemes estimate upstream volumes from occupancies instead of measuring them directly; the advantage of this is that it does not require detectors for all lanes. In both cases, the normal response calculation is overridden whenever high occupancies indicate congestion; in this case, minimum metering rates are employed. Feedback strategies base meter responses on traffic conditions downstream of the ramp. Recent European work has included the development and testing of the occupancy-based feedback rule ALINEA (*12–14*). Real-life tests on European freeways have indicated that ALINEA is superior to demand-capacity and traditional occupancy-based strategies.

The superiority of feedback strategies appears to lie in their quicker response to the transition from uncongested to congested flow. Under uncongested conditions, all local strategies are similar and can be thought of as variations of the following rule:

$$M(t) = Q - \hat{q}(t) + \Delta(t) \qquad \text{for } M_{min} \leq M(t) \leq M_{max} \tag{1}$$

otherwise $M(t) = M_{max}$ or M_{min}, as appropriate, where

$M(t)$ = metering rate for time period t,
M_{max} = maximum metering rate,
M_{min} = minimum metering rate,
Q = output flow target,
$\hat{q}(t)$ = estimated mainline flow arriving from upstream during time period t,
$\Delta(t)$ = estimated difference between the actual ramp output and the metering rate; $\Delta(t) = M(t) - p(t)$, where $p(t)$ is the ramp passage count.

Alternatively, the relationship between the passage count and the metering rate may be expressed as the ratio $r(t) = M(t)/p(t)$, and

$$M(t) = \frac{Q - \hat{q}(t)}{r(t)} \tag{2}$$

as long as $M_{min} \leq M(t) \leq M_{max}$.

In either version, this rule states that the metering rate for time period t is a function of the difference between the target output flow and an estimate of the mainline input flow for the same time interval. The various strategies differ only in the way in which the current upstream flow is estimated. Demand-capacity strategies estimate it on the basis of past upstream volume counts; traditional occupancy-based strategies estimate it on the basis of past upstream occupancy measurements; and feedback strategies estimate it on the basis of past downstream volumes or occupancies. In addition, estimates may involve various types and degrees of data smoothing.

EVALUATION SCHEME

The ultimate goal of local traffic-responsive metering strategies is to reduce delay by maximizing flow through the bottleneck. All are based to some extent on the so-called two-capacity phenomenon, in which maximum uncongested flows have been found to exceed queue discharge rates. Consequently, they are designed to hold uncongested flow as close as possible to its maximum value, and to act quickly to restore uncongested flow whenever flow breakdown does occur. They can be successful only if they can hold uncongested flows above the queue discharge rate. Since recent research indicates that maximum uncongested flow rates exceed queue discharge rates by no more than 5 to 6 percent (*15–19*), metering is required to be quite precise.

The most direct means of evaluating such control strategies is to implement them to determine whether they can produce bottleneck flows that exceed the queue discharge rate over extended periods of time. Where this is not possible, they may be simulated and analyzed by their ability to produce a predetermined target output flow.

In carrying out this evaluation, it is important to consider the time frame. Most traffic-responsive systems use short count intervals, often 1 min or less. The meter's responses are updated with similar frequency, so that it is possible to evaluate the system's ability to produce specific 30-sec or 1-min output volumes. It is not clear, however, what deviations from the target flow are significant at this level of disaggregation. Short-term volume counts normally display a great deal of random variation, and this is true of both queue discharge and uncongested flow. The *Highway Capacity Manual* (*20*) uses 15-min flows to define capacity; other research related to capacities, such as the recent work related to the two-capacity phenomenon, has used intervals of at least 5 or 6 min. Meanwhile, average queue discharge rates, which are the critical point of comparison in evaluating the performance of the system, may often be measured over even longer time intervals.

It seems reasonable to assume that the value of Q used in Equation 1 or 2 will represent flow over a period of at least 5 or 6 min, although the meter's response is updated much more frequently. Consequently, the primary criterion for evaluating the accuracy of a particular strategy is the difference between the output and the target flow, averaged over a period of 5 min or more; a secondary criterion is the variance or standard deviation of the differences measured at the minimum response interval, since this reflects the accuracy of the response at the minimum interval.

Let $e(t)$ represent the response error for time period t; $e(t)$ is defined as the actual output flow minus the flow target, that is

$$e(t) \equiv q_o(t) - Q \tag{3}$$

where q_o represents the actual output flow. Three sources of response error are to be expected. These are (a) predictive error e_q, which results from the difference between q and $\hat{q}$, (b) response error e_r, which results from nominal metering rates falling outside the limits M_{max} and M_{min}; and (c) response error e_p, which results from the differences between $M + \Delta$ or rM and the actual ramp flow. If Equation 2 is used to determine the metering rate, these are

$$e_q(t) = q(t) - \hat{q}(t) \tag{4}$$

$$e_r(t) = \begin{cases} r(t)M_{min} - Q(t) + \hat{q}(t) & \text{if } [Q(t) - \hat{q}(t)]/r(t) < M_{min} \\ r(t)M_{max} - Q(t) + \hat{q}(t) & \text{if } [Q(t) - \hat{q}(t)]/r(t) > M_{max} \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

and

$$e_p = p(t) - r(t)M(t) \quad (6)$$

The overall error in the output is the combination of these errors.

Now consider the ways in which these various sources of error might combine to produce the distribution of flow downstream of the ramp. In cases in which all predicted flows fall within the response limits, flow downstream of the meter is

$$q_o = Q + e_q + e_p$$

If the parameters of the predictive model are correctly estimated, both e_q and e_p should have zero mean. Also, there is every reason to believe that they are statistically independent, so their variances should sum. Thus, the distribution of q_o should have mean Q (the target flow) and variance $VAR(e_q) + VAR(e_p)$.

If some predicted flows fall outside the response limits, matters become more complicated. In this case

$$q_o = Q + e_q + e_p + e_r$$

but e_r does not have zero mean. Instead, it has negative mean when the calculated value of M exceeds M_{max} and positive mean when the calculated value of M is less than M_{min}. Also, e_r is not independent of e_q; instead, they should be negatively correlated. For instance, if $\hat{q}$ is too high, there will be a negative error e_q; meanwhile, there will be increased probability that the calculated value of M will be less than M_{min}, which results in a positive error e_r. Since e_r exists only when the meter's response is M_{max} or M_{min}, the distribution of e_r is a combination of the conditional distribution of q, given that $(Q - \hat{q})/r > M_{max}$ and the conditional distribution of q given that $(Q - \hat{q})/r < M_{min}$. As such, it depends on the dispersion and shape of the distribution of the actual upstream count q and on the relative probabilities that the calculated meter responses fall outside the limits M_{max} and M_{min}.

METER RESPONSE SIMULATION

The main objective of this study was to evaluate a minor proposed modification to the San Diego ramp metering system. The system in question is a multiramp traffic-responsive system, in which all field controllers communicate with a central computer. At the time of the study, the system consisted of approximately 100 controllers; most were grouped into six major subsystems of 7 to 22 controllers each. In addition, there were a number of isolated controllers or small groups. Most controllers are used to operate meters, although a few only collect data.

At present, coordination between the various metering locations consists of a set of flow targets that are intended to hold flows at the bottlenecks to their capacities. These are preset, having been determined over time by trial and error. In a sense, they correspond to the optimization layer in Papageorgiou's proposed system (*5,6*), although no actual optimization is involved. In addition, there is a limited adaptive capability: under certain circumstances, high occupancies will cause restrictive metering rates to be passed upstream.

Local traffic-responsive control is of the demand-capacity variety. Detectors, as is common in demand-capacity systems, are located just upstream of the on-ramps. At present, the upstream flow is calculated as a moving average of several past 30-sec counts, with the number of counts varying by location. Meter cycles (and hence metering rates) vary in discrete steps, with the number of steps depending on the location. Each ramp has maximum and minimum rates, and if the calculated metering rate falls outside these limits, it is set at the appropriate limit. In addition, when high occupancies indicate that a mainline queue is present, this calculation is overridden and the metering rate is set at its minimum value. In contrast to most systems, San Diego does not have maximum ramp queue length constraints; instead, minimum metering rates are set to control the growth of ramp queues.

One alternative to the existing strategy is to base the estimate of $\hat{q}$ in Equation 1 or 2 on data from upstream meters. Under uncongested conditions, average speeds in San Diego have been found to vary only slightly with flow, and there seems to be an emerging consensus that this is true of most North American data (*21,22*). If this is the case, it should be relatively easy to project variations in uncongested flow downstream. It was proposed to do this by using a moving average of flow at an upstream meter, offset from the response interval by the travel time between the two meters. The estimated flow was further modified by a factor intended to account for count biases and expected flows at intervening off-ramps (off-ramp flows are generally not counted in San Diego) and by adding actual or expected flows at intervening on-ramps. Details of the proposed evaluation scheme and the calibration of the parameters of the flow model may be found elsewhere (*23*).

The proposed control system modification was evaluated by simulating the performance of alternative estimates of $\hat{q}$ based on data from three freeway segments in San Diego. Figures 1 and 2 are schematic diagrams showing lane configurations, distances between detectors, and approximate maximum flow rates for selected locations in these sections. The parameters of the models were calibrated for these sections, and actual 30-sec counts from the peak periods of three different days were used to calculate values of $\hat{q}$ at the downstream end of each section. These were compared with actual 30-sec volume counts. The response of a meter at or near the downstream end of each section was then simulated on the basis values of Q set for these locations. In the case of Sections 1 and 3, no bottlenecks were present, and values of Q were set arbitrarily. In the case of Section 2, the Q-value was set to approximate the capacity of the College Avenue bottleneck; it should be noted, however, that what was simulated was the output of the College Avenue meter, although past studies indicate that the bottleneck is probably just upstream of the on-ramp (*15*). The output of the meter was estimated to be

$$\hat{q}_o(t) = q(t) + r(t)M(t)$$

where $q(t)$ represents the actual 30-sec count for time period t; $e(t)$ was then estimated to be $\hat{q}(t) - Q$. Metering rates between the maximum and minimum limits were assumed to vary continuously, thus ignoring the discrete cycle lengths used in practice.

Since the simulated metering rates were not necessarily the same as the actual rates in the data, it was not possible to

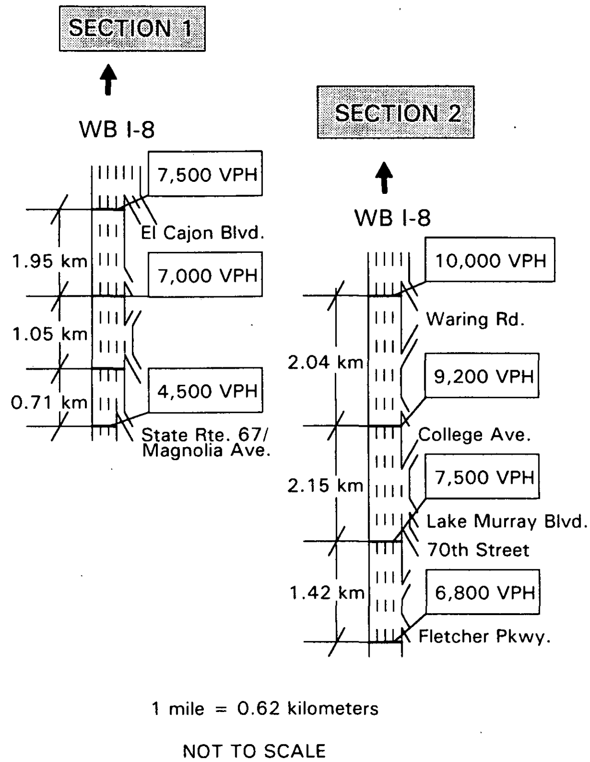

FIGURE 1 Schematic diagrams of Interstate 8 study sections.

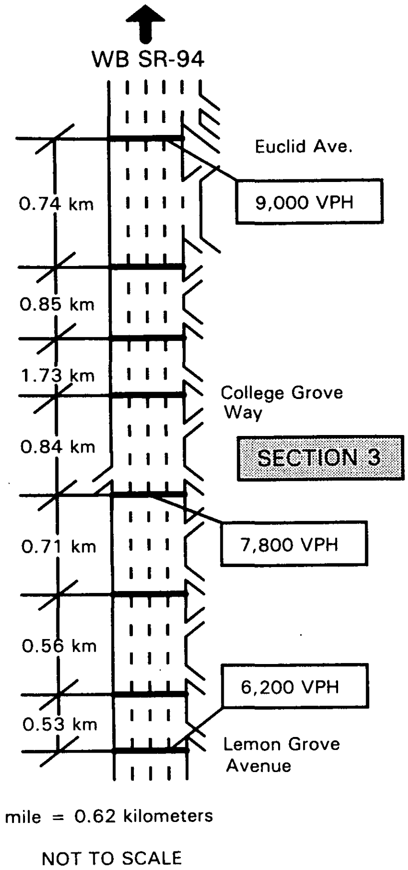

FIGURE 2 Schematic diagram of State Route 94 study section.

estimate e_p. Also, given the complexities of the error distribution in the case in which some values of $\hat{q}$ fall outside the response limits, no attempt was made to analyze the distribution of e_r in detail. Instead, alternative ways of estimating $\hat{q}$ were compared on the basis of the distributions of e_q and of the combined errors $e_q + e_r$. Means and standard deviations were calculated for e_q and the sum $(e_q + e_r)$ for 6-min intervals, and the mean of e_r was calculated by subtracting the mean of e_q from the sum. Also, these quantities were calculated for 30-min intervals during which mainline flows produced by the existing system are nearly constant and at their maximum values.

Estimation schemes based on upstream data were compared with those based on local data. Also, within each category, different averaging intervals were used to determine the effect of data smoothing. Finally, upstream-based schemes using data from different locations were compared.

RESULTS

All models were reasonably good at estimating $\hat{q}$ over periods of 5 to 6 min. Mean values of e_q, when averaged over 6 min, ranged from near 0 to about 1.7 percent of the target flow, and in most cases were less than 1 percent. Given the corresponding variances of e_q, these means do not appear to be significantly different from zero.

Ability to predict individual 30-sec counts, on the other hand, was low in all cases. This is indicated by relatively high standard deviations of the distributions of e_q. When calculated over 30-min intervals, the standard deviation of e_q ranged from 7 percent up to about 20 percent of the mean flow, with most cases in the range of 10 to 15 percent.

Mean values of e_r were highly variable and quite sensitive to the relationship between the average input flow and the target flow, the response range of the meter, and daily variations in the time series of the input flow. On the whole, they tended to be similar for both upstream-based and locally based models. For both types of model, their absolute values tended to increase as averaging intervals decreased.

Absolute values of the mean of e_r, calculated over 30 min, ranged from 0 to around 7 percent of the target flow. They tended to be least when the mean of the input flow was "centered" in the response range of the meter—that is, when the mean input flow was approximately equal to $Q - r(M_{max} + M_{min})/2$—and to increase as the mean input flow approached $Q - rM_{max}$ or $Q - rM_{min}$. The response error e_r tended to dominate the total error in cases in which the mean input flow was not centered in the response range and led to average total errors of roughly 2 to 5 percent of the mean flow over periods as long as 30 min.

EFFECTS OF RESPONSE LIMITS

To understand why the output is so sensitive to response errors, it is necessary to realize that the normal random variation in 30-sec counts almost always exceeds the difference between the maximum and minimum metering rates. Figure 3 is a typical example of a time series of 30-sec counts from the San Diego system. The section in question has four lanes in one direction, and a maximum average flow of around 57 vehicles per 30 sec or 6,800 vehicles per hour. (The comparatively low flow is because the location in question is some distance upstream of the bottleneck.) Note that the individual counts fall in a band that is about 20 to 30 vehicles wide. This corresponds to a variation in flow rate of around 2,400 to 3,600 vehicles per hour. Meanwhile, differences between maximum and minimum metering rates for meters in the sections studied ranged from 1.2 to 8.5 vehicles per 30 sec or 144 to 1,020 vehicles per hour, with a range of around 5 vehicles per 30 sec or 600 vehicles per hour being most common. As a result, meters are responding with maximum or minimum rates most of the time; in the case of the simulations described here, 45 to 96 percent of the responses were either M_{max} or M_{min}, depending on the flow model, the mean input flow, and the response range of the meter.

When the input flow is centered in the meter's response range, M_{max} and M_{min} occur about equally, and the response errors tend to cancel out. When mean input flow approaches $Q - rM_{min}$, M_{min} predominates and the output is biased high; the opposite happens when q approaches $Q - rM_{max}$. Consequently, the main difficulty in providing precise control of bottleneck flows is the limited response of the meter in the face of highly variable mainline flows and not the accuracy of the estimate $\hat{q}$ on a count-by-count basis.

In fact, overly accurate prediction of 30-sec counts is counterproductive, as is evidenced by the fact that the short-averaging-interval upstream-based models, which produced the best estimates of individual 30-sec counts, resulted in the largest response errors. Consider the case of a perfectly accurate estimator of 30-sec flow attempting to respond to a mean input flow of $Q - rM_{min}$. When the 30-sec count exceeds the average value, the meter responds with M_{min}, and the output exceeds Q. When the 30-sec count is less than average, it responds (correctly) with something more than M_{min}, and the output is Q. Consequently, these will be a substantial positive bias in the output. An estimator that was so heavily smoothed that it always estimated the next 30-sec count to be equal to the long-term average of q, on the other hand, would result in roughly half the output counts being less than Q and half being greater, so that the average output would be approximately Q.

In addition to the phenomenon just described, other limitations to the system's ability to respond were observed. For instance, it is normally assumed that actual flow from a metered ramp will be approximately equal to the metering rate or that the two may be related by a fairly stable factor, such as r in Equation 2. In calculating r values for the San Diego system, it was discovered that they not only varied by time of day, but that, at some locations, they also were often significantly less than 1.0 throughout the peak. Of the 16 ramps included in the sections studied here, 6 had r values of less than 0.95 throughout the peak or exceeded 0.95 for no more than 5 or 6 min; several other meters in Section 3 exceeded 0.95 for a total of 15 to 30 min during a 4-hr potential metering period.

This appears to indicate that ramp demand is less than the minimum metering rate at many locations. This was particularly true where ramp spacings were fairly close (as would be expected) and where there were high-occupancy vehicle (HOV) lane bypasses around ramp queues. The San Diego system, in contrast to some other systems, does meter HOV lanes, but it appears that demand for HOV lanes is rarely equal to the minimum metering rate.

In Figures 4 through 6, ramp passage counts are compared with metering rates for a variety of situations. Figure 4 shows the beginning of metering on a fairly high-volume ramp. It

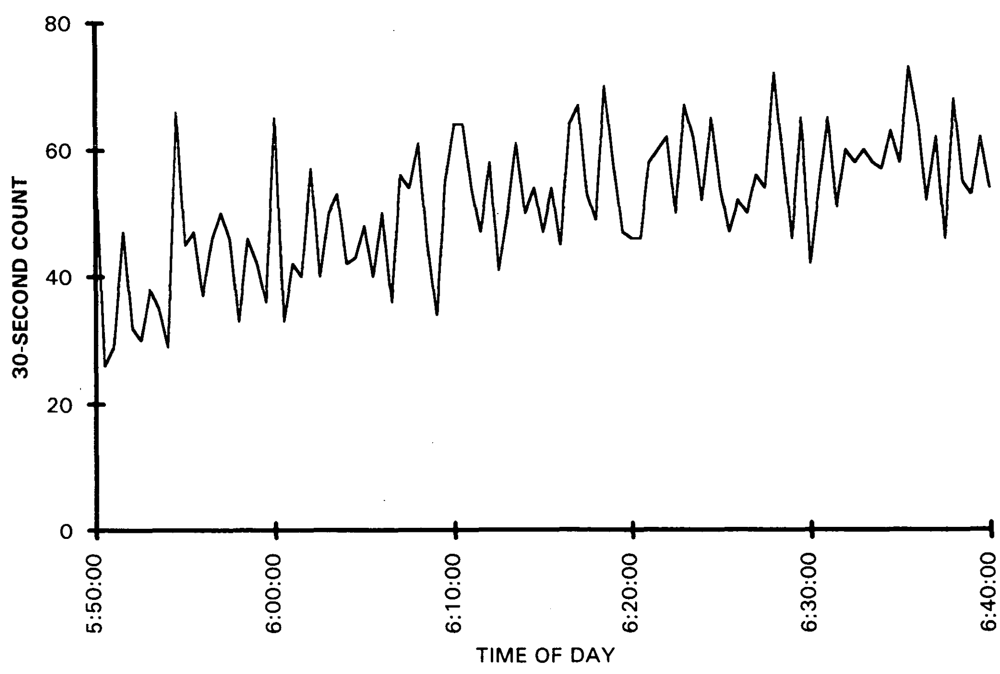

FIGURE 3 30-sec count versus time, westbound Interstate 8 at Fletcher Parkway, June 11, 1991.

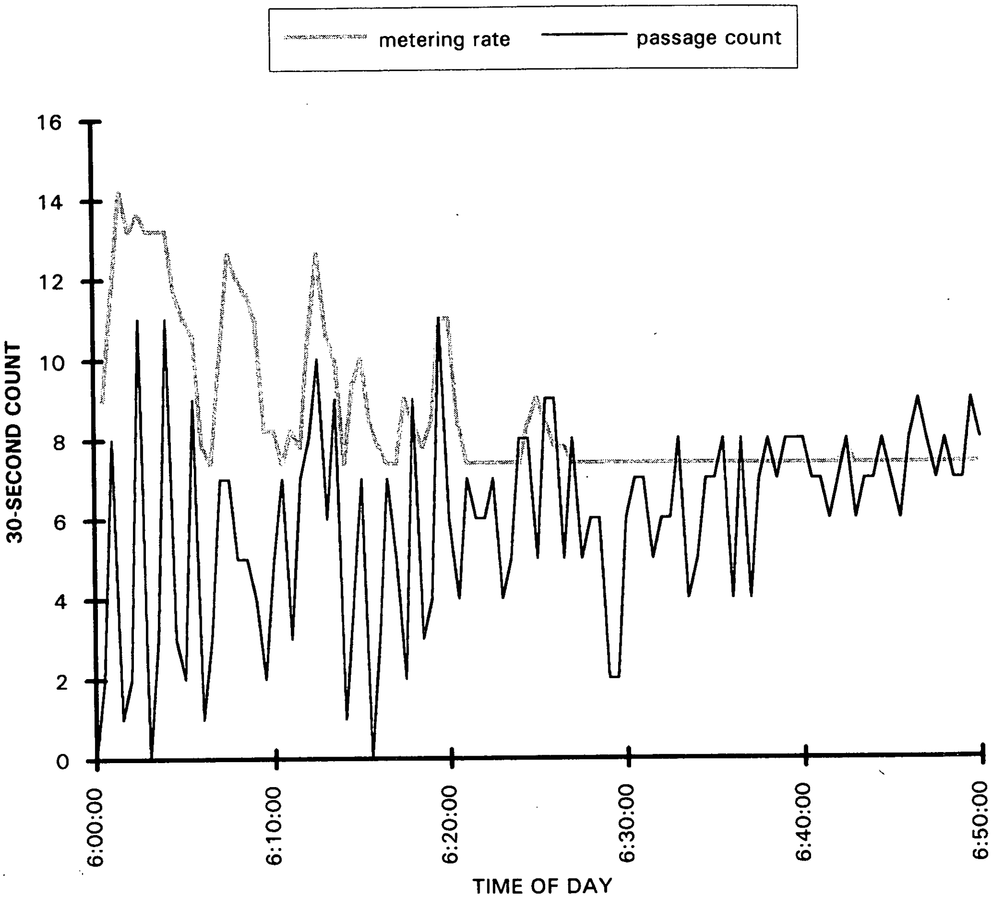

FIGURE 4 Theoretical metering rate versus passage count, Fletcher Parkway on-ramp to Interstate 8, August 5, 1991.

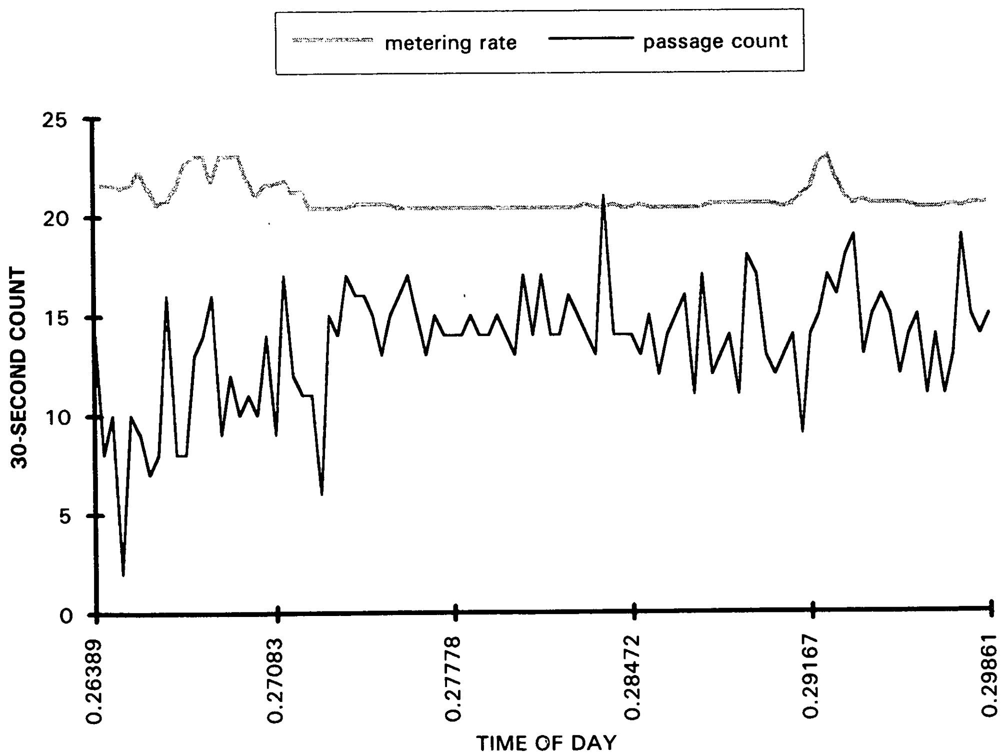

FIGURE 5 Theoretical metering rate versus passage count, 70th Street/Lake Murray Boulevard on-ramp to Interstate 8, August 5, 1991.

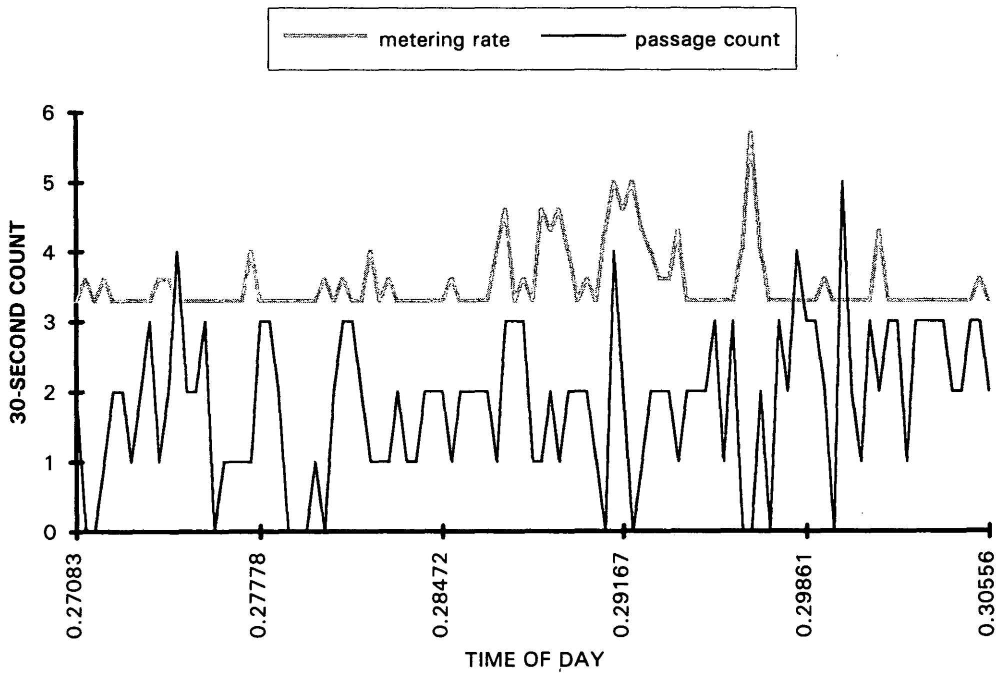

FIGURE 6 Theoretical metering rate versus passage count, College Grove Way on-ramp to State Route 94, August 5, 1991.

takes about 20 min for the metering rate to settle down to its minimum value and an additional 15 min or so for the passage count to settle down to the metering rate; however, it eventually does so. Figure 5 represents a high-volume ramp with an HOV lane, and Figure 6 represents a low-volume ramp with an HOV lane. In both of these cases, the average passage count is substantially less than the metering rate and there is substantial variation on a count-by-count basis.

Finally, a major response limitation in San Diego appears to be a lack of adequate total metering capacity, despite an extensive system. The portions of the system studied involved inbound flow during the morning peak and metering, which extends to the outer edge of the metropolitan area. It was observed that all meters throughout this portion of the system were responding with their minimum rates during the most intense portions of the peak, but that this was not always sufficient to prevent flow breakdown. Obviously, if all meters must be at their minimum rates for extended periods of time to prevent flow breakdown at the bottlenecks, there can be no real traffic-responsiveness during those periods.

IMPLICATIONS FOR METERING STRATEGIES

The preceding section identified three ways in which the San Diego ramp metering system is limited in its ability to respond to variations in traffic flow. These limitations, which may well apply to other traffic-responsive multiramp systems, include insufficient total metering capacity, a significant number of ramps at which demand is less than the minimum metering rate, and meter response ranges that are considerably less than the normal variation in mainline volumes at the minimum count interval. The following discussion is focused on the last of these, since not much can be done about the other two unless minimum metering rates can be reduced.

A traffic-responsive metering strategy may be adjusted in three ways to reduce the output biases that result from frequent use of maximum or minimum metering rates. They are as follows: (a) set flow targets at the upstream meters so that the input flow at the bottleneck will be centered in the meter's response range; (b) smooth the flow estimate $\hat{q}$ to reduce the response to short-term input flows that do fall within the response range; or (c) use an offset value for the flow target (i.e., use some value other than the actual output flow target for Q in Equation 1 or 2).

Of these three modifications, only the first shows much promise of success. The simulations showed that where input flows were not well-centered in the response range, smoothing of flow estimates based on local data by increasing the averaging interval from 30 sec to 6.5 min will normally reduce the response error by about 30 to 50 percent. The remaining errors, however, are still on the order of 2 to 5 percent of the target flow, and there will almost always be an increase in the variance of the output.

The possibility of using offset flow targets was not investigated in detail. Offset targets could be determined by simulations in which the target in the response calculation was varied until the desired mean output was achieved. This would be unlikely to produce consistently good results, however, because the simulations also showed considerable variation in the response error from day to day when input flows were not centered in the response range.

Setting upstream flow targets centered in the meter's response range to provide input flows at the bottleneck could be somewhat more effective. Based on the simulations performed here, this method shows promise of being able to hold mean output flows (averaged over 30 min) to within 1.5 percent of their targets. In order to do this, it would probably be necessary to set all upstream targets based on the assumption that the mean metering rate at every ramp downstream was approximately halfway between M_{max} and M_{min}. By doing this, flows could be held close to their targets throughout the system.

Unfortunately, this could create a another problem if there were an attempt to implement the optimization layer of Papageorgiou's hierarchical control system (*5,6*). In this scheme, flow targets are set based on the steady-state linear programming formulation of Wattleworth (*7,8*), which is unlikely to result in metering rates halfway between M_{max} and M_{min}. Obviously, there may be a trade-off between the objective of precise control of flow at the bottleneck and that of minimization of delay for the system as a whole, which is the basis of the Wattleworth strategy.

CONCLUSION

Described in this paper is a study in which simulations of ramp meter responses were used to explore the feasibility of replacing the locally based moving-average estimates of mainline flow used by the San Diego ramp metering system with estimates based on upstream data. These simulations showed that choice of estimation methodology made little practical difference in the performance of the system and that the major problem in using ramp metering to provide precise control of bottleneck flows is the limited ability of the system to respond and not the accuracy of the flow estimate. The most important limitation appears to be the limited range of response of individual meters, which is typically much less than the normal variation in flow at the minimum count interval of 30 sec used in San Diego. This leads to a situation in which meters respond with their maximum or minimum rates much of the time and in which mean output flows will be biased unless mean input flows are centered in the response range. Other response limitations include a comparatively large number of ramps at which demands are less than minimum metering rates and insufficient total metering capacity.

The most promising way of preventing biases in the mean output of meters is to set flow targets for upstream meters to provide mean mainline input flow at the bottleneck approximately equal to the output target minus $(M_{max} + M_{min})/2$. This strategy will normally require that mean metering rates at all meters be approximately halfway between M_{max} and M_{min}. It would be ineffective for systems such as that in San Diego, however, because at present all meters are required to operate at their minimum rates for much of the peak in order to hold flows at bottlenecks close to their capacities. It is also likely that this strategy conflicts with any strategy that sets metering rates based on some other criterion, such as minimization of delay.

These findings raise several issues requiring further research. One of these is how best to resolve conflicting objectives in hierarchical ramp metering control systems. A second issue is that of the relationships between measurable (and potentially controllable) characteristics of the traffic stream and flow breakdown. For instance, how important to the process of flow breakdown is the variation in volume counts over intervals such as 30 sec as opposed to the mean flow over periods of 5 or 6 min? The goal of this research should be clarification of the goals of the control system in terms of the type of flow it should be attempting to produce. A third issue is how metering strategies other than those simulated here are affected by response limitations. For instance, is the reported superiority of the occupancy-based feedback law ALINEA due to reduced vulnerability to the effects of response limits (which seems unlikely) or to some other cause, such as faster response to the onset of congestion?

ACKNOWLEDGMENTS

This research was funded by the State of California Department of Transportation, District 11. The contents of this paper reflect the views of the author, who is responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the State of California. This report does not constitute a standard, specification, or regulation. Special thanks are due to research assistant Sadam Panduranga.

REFERENCES

1. D. R. Drew. Gap Acceptance Characteristics for Freeway-Ramp—Surveillance and Control. In *Highway Research Record 157,* HRB, National Research Council, Washington, D.C., 1967, pp. 108–135.
2. L. Isaksen and H. J. Payne. Suboptimal Control of Linear Systems by Augmentation with Application to Freeway Traffic Regulation. *IEEE Transactions on Automatic Control,* Vol. AC-18, 1973, pp. 210–219.
3. N. B. Goldstein and K. S. P. Kumar. A Decentralized Control Strategy for Freeway Regulation. *Transportation Research,* Vol. 16B, 1982, pp. 279–290.
4. D. P. Looze, P. K. Houpt, N. R. Sandell, and M. Athans. On Decentralized Estimation and Control with Applications to Freeway Ramp Metering. *IEEE Transactions on Automatic Control,* Vol. AC-23, 1978, pp. 268–275.
5. M. Papageorgiou. A Hierarchical Control System for Freeway Traffic. *Transportation Research,* Vol. 17B, 1983, pp. 251–261.
6. M. Papageorgiou. Multilayer Control System Design Applied to Freeway Traffic. *IEEE Transactions on Automatic Control,* Vol. AC-29, 1984, pp. 482–490.
7. J. A. Wattleworth and D. S. Berry. Peak Period Control of a Freeway System—Some Theoretical Investigations. In *Highway Research Record 89,* HRB, National Research Council, Washington, D.C., 1965, pp. 1–25.
8. J. A. Wattleworth. Peak-Period Analysis and Control of a Freeway System. In *Highway Research Record 157,* HRB, National Research Council, Washington, D.C., 1967, pp. 10–21.
9. A. D. May. A Proposed Dynamic Freeway Control System Hierarchy. *Proc., 3rd IFAC/IFIP/IFORS International Symposium on Control in Transportation Systems,* Columbus, Ohio, 1976, pp. 1–12.
10. M. Papageorgiou. A New Approach to Time-of-Day Control Based on a Dynamic Freeway Control Model. *Transportation Research,* Vol. 14B, 1980, pp. 349–360.
11. R. L. Pretty. Control of a Freeway System by Means of Ramp Metering and Information Signs. In *Highway Research Record 388,* HRB, National Research Council, Washington, D.C., 1972, pp. 62–69.
12. H. Hadj-Salem, J.-M. Blosseville, and M. Papageorgiou. ALINEA: a Local Feedback Control Law for On-Ramp Metering—A Real Life Study. *Proc., 3rd IEE International Conference on Road Traffic Control,* London, England, 1990, pp. 194–198.

13. S. Smulders and F. Middelham. *Isolated Ramp-Metering: Real Life Study in the Netherlands.* DRIVE-Office, Brussels, Belgium, 1991.
14. M. Papageorgiou, H. Hadj-Salem, J.-M. Blosseville. ALINEA: A Local Feedback Control Law for On-Ramp Metering. In *Transportation Research Record 1320,* TRB, National Research Council, Washington, D.C., 1991, pp. 58–64.
15. J. H. Banks. Flow Processes at a Freeway Bottleneck. In *Transportation Research Record 1287,* TRB, National Research Council, Washington, D.C., 1990, pp. 20–28.
16. J. H. Banks. The "Two-Capacity" Phenomenon at Freeway Bottlenecks—A Basis for Ramp Metering? In *Transportation Research Record 1320,* TRB, National Research Council, Washington, D.C., 1991, pp. 83–90.
17. J. H. Banks. The "Two-Capacity" Phenomenon: Some Theoretical Issues. In *Transportation Research Record 1320,* TRB, National Research Council, Washington, D.C., 1991, pp. 234–241.
18. F. L. Hall and K. Agyemang-Duah. Freeway Capacity Drop and the Definition of Capacity. In *Transportation Research Record 1320,* TRB, National Research Council, Washington, D.C., 1991, pp. 91–98.
19. T. Urbanik II, W. Hinshaw, and K. Barnes. Evaluation of High-Volume Urban Texas Freeways. In *Transportation Research Record 1320,* TRB, National Research Council, Washington, D.C., 1991, pp. 110–118.
20. *Special Report 209: Highway Capacity Manual.* TRB, National Research Council, Washington, D.C., 1985.
21. J. H. Banks. Freeway Speed-Flow-Concentration Relationships: More Evidence and Interpretations. In *Transportation Research Record 1225,* TRB, National Research Council, Washington, D.C., 1989, pp. 53–60.
22. F. L. Hall, V. R. Hurdle, and J. H. Banks. A Synthesis of Recent Work on the Nature of Speed-Flow and Speed-Occupancy (or Density) Relationships on Freeways. Presented at 71st Annual Meeting of the Transportation Research Board, Washington, D.C., 1992.
23. J. H. Banks. *Predictive Flow Model for Ramp Metering.* Final Report, Civil Engineering Report Series 9210. San Diego State University, San Diego, Calif., 1992.

Publication of this paper sponsored by Committee on Freeway Operations.

Development of a Freeway Traffic Management Project Through a Public-Private Partnership

STEVEN Z. LEVINE, WILLIAM R. MCCASLAND, AND DENNIS G. SMALLEY

If intelligent vehicle-highway system (IVHS) and traffic management systems are to be successful in the future, a good relationship must exist between the public and private sectors. To foster this relationship in the Houston area, a public-private sector partnership was organized by the Texas Department of Transportation. The initial objective of this group was to improve the sources and use of real-time travel information. This partnership consisted of public-sector transportation agencies, commercial transportation companies, and traffic service organizations. The initial project of the partnership was the development and operation of a cellular phone probe project as a source of real-time travel information. This information was then used by commercial transportation companies and traffic service organizations. This project demonstrated how the public, public agencies, and the private sector can have effective roles in improving the sources and use of travel information.

To obtain the maximum benefits from intelligent vehicle-highway systems (IVHS) and traffic management systems (TMS), the public and private sectors must have good working relationships. To encourage the formation of a cooperative relationship at the local level, the Houston Office of the Texas Department of Transportation (TxDOT) organized a partnership of public agencies and private-sector companies with the initial objective of increasing the sources and the use of real-time travel information. Public-sector participants were the City of Houston Aviation, Police, and Transportation departments; Harris County Metropolitan Transit and Toll Road authorities; Texas Department of Transportation; and Texas Transportation Institute. Private-sector participants were Federal Express, Houston Lighting and Power, METRO Traffic Control, ATE Bus Lines, Shadow Traffic, and Traffic Central. During monthly meetings the partners defined issues that provided the basis for a short-term joint public-private implementation project in an area called the North Houston Corridor. These issues included the following:

- Travel time information for the study area was needed to operate eight changeable message signs (CMSs). The schedule for installation of these signs was accelerated when the Harris County Toll Road Authority (HCTRA) received funding.
- Travel time information was scheduled to be provided by TMSs, which would be operational within 2 to 3 years.
- In the North Houston Corridor one of the major roadways, the US 59 Eastex Freeway, was to be reconstructed during a 5-year period. The monitoring of roadway and traffic conditions was essential for the implementation of congestion mitigation efforts.
- Although the original purpose of the partnership was of a planning and technology sharing nature, it was decided that a short-term, successful implementation project for demonstration purposes would be needed to maintain the interest of all partners.

The North Houston Corridor offered the best potential for satisfying the requirements raised in these issues and for fostering the public- and private-sector cooperation needed for implementing and participating in the project. The project involved use of a technology for tracking vehicles through the corridor so that travel times could be monitored in real time. Automatic vehicle identification (AVI) and automatic vehicle locator systems were considered the best technologies, but a technique using manual data collection and transmission was selected for the first phase to meet the schedules for early implementation. Cellular telephones were issued to 200 volunteers who were asked to serve as traffic reporters (probes) during their commute trips to and from work. The experiences in involving the public, public agencies, and the private sector in the development and conduct of the study are reported in this paper. The results of the cellular telephone demonstration study as a source of travel time information are also discussed.

FACTORS IN SELECTING THE PROJECT CORRIDOR

The North Houston Corridor study area is defined by the three parallel radial freeways: IH 45 North Freeway, the Hardy Toll Road, and US 59 Eastex Freeway (Figure 1). The High-occupancy vehicle (HOV) lane on IH 45 was also included in the study. Several factors favored the selection of this corridor:

- The three parallel freeways are spaced such that they can serve as alternate routes for each other.
- The US 59 Eastex Freeway was scheduled for reconstruction beginning in 1992, and installation of a TMS would not begin until completion of construction in 3 to 4 years.
- Eight electronic CMSs are installed in the corridor for traffic diversion operations (Figure 2). These signs were in-

S. Z. Levine, Texas Department of Transportation, P.O. Box 1386, Houston, Tex. 77251. W. R. McCasland and D. G. Smalley, Texas Transportation Institute, 701 North Post Oak, Suite 430, Houston, Tex. 77024.

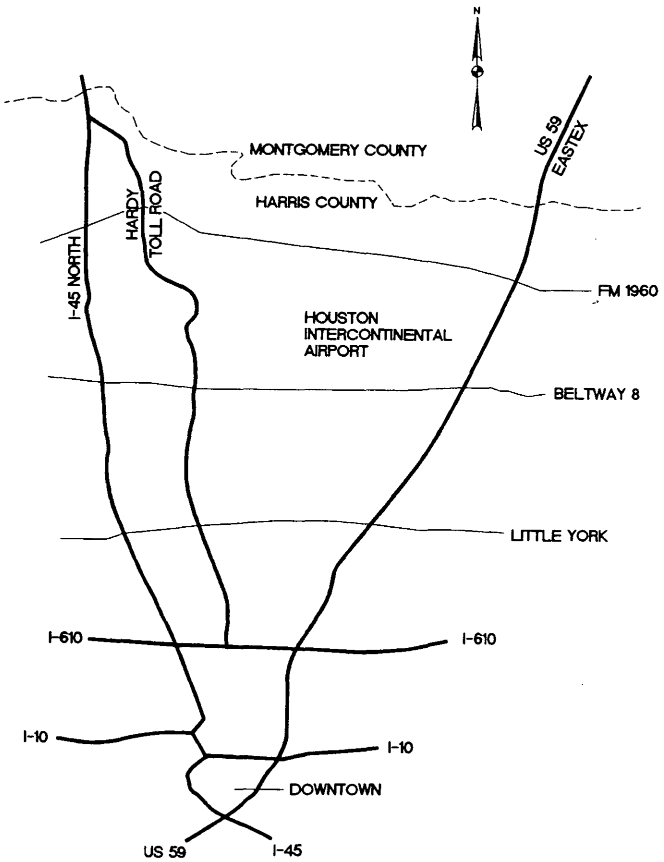

FIGURE 1 North Houston Corridor.

stalled with funding assistance from HCTRA, which operates the Hardy Toll Road.

• The TMS being installed on IH 45 North Freeway is not scheduled to be operational for 2 years. No traffic management or monitoring systems are planned for the Hardy Toll Road.

• The North Houston Corridor serves several major activity centers, including
 - Houston Intercontinental Airport,
 - Woodlands and Kingwood residential communities,
 - Downtown Houston and the Greenway Plaza employment centers, and
 - Houston Medical Center.

• The North Houston Corridor carries a high volume of commercial traffic.

DEVELOPING THE CELLULAR TELEPHONE PROJECT

Several developments in transportation should be of interest to the private sector: (a) the potential impacts of the Clean Air Act on the operation of businesses; (b) the potential for new business opportunities resulting from the development of IVHS technologies; (c) the improvement in the efficiencies of operations with the availability of real-time information on traffic conditions and navigation information for guidance. Therefore the development of the public-private partnership was successful in enlisting companies representing the following business groups:

- Commercial transportation companies, including contract carriers, taxi cab companies, and package and long-haul trucking companies;
- Traffic advisory service companies;
- Cellular telephone companies;
- Utility companies; and
- Greenway Plaza Property Transportation Center.

The discussions conducted at the monthly meetings of the partnership were useful in defining the interests and potential roles of the various companies in developing the IVHS program in Houston. For example, although the commercial transportation companies were interested in the development of IVHS and expected to participate in future IVHS projects, they were not interested in making financial investments in demonstration and development projects at this time. Further, it was determined that the number and time of trips by commercial vehicles in the study corridor during the commute time periods would not provide adequate coverage in the collection of real-time travel time data.

The traffic advisory service organizations were most interested in participating in the project because they had immediate needs for the information to be collected and could recognize the potential business that would result from the implementation of IVHS. It was determined that their support role would be to provide manpower and equipment to operate the system instead of straight financial support.

The need for the implementation of a short-term project to examine and demonstrate the benefits of real-time information became evident. The private sector needed assurance that their operations would realize real benefits. The use of cellular telephone technology with volunteer probe vehicles was determined to require the least capital expenditure. However there were major operating costs, primarily the cellular telephone air time for the traffic reports, for which private-sector support was necessary.

Cellular Telephone System

The probe study required 200 vehicles equipped with cellular telephones to provide the desired coverage of the four routes. The reporters (probes) would make 10 telephone calls per day, the cost for which would accrue to $180,000 annually for air time. Other costs include the 200 cellular telephones and the monthly charges for cellular telephone service.

The partnership obtained support from the Houston Cellular Telephone Company to provide free air time for project calls. The sponsoring agency, TxDOT, funded the cost of the telephone equipment and the monthly access charge for telephone service, which was $29.75 per month per telephone.

Office Staff

The telephone calls were received by three operators who recorded the information directly into the computer system. Two of the operators were provided by the two traffic advisory

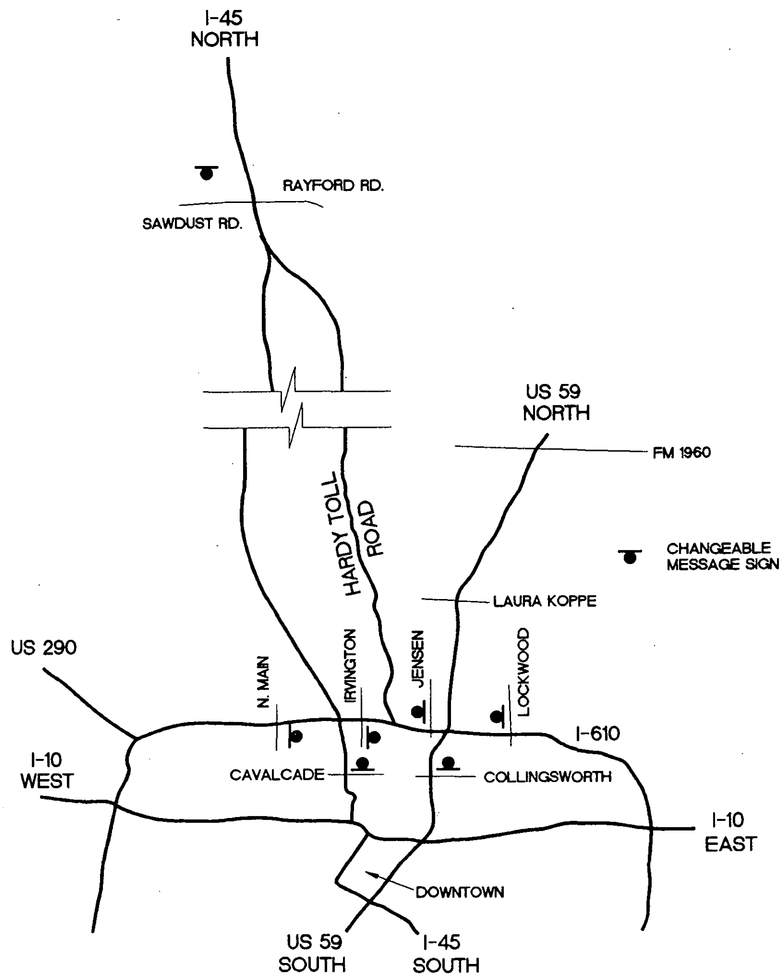

FIGURE 2 CMS locations.

services in Houston, METRO Traffic and Shadow Traffic, which market traffic information to radio and television stations. The third operator was provided by the study. A third traffic advisory service, Infobanq, provided equipment in lieu of personnel support. Supervision of the operators was provided by the Texas Transportation Institute (TTI) under contract to TxDOT.

Office Facilities

Office space of approximately 1,000 ft^2 was required to house the operators and computer equipment. Senterra Development Corporation, manager of the Greenway Plaza Properties, provided the office space at cost (approximately 50 percent of the leasing rate). An additional 1,000 ft^2 was acquired to house the TxDOT's interim communications center, which uses the real-time information to operate the CMSs and dispatches emergency vehicles to incidents.

Computer Network

A network of six computers, three for data input by the operators, are required to receive, process, and disseminate the information. Some additional equipment, such as printers, fax machines, and networking devices are also needed. This equipment was furnished by the Southwest Regional Transportation Center of the Texas A&M University in exchange for the opportunity to conduct several studies that involve the information obtained in the cellular telephone project.

Project Management

A project staff was required to develop the project, acquire the equipment, furnish the office, develop the computer software, and supervise the operators, probes, and other personnel involved in the study. TTI provide this function under contract to TxDOT. An existing interagency agreement between TTI and TxDOT was used, which enabled this project to become operational within 1 year.

Although support by the private sector was eventually received, it became apparent that the public sector's perception of what the private sector's role should be in such a project was not accurate. For example, because cellular telephones are often advertised at prices as low as $29.00 per unit, the staff expected that the equipment could be obtained at no cost to the project. This proved to be incorrect, and through a competitive bid procedure, 200 units were purchased at a cost of $105 per unit. Both cellular telephone companies in Houston were invited to participate in the project, but only Houston Cellular elected to participate. Houston Cellular provided free air time for the traffic reports and provided a good rate schedule for personal calls equal to a government plan. However the project was required to pay the monthly access fees of $29.50 per month per telephone for the 200 units.

RECRUITMENT OF VOLUNTEER TRAFFIC REPORTERS (PROBES)

Tracking vehicles along streets and highways has been a standard technique for measuring travel characteristics. The floating car method with various techniques for recording the time intervals between check points has been improved over the years with the application of automatic sensing devices and recorders on the vehicle. Laptop computers that can plug into the vehicle electronics have made the study technique practically independent of the operator of the vehicle. However to get real-time information over a large area for a long time period, the information must be collected at frequent intervals from many vehicles. It is impractical to equip and to operate many vehicles on a permanent basis to collect this information. However the development of automatic systems that can identify a vehicle or determine its location is moving forward because there are other applications for the data collected from these systems. These applications include bus locator systems and small fleet management systems to aid in the dispatching functions, vehicle locator systems for the retrieval of stolen vehicles, and identification systems for automatic collection of tolls. Other IVHS applications that are becoming more important include vehicle navigation systems, advanced information systems, and TMSs.

In the cellular telephone study for the North Houston Corridor, the data requirements were time of day, vehicle identification, and vehicle location at intervals of 3 or 4 mi on the four roadways, which were approximately 20 mi long. The proposed average time interval between vehicles was 5 min during the commute time periods, and 2 to 3 min during the critical peak periods. The challenge to the study staff was to contact potential volunteers, explain the study, and enlist 200 volunteers for a minimum of 1 year.

The study staff contacted several major employers in Houston to solicit their help in contacting their employees. Arrangements were made to visit the companies to present a brief slide presentation and to obtain the names of people interested in participating in the project. The presentation appealed to their civic duty to help solve the traffic congestion problem, but the project offered real incentives to compensate them for their time and efforts in making the reports:

- A mid-priced cellular telephone (Motorola Model MC310 or NEC Model 3800, installed or transportable);
- Free installation;
- No monthly access charges for 1 year;
- Free air time for traffic-related calls; and
- Government rates for air time charges for personal calls.

Some of the companies, agencies, and groups that participated in the program include Exxon, First City Bank, Houston Medical Center, and TxDOT. Many individuals were contacted through newspaper advertisements and word of mouth. The major problem in obtaining the 200 volunteers was the distribution of trips in time and space. Table 1 depicts the distribution of the probes by time of day and roadway. It was difficult to enroll persons for some time slots on some roadways. Another problem that eliminated several potential participants was that each volunteer had to have a satisfactory credit rating to open an account with the Houston Cellular Telephone Company.

Even though people may have been contacted through their employer, it was important to emphasize that this was a voluntary effort, totally separate from the company. It was difficult to maintain the interest of the 200 persons for a 1-year period. Monthly newsletters were developed to give the probes information on the progress of the study, the activities of the probes (how many incidents were reported, for example), and activities on the level of their participation. Recognizing that people do not travel to work each weekday for 1 year, the project established a 70 percent goal. To qualify for the free monthly access and for the telephone at the end of the year, the volunteers were required to report 70 percent of the weekdays during the year.

The project staff had to work continuously with the cellular telephone company to keep the billing and the credits for access charges and air time up to date. Also, several of the original 200 probes left the program for various reasons. Many persons were transferred to other locations outside the corridor area. These persons were given the opportunity to either purchase the telephones for their personal use or to return them to the project. Telephones were issued to replacement volunteers to maintain the coverage.

Although the project has been judged to be a success, the amount of management effort required to set up and to maintain the level of participation was more than anticipated. The experience gained in this project will be valuable in establishing the fleet of vehicles that will use transponders in an AVI system to collect the same type of information.

RESULTS

Table 2 presents the summary of information developed from the data collected during a 15 min time period on a particular

TABLE 1 Probe Distribution by Time of Day and Roadway

Distribution by Time of Day			
AM	Number	PM	Number
6:00- 7:00	106	3:00-4:00	18
7:00- 8:00	73	4:00-5:00	58
8:00- 9:00	18	5:00-6:00	89
9:00-10:00	3	6:00-7:00	35

Distribution by Roadway	
Roadway	Number
IH 45 Freeway (North)	95
IH 45 HOVL	14
Hardy Toll Road	20
US 59 Freeway (North)	71
Total	200

TABLE 2 Houston Intelligent Transportation System Travel Summary and Incident Information

Travel Summary for 21 July 1992 at 4:45 p.m.		
Roadway	Distance (miles)	Travel-Time (min)
North Freeway (IH 10 to Holzwarth)	19.5	28.9
North Freeway HOV Lanes	10.8	14.4
Hardy Toll Road (No. Toll to 1960)	16.8	17.0
Eastex Freeway (IH 10 to FM 1960)	16.5	28.4

Incident Information

Time Reported	Freeway	Cross Street	Type	Comments	Status
1.16:35	59/NB	Laura Koppe	Accident	2 cars blocking left lane	HPD on scene

Note: This travel time and incident information is automatically faxed every 15 minutes. If you fail to receive any fax, please contact us at 840-9470.

Source: Texas Transportation Institute
3800 Buffalo Speedway
Phone: 840-9470

day from the cellular phone probes. These summaries are transmitted to the traffic service organizations and the TxDOT's Interim Communication Center for further dissemination.

The travel patterns in the North Houston Corridor indicate that the peak traffic flows occur from 6:00 to 8:00 a.m. and 4:00 to 6:00 p.m. Fortunately these were the times that the study staff enrolled the majority of the probes. For a typical month, 28,000 calls were made to determine travel times and 230 calls were made to report incidents.

Although commercial transportation companies are interested in the use of real-time traffic information, they are not yet interested in investing financial resources for this purpose. The volumes and trip times of the commercial transportation companies reveal that they are more interested in information concerning incidents and the resultant congestion (nonrecurrent congestion) instead of the recurrent congestion of the peak periods. The benefits of the real-time travel time information to commercial transportation companies is under evaluation.

There is also some misperception about the potential contribution by the private sector to the cooperative projects by the public sector. Even though the Houston Cellular Telephone Company made a significant contribution by providing free air time, the company received 200 new customers recruited by the project. The free air time is significant when measured in the manner that it is charged by cellular companies. Each call is charged a minimum of 1 min, regardless of the length of the call. Most of the calls to report the vehicle identification number and the location station require no more than 15 sec. Therefore the $180,000 worth of free air time is by actual time worth approximately $50,000.

The traffic service organizations are interested in becoming partners in the collection and processing of traffic information because that is the very essence of their business. This relationship between the private and public agencies in traveler information should continue to expand.

The use of TTI of the Texas A&M University System as the management agency simplified the project administration and enabled the demonstration project to move forward at an accelerated pace. However these types of projects will become the responsibility of the public operating agencies in the future, and procedural changes will have to be made to accomplish the implementation.

CONCLUSIONS

The use of the cellular telephone probes was an acceptable and reliable source of real-time traffic information. Errors in reporting were small. Test runs by the study staff confirmed the accuracy of the reports. There will be a place in future systems for the use of traffic reports from on-the-scene reporters, even though there may be automatic systems for tracking and locating vehicles. One of the major benefits of the cellular system is the reporting and description of incidents that require the dispatch of emergency vehicles.

The improvement of the real-time information from the cellular telephone probes did not result in an increase in the use of the CMSs in the corridor. These signs, operated by TxDOT from the Interim Communication Center, are used to alert motorists of major incidents and to advise the use of alternate routes. During the 9 months of the study and the 4 to 5 hr per day of the probe study, the use of the CMSs as compared with the previous year did not change significantly. As the coverage of traffic monitoring systems approaches 24 hr, the usage of the signs should increase. In the future the signs will be to advise motorists of the best routes under normal conditions.

The participation by the private sector is dependent on some specific benefits that may be realized in the short term. The office complex provided space at cost. The cellular telephone company provided free air time but collected 200 new customers with monthly access fees. It was difficult to maintain the participation of the private individuals for the full year, even though they received free telephones and paid no access fees.

FUTURE PLANS FOR THE IVHS PARTNERSHIP

Several agencies are installing AVI systems in the Houston area. HCTRA installed an AVI system for the Hardy and

Sam Houston toll roads in September 1992 for automatic toll collection. TxDOT installed an AVI System on the IH 10 Katy Freeway, US 290 Northwest Freeway, and the IH 45 North Freeway HOV lanes and main lanes in early 1993 to measure real-time travel times between the HOV lanes and main lanes. These installations, which will replace the cellular telephone probes on IH 45 North Freeway and the Hardy Toll Road, will be used to enhance the use of public transportation facilities (HOV lanes).

The City of Houston Aviation Department has proposed an AVI system for monitoring traffic flow and for automatic fare collection of commercial vehicles at the two airport facilities. The progress on these three AVI projects has been monitored and coordinated through the meetings of the public-private partnership.

All three of the agencies will probably use the same AVI technology to ensure compatibility and to increase the coverage of vehicles equipped with transponders.

An interim traffic monitoring system will be installed on the US 59 Eastex Freeway during the reconstruction project. The speed detection system will replace the cellular telephone probes in fall 1993.

A part of the IVHS "SMART Commuter Project" will use the real-time travel time information collected with the AVI and cellular telephone probes. Certain commuters will be provided with the information in various forms and locations to encourage an increase in the use of the buses and carpools on the HOV lanes. The SMART Commuter Project will begin in fall 1993.

The Houston area has been designated as a priority corridor under the Intermodal Surface Transportation Efficiency Act of 1991. One of the proposed projects is to expand the AVI network by supplementing the AVI installations along the two toll roads with additional AVI traffic monitoring stations and to monitor the park-and-ride facilities supporting the HOV lanes. Additional AVI systems are proposed for the IH 610 Loop Freeway from the US 59 Eastex Freeway to the US 59 Southwest Freeway and on the Southwest Freeway from the Sam Houston Toll Road to downtown (Figure 3). If funding from the Congestion Mitigation and Air Quality Program is approved, installations should be completed in 1993.

Initially all AVI systems will be accessed by telephone lines. After the fiber optics trunk lines are installed for the TMSs, the telephone communications will be converted to the state-owned fiber network.

Initially all of the real-time travel time information will be received and distributed by the TxDOT Interim Communication Center. This center will be replaced by an interagency central control facility (CCF) to be constructed in 1994. Plans are being made to include the private traffic service organizations in the CCF to assist in the management and distribution of traffic and transit information, but this arrangement is contingent on interpretation of state law concerning the joint use of public facilities by private companies.

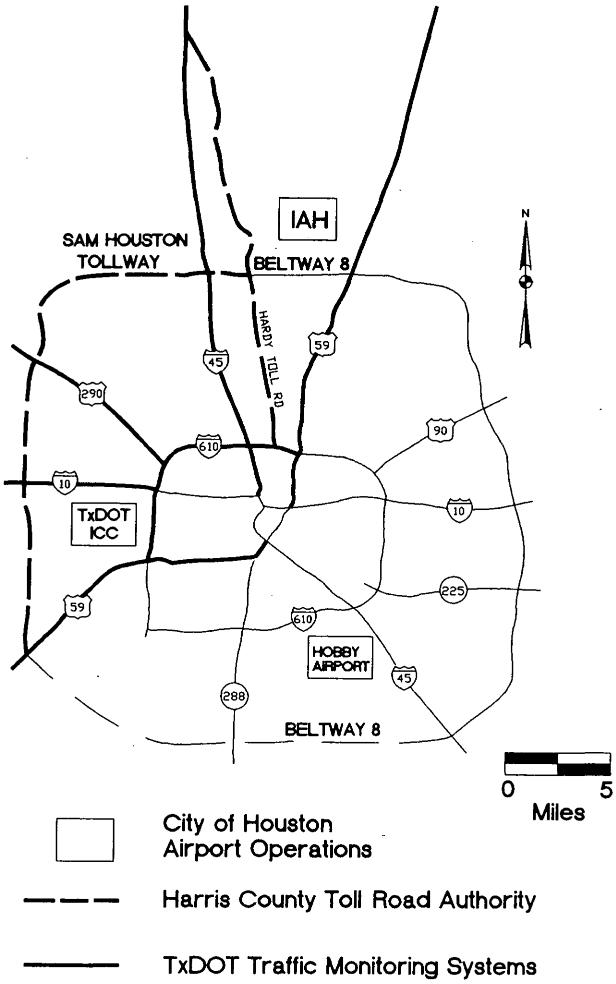

FIGURE 3 Proposed AVI systems.

Future applications of real-time monitoring systems will include commercial fleets, thus expanding the coverage into time periods that will benefit their operations. With the use of automatic systems the interaction with the drivers to report locations is eliminated. In addition, multiple use of the AVI system for toll collection and other restricted area monitoring will enhance the use by the private sector.

Publication of this paper sponsored by Committee on Freeway Operations.

Evaluation of the Seattle I-5 North High-Occupancy Vehicle Lane 2+ Occupancy Requirement Demonstration

KATHERINE F. TURNBULL, CYRUS G. ULBERG, AND LESLIE N. JACOBSON

High-occupancy vehicle (HOV) facilities allow for a great deal of operational flexibility. This flexibility has often been noted as one of the advantages of HOV lanes. The ability to change the vehicle occupancy level required for use of the HOV facility reflects this flexibility. Increasing or decreasing the occupancy requirement provides one approach to managing the demand on the facility and thus the number of vehicles and people using the facility. The impact of increasing and decreasing vehicle occupancy requirements has been a topic of discussion at recent national HOV conferences and other meetings. Although a few areas, most noticeably Houston, have documented the impact of changes in HOV lane vehicle occupancy levels, more information is needed to fully understand the consequences of these changes. This paper summarizes the impact of one such change: the demonstration project in which the minimum vehicle occupancy requirement on the I-5 North HOV lanes in Seattle was lowered from three or more persons per vehicle to two or more persons per vehicle. The impact of this change on the I-5 North HOV lanes and the general-purpose lanes is documented in this paper. The information presented assists in enriching the understanding of the consequences of changing HOV lane vehicle occupancy requirements.

On July 29, 1991, the Washington State Department of Transportation (WSDOT) initiated a 6-month demonstration project in which the minimum vehicle occupancy requirement on the I-5 North high-occupancy vehicle (HOV) lanes was lowered from three or more persons per vehicle (3+) to two or more persons per vehicle (2+). The demonstration was undertaken to determine the impacts of reducing the occupancy requirement from 3+ to 2+ on the operation of the HOV lanes and the freeway general-purpose lanes.

The results of this demonstration, as presented in this paper, should be of interest to transportation planners and engineers, transit personnel, and others. The operational flexibility associated with HOV lanes, including the ability to change the vehicle occupancy level required to use the facility, represents an often-noted advantage of HOV facilities (*1,2*). Advantages of increasing and decreasing HOV occupancy requirements have been discussed at recent HOV conferences and other meetings (*3*). Few quantifiable data exist outside of Houston (*4*) on the impact of changes in vehicle occupancy requirements on the HOV lane and the general-purpose lanes. An evaluation program, conducted by the University of Washington and the Texas Transportation Institute (TTI) under contract to WSDOT, was undertaken to examine these impacts (*5*). The results from this evaluation are summarized here. By providing additional insights into these impacts on the basis of the I-5 North demonstration, this paper enhances the current level of understanding and enriches the available information on the impacts of changes in vehicle occupancy requirements on HOV lanes.

K. F. Turnbull, Texas Transportation Institute, Texas A&M University, College Station, Tex. 77843. C. G. Ulberg and L. N. Jacobson, Washington State Transportation Center, 4507 University Way NE, Suite 204, Seattle, Wash. 98105.

PURPOSE OF DEMONSTRATION AND EVALUATION

The demonstration was initiated to determine the impacts reducing the occupancy requirement from 3+ to 2+ persons per vehicle would have on the operation of the I-5 North HOV lanes and the freeway general-purpose lanes. The location of the I-5 North concurrent flow HOV lanes is shown in Figure 1. The southbound HOV lane is 7.7 mi long, and the northbound HOV lane is 6.2 mi long. The HOV lanes were first opened in 1983.

Specifically, the demonstration was intended to determine whether the change in occupancy requirement maintained or contributed to meeting the objectives of the Washington State HOV program (*6*). The I-5 North Demonstration Project Steering Committee, composed of representatives from WSDOT, Seattle Metro, Community Transit, Pierce Transit, the City of Seattle, Washington State Patrol, and FHWA, provided guidance for the evaluation.

The evaluation was focused on the impact the change in the vehicle occupancy requirement on the I-5 North HOV lanes had on achieving the objectives established by WSDOT for the freeway HOV lane system. The objectives and measures of effectiveness developed by WSDOT, as identified in the following paragraphs, were used to guide the evaluation (*6*):

> Preamble:
> By satisfying the following overall objectives the HOV system is successfully providing mobility choices consistent with the mission of the Washington State Department of Transportation and the goals of recent growth management, commute trip reduction, and air quality legislation. Critical to the success of the HOV system is public support. These objectives and all decisions regarding the system must reinforce public acceptability of and support for HOV facility development.
> The overall objectives of the HOV system are to accomplish the following:
> • Improve the capability of congested freeway corridors to move more people by increasing the number of persons per vehicle.

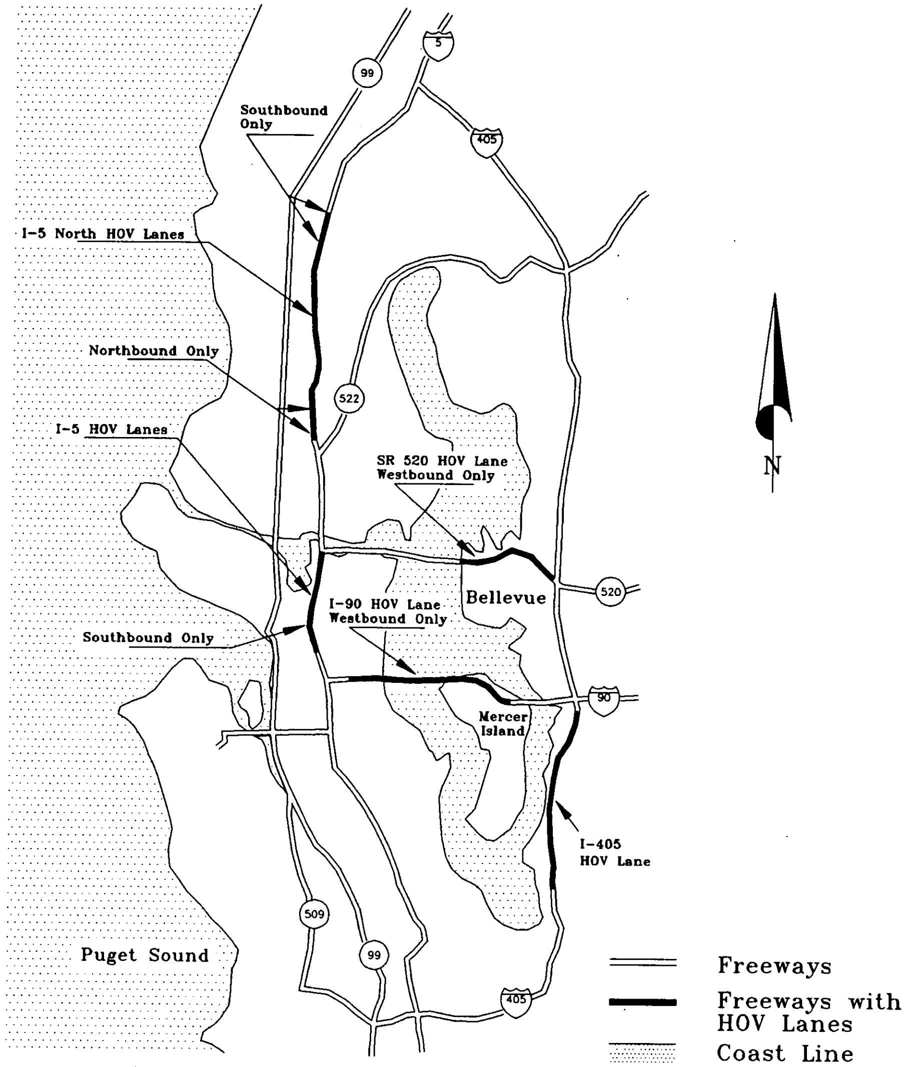

FIGURE 1 Location of I-5 North HOV lanes.

- Provide travel time savings and a more reliable trip time to high-occupancy vehicles utilizing the facilities.
- Provide safe travel options for high-occupancy vehicles without unduly affecting the safety of the freeway general-purpose mainlanes.

Measures of effectiveness used to determine the impact of the HOV system include the following:

- Person throughput,
- Vehicle occupancy,
- Comparative and absolute general-purpose and HOV lane travel times,
- Travel time reliability, and
- Accident rates.

First, the impact of the occupancy requirement change on the HOV lanes was examined. This included an analysis of changes in use levels, travel times, travel-time reliability, and bus ridership levels. The impacts on the general-purpose lanes and the total freeway were also examined. Changes in vehicle volumes in the general-purpose lanes were examined, along with travel time and accident data. The changes in average vehicle occupancy levels and person throughput were also analyzed. Public perception of the change and the HOV lanes in general was also measured through surveys of bus riders, carpoolers, and motorists.

A variety of data collection activities was conducted to support the analysis. Travel-time surveys, using the license plate methodology, were taken of vehicles using both the HOV and general-purpose lanes. Use levels and lane vehicle volumes were obtained from loop detector data gathered as part of the ongoing WSDOT monitoring program. Accident data were obtained from the State Patrol. Bus ridership levels and park-and-ride lot use rates were provided by Community Transit. Special surveys were conducted of Community Transit riders, carpoolers, and motorists to provide additional information on the impact of the change in the vehicle occupancy requirement and general attitudes toward the HOV lanes.

It is important to note that the change in the I-5 North HOV lane occupancy requirement did not occur in a vacuum. Thus, there are a number of limitations with evaluation. First, many factors other than the change in the HOV occupancy requirement may have influenced travel in the corridor during the demonstration. Identifying and evaluating the impact of these confounding variables, such as the introduction of the U-Pass program at the University of Washington, was difficult. Second, individuals' travel habits change over time, and changes in modes, travel times, and routes could be expected in any corridor. However, while individuals' travel habits change, it takes time to establish new routines and patterns. Thus, changes in travel characteristics may not emerge immediately after the implementation of a demonstration. The analysis was further limited by the availability of data, especially for the period immediately preceding the start of the demonstration.

Even with these limitations, the evaluation provides an indication of the general trends and impacts of the reduction in the I-5 North HOV lane occupancy requirement on the HOV and general-purpose lanes. Thus, results of this analysis should be of interest to other areas considering changing the vehicle occupancy requirements on HOV lanes. Further, the results enrich the general understanding of the impact changes in occupancy levels have on the operation of the HOV and general-purpose lanes, overall vehicle occupancy levels, and other factors.

ORGANIZATION OF PAPER

The remainder of this paper presents the results from the 6-month evaluation of the demonstration project in which the minimum vehicle occupancy requirement on the I-5 North HOV lanes was lowered from 3+ to 2+ persons per vehicle. To accomplish this, the remainder of the paper is divided into two major sections. The next section provides an overview of the data collection activities undertaken as part of the evaluation, discusses some of the limitations associated with the evaluation, and presents a summary of the major results. The paper concludes with a discussion of how these results may be of use on a national level and ideas for improving future evaluations.

EVALUATION RESULTS

This section presents a summary of the results from the evaluation of the reduction in the vehicle occupancy requirement on the I-5 North HOV lanes. The information used in the analysis was obtained through special surveys and from ongoing WSDOT monitoring efforts. The availability of data from these different sources varies. The WSDOT lane volumes were available in computerized format for recent years. However, data older than 1990 were available only on microfiche, making it more difficult to obtain and analyze. Due to the quick implementation of the demonstration, travel-time surveys and vehicle occupancy counts were conducted for only 4 days in July before the change was made. Thus, the ability to fully analyze many impacts during the evaluation was limited by the available data.

The major focus of the analysis was on the impact the change in the vehicle occupancy requirement had on traffic levels and traffic conditions during the morning and afternoon peak hours and peak periods. For the purposes of the evaluation, the morning peak hour was defined as 7:00 a.m. to 8:00 a.m.; the morning peak period was defined as the 3-hr period from 6:00 a.m. to 9:00 a.m.; the afternoon peak hour was defined as 5:00 p.m. to 6:00 p.m.; and the afternoon peak period was defined as the 3-hr period from 3:30 p.m. to 6:30 p.m. These times correspond to the periods when the greatest demands, and thus the greatest vehicle volumes, are typically placed on the general-purpose freeway and HOV lanes.

The major results from the evaluation are summarized in Table 1 and discussed in more detail next. Information is presented first on the impact the change had on the I-5 North HOV lanes. This includes a discussion of changes in use levels, travel times, travel-time reliability, and bus ridership levels. This is followed by a discussion of the impact on vehicle volumes, travel times, and travel-time reliability for the general-purpose lanes. The changes in overall vehicle occupancy levels and person movement are discussed next. This is followed by a brief review of available information on accident rates. This section concludes with a summary of the public perception concerning the change.

Impact on I-5 North HOV lanes

HOV Lane Use Levels

Data on the HOV lane use levels were obtained from the WSDOT ongoing monitoring program on the I-5 North freeway. Loop detectors located in the pavement at selected sites along the HOV and general-purpose lanes record the number of vehicles passing over the detectors. Data from 1990 to the present were available through the WSDOT computer system. However, data from before 1990 were stored on microfiche, making it harder to compile and analyze those data. For the purpose of the analysis, HOV lane use trends in 1990 and 1991 at the 3+ vehicle occupancy level were compared with those during the 6-month demonstration period when the 2+ occupancy requirement was in effect. HOV lane volumes were examined for three southbound and two northbound data stations.

In general, the peak-hour and peak-period volumes in the HOV lanes recorded at each of the three station more than doubled after the change to the 2+ vehicle occupancy requirement. However, slightly different trends emerged in the morning and afternoon periods. In the morning peak hour–peak direction, HOV volumes increased significantly with the change to the 2+ requirement and continued to increase during the demonstration period. HOV volumes increased from approximately 500 vehicles an hour to between 1,200 and 1,400 vehicles an hour. This trend may be seen by the information from one data station, shown in Figure 2. Further, the number of vehicles using the HOV lane continued to increase during the demonstration. The vehicle volumes in August and September, immediately after the change, averaged between 950 and 1,100 vehicles during the morning peak hour. However, the peak-hour HOV volumes continued to increase steadily, averaging between 1,300 and 1,400 vehicles in November and December. Similar trends were reflected in the morning peak-period volumes, which more than doubled initially and continued to increase during the demonstration.

The afternoon volumes in the HOV lane also more than doubled for both the peak hour and the peak period. The

TABLE 1 Summary of General Trends Associated with I-5 North 2+ HOV Lane Demonstration

Measures of Effectiveness and Related Policies	General Trends in Change from 3+ Persons per Vehicle to 2+ Persons per Vehicle
Person Throughput	
HOV Lane Vehicle Volumes	Significant increase in peak hour volumes from approximately 400-500 vehicles to 1,200 to 1,400 vehicles or approximately 200%.
HOV Lane Person Volumes	Peak hour person volumes increased by approximately 1,200 or 35%.
General-Purpose Lane Vehicle Volumes	Remained approximately the same.
General-Purpose Lane Person Volumes	Remained approximately the same.
Bus Ridership Levels	Community Transit ridership to the University has increased. This appears to be result of the U-Pass Program. Ridership to downtown Seattle was stable or slightly declining.
Total Person Throughput	Increased during demonstration period by approximately 12%.
Vehicle Occupancy	
HOV Lane	Declined, due to more 2 person carpools and fewer 3 person carpools.
HOV and General-Purpose Lanes	Vehicle occupancy levels for the combined HOV and general-purpose lanes increased after the start of the demonstration, but have declined since to about the same level (1.2 persons per vehicle) as in 1989-1990.
Travel-Times and Travel-Time Reliability	
HOV Lane Travel Times	Morning peak hour travel times have remained about the same, while afternoon peak hour travel times have increased.
General-Purpose Travel Times	Morning peak hour travel times have decreased and afternoon peak hour travel times have increased.
HOV Lane Time Savings over General Purpose Lanes	HOV lane travel time savings decreased in the morning peak hour, but increased in the afternoon peak hour. These changes are due primarily to changes in travel speeds in the general-purpose lanes.
HOV Lane Travel Time Reliability	HOV travel time reliability has declined, especially in the afternoon peak hour.
General-Purpose Lanes Travel Time Reliability	Travel time reliability in the general-purpose lanes has remained about the same in the morning, but appears to have declined slightly in the afternoon peak hour.
Bus Service On-Time Performance	On-time performance appears to have declined slightly in the afternoon.

(continued on next page)

volumes in November and December averaged between 1,300 and 1,550 vehicles. This represented a significant increase from the average of 400 to 600 vehicles recorded in March through June, 1991, before the demonstration. In general, the afternoon peak-hour HOV lane volumes were higher than those recorded during the morning peak hour. However, contrary to the morning trend of increasing growth in vehicle volumes, the afternoon data indicated a slightly different trend. Some of the highest volumes were recorded in August, immediately after the change to the 2+ occupancy requirement. From September to December, the HOV lane volumes were more erratic, and a clear picture of increasing volumes, such as noted in the morning, did not emerge. This trend may have been caused by initially higher afternoon volumes with little capacity for additional vehicles.

Travel Times

Travel times for vehicles using the HOV lane were measured by recording the license plate numbers of vehicles in the lane at two locations in the corridor. The license plate information was recorded using a microcomputer that also recorded the time at locations close to the beginning and end of the HOV

TABLE 1 (*continued*)

Measures of Effectiveness and Related Policies	General Trends in Change from 3+ Persons per Vehicle to 2+ Persons per Vehicle
Accident Rates	
HOV and General-Purpose Lanes Accident Rates	No discernible trends were identified associated directly with the change. However, the areas downstream from both the northbound and southbound HOV lane reflect an increasing accident rate, which started before the implementation of the demonstration.
Public Perception	Surveys of bus riders, carpoolers, and motorists indicate that all three groups think the I-5 HOV lane is a good transportation improvement and between 59% to 74% think it is sufficiently utilized. 39% of the bus riders, 83% of the carpoolers, and 89% of the motorists survey felt permitting 2+ carpools to use the I-5 North HOV lane was a good move. 47% of the bus riders and 23% of the carpools indicated that travel times seem longer with the 2+ requirement, 23% said buses are not on schedule, and 5% reported missing connections. 23% of the carpoolers reported longer travel times and 21% reported problems entering or exiting the lane.
Other Federal, State, and Local Policies	
Transportation Policy Plan for Washington State	This plan emphasizes the movement of people rather than vehicles and advocates the provision of cost-effective alternatives to one-person vehicles, including transit and ridesharing. To the extent that the 2+ demonstration has resulted in more vehicles on the facility and lowered the overall occupancy vehicle rate, the results are counter to this plan.
Clean Air Act and Commute Trip Reduction Act	To the extent that more vehicles are moving through the I-5 corridor, due to the reduction in 3+ carpools, and to the extent that a degradation in the travel times and travel time reliability for HOVs have occurred, the demonstration is less supportive of these acts than the 3+ requirement.
Growth Management Act and PSCOG's Vision 2020	The demonstration evaluation did not examine possible land use or growth management impacts of the change, which would occur over a long time period. The impact on the Growth Management Act and other policies should be examined more closely.

lanes. The license plate numbers were then matched by computer and the travel time computed. Because of the short time period before the demonstration project was implemented, only 4 days of "before" data were available for July 1991.

Overall, the travel times in the segment of the HOV lane included in the survey during the morning peak hour did not change significantly with the 2+ requirement. The average travel time was 5.82 min before the demonstration and 5.80 min after, even with the increase in the number of vehicles in the lane. During the afternoon peak hour, the travel time in the HOV lane increased from 7.50 min before the demonstration to 7.98 after.

The travel-time savings for HOVs during the morning peak hour decreased during the demonstration. This decline was due primarily to a decrease in travel times for vehicles in the general-purpose lanes. Thus, while the travel times for vehicles in the HOV lane remained relatively constant, the travel times in the general-purpose lanes decreased, resulting in less difference between travel times for vehicles in the HOV and general-purpose lanes.

The travel-time savings for HOVs during the afternoon peak hour were greater during the demonstration period than those experienced when the 3+ occupancy requirement was in effect. On Mondays through Thursdays, the travel-time savings increased from 1.58 min to 3.44 min, and on Fridays the increase was from 6.12 min to 9.66 min. Because the travel times in the HOV lane during the afternoon peak hour did not decrease significantly, the increased travel-time savings resulted primarily from longer travel times in the general-purpose lanes, which occurred during the demonstration.

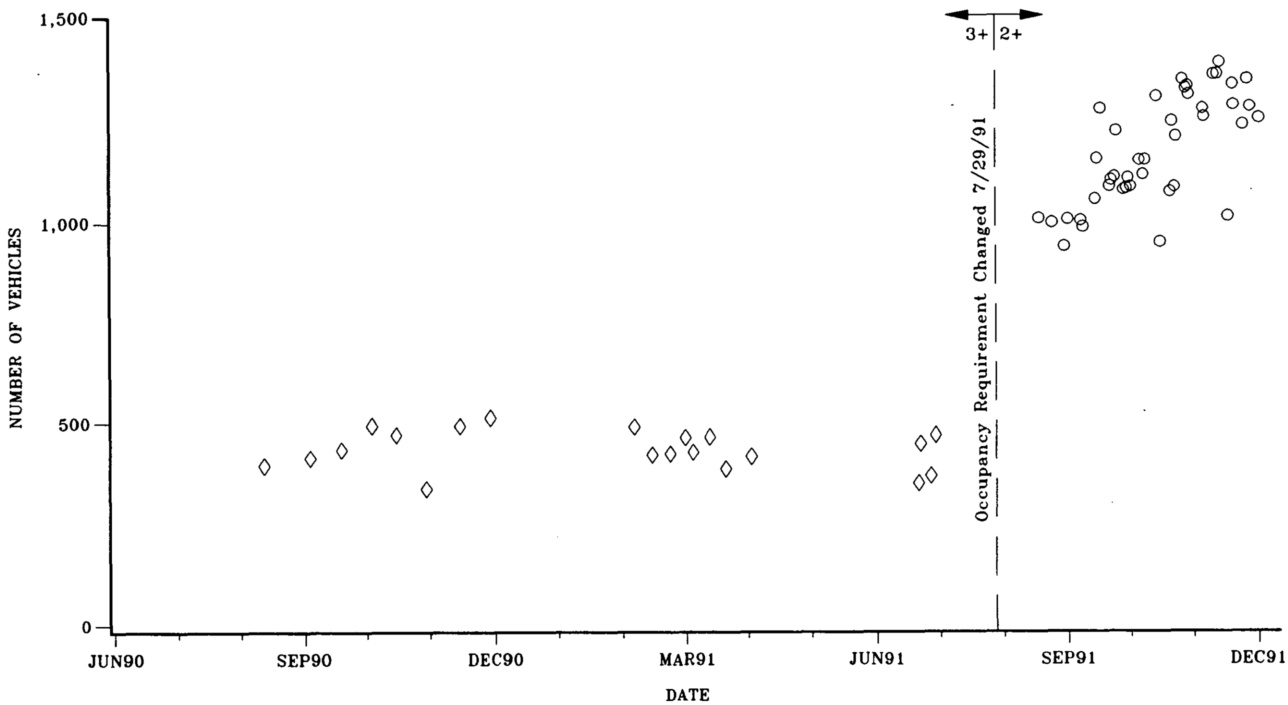

FIGURE 2 Morning peak-hour HOV volumes, 145th Street.

Travel-time Reliability

Another important factor influencing the use of HOV lanes is the travel-time reliability offered by these facilities. Thus, not only is the actual shorter travel time important to HOV lane users, the ongoing reliability of these savings is also important. The influence of the change in the vehicle occupancy requirement on travel-time reliability was measured by calculating the standard deviation of speeds in the HOV lanes before and after the change. This measure provided an indication of the reliability of travel times for HOVs.

The analysis indicated that travel times were more reliable in the morning than in the afternoon. However, the increase in the standard deviation experienced in both time periods indicated that travel-time reliability suffered during the demonstration. Thus, travel times in the HOV lanes were not as reliable with the 2+ occupancy requirement as they were with the 3+ requirement. The analysis indicated that this change was strong and statistically significant.

Bus Ridership Levels

Bus service in the I-5 North corridor is provided by both Seattle Metro and Community Transit. However, because of the location of the HOV lanes and the orientation of service, only Community Transit buses use the I-5 North HOV lanes in the study area on a regular basis. Information on the number of buses using the lane, ridership levels, and use of park-and-ride lot was obtained from Community Transit. In addition to the examination of trends in bus use on I-5 North, comparisons were also made of bus volumes and ridership levels associated with HOV facilities in other metropolitan areas in North America.

Community Transit operates a total of 19 weekday routes in the I-5 North Corridor. Of these, 14 routes are oriented toward downtown Seattle, 4 routes are oriented toward the University of Washington, and 1 route provides service to the North Seattle Community College. A total of approximately 109 inbound and 116 outbound buses provide weekday service to downtown Seattle, whereas 48 inbound and 44 outbound buses serve the university. One inbound and one outbound bus serves the Community College. Of these, approximately 59 buses operate during the morning peak hour, and 40 operate during the afternoon peak hour. Service is oriented from both neighborhood areas and park-and-ride lots located in Snohomish County. In addition, connections are provided to both the Edmonds and Mukilteo ferry service.

This represents a significant level of bus service for an HOV facility. The bus volumes on the I-5 North HOV lane rank first in a comparison with other concurrent flow HOV lanes in the United States. Further, the bus volumes are higher than a number of the exclusive freeway HOV facilities. Thus, buses represent a significant component in the vehicle mix on the I-5 North HOV lane.

Overall, approximately 10,000 daily riders were carried to downtown Seattle and the University of Washington on Community Transit buses using the I-5 North HOV lane. During the morning peak hour, some 2,500 riders were carried on buses using the HOV lane. The morning peak-hour bus passenger volumes on the I-5 North HOV lane compare favorably with levels on other HOV lanes in North America. The I-5 facility carries the second largest number of bus riders in the morning peak hour of the concurrent flow HOV lanes and records higher bus passenger volumes than some of the exclusive freeway HOV lanes.

Ridership on the Community Transit systems has grown dramatically during the past 5 years. In 1986, the daily average ridership on the commuter routes to downtown Seattle was approximately 3,400 passengers. By 1990, this figure had increased to some 7,400 passengers. During 1991, ridership leveled off, however. The ridership trends for the months of August 1991 through January 1992 were examined for routes oriented toward downtown Seattle. While ridership remained constant with the levels from the previous year in August and December, ridership in September, October, November, and January was slightly below the levels for the same months the previous year.

It appears that this stabilization and slight decline may be attributable to a number of factors. First, most runs are at capacity with little room for new riders. Second, the general slowdown in the economy may have resulted in fewer bus riders in the corridor. Third, some bus riders may have changed to carpooling with the lower occupancy requirement. Given all these factors, it is difficult to determine the exact impact the occupancy requirement change had on bus ridership levels on service to downtown Seattle.

Community Transit service to the University of Washington experienced significant ridership increases during fall 1992. The average daily ridership during the school year increased from 1,593 passengers in February to 2,739 passengers in October. This trend, which was also experienced on Seattle Metro service to the university, appears to be the result of the new U-Pass Program. To accommodate the increased ridership, articulated buses replaced regular buses on some routes.

Impact on I-5 North General-Purpose

Vehicle Volumes

Vehicle volumes for the I-5 North general-purpose lanes were obtained from the WSDOT ongoing monitoring program on the I-5 North freeway described previously. The examination of the vehicle volumes in the general-purpose lanes before and during the demonstration indicated a good deal of variability between days. Vehicle volumes for the three-lane section of the facility averaged between 4,000 and 5,000 vehicles for the morning peak hour, with slightly higher volumes during the afternoon peak hour. The vehicle volumes were generally similar during the complete time period, however, and no significant changes were discernible during the demonstration.

Travel Times

Travel times were estimated for vehicles in the general-purpose lanes using the same methodology described previously. The results of this analysis indicated that on average, travel times declined during the morning peak hour and increased during the afternoon peak hour during the demonstration. During the morning peak hour, travel speeds for vehicles in the general-purpose lanes decreased by almost 2 min. During the afternoon, however, travel times increased by 2 to 3 min. Field observations indicated that the increase in the afternoon travel time appears to be caused in part by traffic congestion resulting from the lane drop and merging occurring at the north end of the HOV lane.

Changes in Vehicle Occupancy Levels

The changes in personal vehicle occupancy levels on the I-5 North HOV facility were also examined. This analysis did not include bus ridership, which was examined separately. Personal vehicle occupancy levels on the I-5 facility were measured in 1989 and 1990 as part of the WSDOT Vehicle Occupancy Monitoring Project. Vehicle occupancy information was also collected for 4 days in July 1991, before the start of the demonstration. Similar surveys were conducted during the first 5 months of the demonstration.

Overall, the percentage of two-person carpools was higher in the afternoon peak period. This relates to the fact that there are more nonwork trips, which tend to have higher occupancy levels, in the afternoon. The results of the analysis further indicated that there was as increase in two person vehicles from 1989 to July 1991. After the initiation of the 2+ demonstration, two-person carpools increased from some 10.5 percent to 16.5 percent during the morning peak period. A similar, although smaller, increase also occurred during the afternoon peak period. The number of two-person carpools declined during the October to December time period, however, returning to approximately the same percentage as the period before the demonstration.

A different trend was found in three-person carpools. The percentage of three-person carpools on the I-5 North facility has historically averaged about 4 percent. During August and September, the first 2 months of the demonstration, the percentage of three-person carpools remained about the same. The percentage of three-person carpools dropped off considerably from October through December, however. In December, three-person carpools accounted for approximately 1 percent of the morning peak-period volumes.

Trends in the percentage of single-occupant vehicles (SOVs) on the I-5 North freeway were also examined. SOVs represented the largest percentage of vehicles using the facility both before and during the demonstration. The percentage of SOVs in the morning peak hour decreased slightly after the start of the demonstration, reflecting the increase in two-person carpools noted earlier. However, reflecting the mirror image of the trends described earlier, the percentage of SOVs increased again after October to a level slightly below the previous high in 1989 and 1990. The average vehicle occupancy level during the morning peak period in December was approximately 1.2 persons per vehicle, the same level as in 1989 and 1990.

Person Movement

The change in the vehicle occupancy requirement influenced not only the number of vehicles using the HOV lane, but also the number of people using the facility. On the basis of the Community Transit ridership information, it appears that ridership to downtown Seattle was slightly lower than the previous year, whereas ridership to the University of Washington increased significantly. In order to determine the impact the change in the occupancy requirement had on the person-movement levels in the I-5 North HOV lanes and the general-purpose lanes, a number of factors were examined. These factors included the increase in vehicles using the lane, changes in the observed vehicle occupancy levels, and the results from the surveys of bus riders, carpoolers, and motorists.

The results of this analysis indicated that the person throughput for the total facility was greater during the demonstration period. A slight decline in the person throughput was noted during the demonstration period, however. Thus, in general, more people were being moved on the facility with the 2+ occupancy requirement than on the one with the 3+ occupancy requirement.

Accident and Safety Impacts

One of the objectives of the WSDOT freeway HOV system policy states that the HOV lanes should provide safe travel options for HOVs without unduly affecting the safety of the freeway general-purpose lanes. The intent of the safety analysis was to determine whether the reduction in the vehicle occupancy requirement altered the safety levels on the segment of I-5 North under study. Information from the WSDOT accident data base and accident reports from the Washington State Patrol were used for the analysis, which focused on identifying changes in safety conditions on both the HOV and general-purpose lanes in the study area and examining possible causes if any changes were found. The analysis of the accident information did not identify any specific trends or variations in the previous patterns that could be associated with the reduction in the vehicle occupancy requirement on the I-5 North HOV lanes.

Public Perception

The preamble to the freeway HOV system objectives outlined by WSDOT notes that public support is critical to the success of the HOV system. Previous surveys (*7*) conducted by WSDOT and Seattle Metro have indicated that the HOV lanes have been received positively by users, nonusers, and the general public in the Seattle area. Surveys of bus riders, carpoolers, and motorists in the I-5 North corridor were conducted to obtain additional information on the perceptions of these user groups concerning the change in the occupancy requirement.

A survey of Community Transit bus riders was conducted on Thursday, November 21, 1991. Approximately 1,300 surveys were distributed; 925 were returned, yielding a response rate of about 71 percent. License plates were recorded in November of carpools in the HOV lane and motorists in the general-purpose lanes. Surveys were then mailed to those two groups. A total of 534 surveys was mailed to motorists; 160 were returned, yielding a response rate of 30 percent. Six hundred surveys were mailed to carpoolers; 57 completed forms were returned, a response rate of 10 percent. The low response rate for the carpool and motorist surveys may be due to the holidays. As a result of the low response rate, it was not possible to draw statistically significant conclusions from the carpool and motorist surveys. However, the results were used to identify general trends.

The results of the surveys of Community Transit bus riders, carpoolers, and motorists indicated a general support for the HOV lanes, but showed a mixed response to the reduction in the occupancy requirement on the I-5 North HOV lanes. More than 90 percent of the bus riders and carpoolers and 82 percent of the motorists surveyed believed that the I-5 HOV lanes were good transportation improvements. A lower percentage, ranging from 42 percent to 74 percent, indicated that they felt the lanes were currently being sufficiently used. Slightly more than 50 percent of the motorists surveyed believed that traffic conditions in the general-purpose lanes had improved since the change in the occupancy requirement was made, and only 15 percent reported encountering problems in the general traffic lanes that may have resulted from the lower occupancy requirement. Almost 50 percent of the carpool respondents indicated that they had not encountered any difficulties with the HOV lanes since the change.

Some 23 percent of the carpoolers responding to the questions indicated that travel times appeared longer, however, and 21 percent reported problems entering or exiting the lane. An even larger number of bus riders reported problems after the occupancy change was made. Some 47 percent indicated that travel times appeared longer, 23 percent reported buses were not always on schedule, and 5 percent noted they had missed connections. More problems were noted by bus riders for afternoon trips than morning trips. Finally, the response to the question on whether permitting 2+ carpools was a good move varied by user group. Approximately 92 percent of the motorists and 83 percent of the carpoolers surveyed indicated strong support for the change, whereas only 39 percent of the bus respondents favored the change.

Many of the respondents to all three surveys provided additional comments. A number of bus riders indicated problems with slower travel times and many strongly supported a return to the 3+ occupancy requirement. These concerns were also reflected in a petition received by Community Transit in September and telephone calls from bus riders complaining about the negative impacts on bus service since the change. A number of comments on all three surveys strongly supported extending the HOV lanes further in the northbound direction.

CONCLUSIONS

A number of observations may be made concerning the 6-month evaluation of the 2+ demonstration project on the I-5 North HOV lanes. These observations are related to both the demonstration and the evaluation process and the results of the 6-month evaluation. These observations, which are discussed briefly in this section, should be of interest on both a local and national level.

First, the evaluation of the I-5 North 2+ demonstration was limited by data availability because the demonstration was implemented on short notice. Consequently, few "before" data were collected, which limited the ability for meaningful comparisons to be made with the data collected during the demonstration. The importance of adequate "before" and "after" data has been stressed in other studies (*4,8*). The I-5 North evaluation further supports this need.

Additional data-related problems were encountered during the study. Although historical data on vehicle volumes, accidents, and other information needed in the evaluation have been maintained, obtaining it was often difficult. This further supports the importance of maintaining ongoing transportation data bases with easy access. The WSDOT monitoring system is one of the better systems in the United States, and improvements are being made to facilitate access to the data. This will benefit future evaluations and studies.

The results of the 2+ demonstration indicated ambiguous evidence concerning the relationship between the 2+ carpool definition and the WSDOT HOV System Policy objectives and other related policies. This was due to the limitations of the evaluation just noted, the short time period covered by the evaluation, and the ambiguity of some of the HOV System Policy objects and the lack of specific measures of effectiveness. The difficulty in drawing conclusions on some of the objectives is discussed here briefly.

The first objective of the HOV System Policy is to "improve the capability of congested freeway corridors to move more people by increasing the number of persons per vehicle." The demonstration results indicated that more people were traveling on the I-5 facility during the peak period during the demonstration. However, the number of persons per vehicle was lower during the October to December period than it was in the week before the demonstration started. The percentage of SOVs increased, and the percentage of 3+ carpools decreased. The percentage of two-person carpools increased initially, but went back down to predemonstration levels. Data also indicated that on a monthly basis Community Transit ridership levels to the downtown during the demonstration were approximately the same or slightly lower than those of the previous year.

A number of questions arise concerning these data, however. The decrease in 3+ person carpools may be attributable to the change in the occupancy requirement. However, the increase in SOVs may come from several sources, including shifts from parallel arterials, shifts from earlier and later time periods, and latent demand filling the additional capacity. Ridership on Community Transit increased to the university at a level comparable to the increase experienced by Metro Transit that was attributable to the U-Pass. The lower ridership levels from the previous year to downtown Seattle may have been partially attributable to the lack of increase in service, the general downturn in the economy, and poor weather conditions in the previous year, which resulted in some of the highest ridership levels recorded by Community Transit. However, survey data and other responses from bus riders show that they did notice a decline in the service reliability for buses using the I-5 HOV lanes.

The second objective in the HOV System Policy is to "provide travel time savings and a more reliable trip time to high-occupancy vehicles utilizing the facilities." During the demonstration, HOVs still saved time, on the average, by using the HOV lanes. However, the travel-time savings in the morning decreased, and the increase in travel-time savings in the afternoon from a decrease in speed in the general-purpose lanes. Furthermore, reliability of travel time was degraded.

During the demonstration, conditions in the morning and afternoon peak periods were quite different. During the morning peak hour, travel times for HOVs did not change significantly. Travel times in the general-purpose lanes actually decreased during a season when increased congestion is normal. This improvement in general-purpose travel times may have been due to a combination of an increase in transit usage to the university, resulting from the U-Pass, and the ability of two-person carpools to use the HOV lanes.

For the afternoon, the results were different, with travel times in both the HOV lanes and the general-purpose lanes increasing. The major difference in the two time periods is partially a result of an existing bottleneck at the north end of the corridor, in the northbound direction. At this point, the HOV lane ends, a general-purpose lane drops, and the resulting merging often causes traffic to back up. The impact of this bottleneck was exacerbated during the demonstration by the large volumes of vehicles in the HOV lane. The increase in volume may be due partially to increased demand historically experienced in the fall. Unfortunately, comparable travel-time information was not available from previous years to compare the change to the normal seasonal change in congestion.

Travel-time reliability for HOVs declined during the demonstration. The standard deviation for travel speed doubled in the morning and increased by approximately 50 percent in the afternoon. On-time performance for Community Transit buses declined compared with the first part of the year. Bus riders responding to the surveys noted that travel-time reliability, especially in the afternoon, was worse during the demonstration, and additional complaints have been received by Community Transit.

It appears that travel-time reliability declined even when the average travel time did not change significantly. The reason for this was that with a 3+ carpool definition and vehicle volumes no higher than 500 to 600 in the peak hour, vehicles in the HOV lane have virtually always enjoyed free-flow conditions. With the additional vehicles in the HOV lane resulting from the change to the 2+ definition, traffic volumes reached levels at which the flow occasionally broke down.

The information presented in this paper should be of benefit on both a local and a national level. At the local level, the results of the 6-month demonstration provide useful data for making future decisions on to the vehicle occupancy levels for the I-5 North HOV lanes and for other HOV facilities in the Seattle area. On a national level, the information provides an enriched understanding of the consequences of changing HOV lane vehicle occupancy requirements. As such, this paper further enhances the common body of knowledge on the use and impacts of HOV facilities.

ACKNOWLEDGMENTS

The evaluation of the Seattle I-5 North HOV Lane 2+ Occupancy Requirement Demonstration Project was a joint effort of the Washington State Transportation Center at the University of Washington and TTI, a part of The Texas A&M University System. A number of individuals at both institutions assisted on the different elements of the evaluation, including Gary Farnsworth and Graciela Etchart of the Washington State Transportation Center and Russell Henk and David Schrank of TTI. In addition, Nicole Beasley of TTI provided word processing support for the preparation of this manuscript. The assistance of these individuals is both acknowledged and greatly appreciated.

REFERENCES

1. K. F. Turnbull and James W. Hanks, Jr. *A Description of High-Occupancy Vehicle Facilities in North America.* U.S. Department of Transportation, 1990.
2. D. L. Christiansen. *High-Occupancy Vehicle System Development in the United States: A White Paper.* TTI, College Station, Tex., 1990.
3. *Transportation Research Circular 384: Fifth National High-Occupancy (HOV) Facilities Conference: HOV Facilities—Coming of Age.* TRB, National Research Council, Washington, D.C., 1991, 303 pp.
4. D. L. Christiansen and D. E. Morris. *An Evaluation of the Houston High-Occupancy Vehicle Lane System.* TTI, College Station, Tex., 1991.
5. Washington State Transportation Center and TTI. *I-5 North High-Occupancy Vehicle Lane 2+ Occupancy Requirement Demonstration Evaluation.* WSDOT, Seattle, 1992.
6. *Washington State Freeway HOV System Policy: Executive Summary.* WSDOT, Olympia, 1991.
7. R. Kinchen et al. *HOV Compliance Monitoring and Evaluation of the HERO Hotline Program.* Report WA-RD205.1. WSDOT, Seattle, 1990.
8. K. F. Turnbull, R. H. Henk, and D. L. Christiansen. *Suggested Procedures for Evaluating the Effectiveness of Freeway HOV Facilities.* U.S. Department of Transportation, 1991.

Publication of this paper sponsored by Committee on High-Occupancy Vehicle Systems.

Equilibrium Traffic Assignment with High-Occupancy Vehicle Lanes

RONALD EASH

Traffic forecasts were prepared by the Chicago Area Transportation Study (CATS) for a proposed high-occupancy vehicle (HOV) lane in northeastern Illinois as part of a Stevenson Expressway HOV lane feasibility study. The traffic assignment procedure developed for this project is outlined. Problems encountered in adapting traffic assignment to highway networks with HOV lanes are first presented. Then, network equilibrium concepts are used to illustrate driver behavior when an HOV lane is available. It is concluded that an equilibrium traffic assignment procedure may be applied to HOV lane traffic assignment. Remaining sections of the paper deal with particular issues in applying the HOV lane traffic assignment in the Stevenson Expressway project. Coding of different HOV lane alternatives in CATS traffic assignment network is discussed and network coding for a concurrent HOV lane illustrated. Trip tables are adjusted for altered mode and vehicle occupancy choices caused by the time savings between the HOV lane and general purpose expressway lanes. The HOV lane traffic assignments were accomplished with software from the Urban Transportation Planning System. The resulting HOV lane traffic assignment procedure is only slightly more complex than normal equilibrium traffic assignment. Results from one of the project's HOV lane traffic assignments are reported in the final section.

A review of potential high-occupancy vehicle (HOV) lane locations in northeastern Illinois was completed in 1990 as part of the region's Operation GreenLight congestion relief program (*1*). Three expressway corridors were identified as the most promising locations for an HOV lane. Two radial corridors, the Kennedy (I-90/94) and Stevenson (I-55) expressways, and a suburban circumferential corridor (I-290 and I-355 tollway) were selected (*2*). The map of the region's expressway network in Figure 1 shows the locations of these corridors.

The three corridors were screened using general criteria to evaluate the feasibility of an HOV lane. Evaluation criteria included the following: (a) existing traffic congestion and average speed; (b) potential HOV lane travel time savings; (c) HOV traffic based on current two-person carpools; (d) feasibility of expanding or reducing existing roadway capacity; and (e) impact on parallel transit services.

After this preliminary evaluation, the Stevenson Expressway (I-55) corridor was selected for additional study. This is the southwest radial expressway corridor shown in Figure 1. A consultant was then contracted by the Illinois Department of Transportation in mid-1992 to perform a feasibility and preliminary design study of an HOV lane in the Stevenson Expressway corridor. A draft report was completed in early 1993 (*3*).

Chicago Area Transportation Study, 300 West Adams Street, Chicago, Ill. 60606.

The Chicago Area Transportation Study (CATS) provided the consultant with forecasts of HOV traffic. The traffic assignment procedure developed at CATS for these HOV traffic estimates is described here. With the exception of one program written at CATS, the HOV lane traffic assignment procedure is applied with existing models and Urban Transportation Planning System (UTPS) model software (*4*). Although not formally proven in this paper, the HOV lane traffic assignment appears consistent with equilibrium traffic assignment.

HOV LANE TRAFFIC ASSIGNMENT ISSUES

A number of complex issues are associated with traffic assignment on HOV lanes. An HOV lane traffic assignment requires at least two vehicle trip tables, one for single-occupant vehicles (SOVs) and trucks and one for carpool vehicles and other HOVs. Additional HOV trip tables for different sized carpools may be needed to determine the most desirable vehicle occupancy requirements for the HOV lane.

The coded highway network on which these trip tables are assigned includes conventional roadways that may be traveled by all vehicles and HOV links that may be used only by carpools. Characteristics of different HOV lane designs must be specified and accurately coded in the network.

An HOV lane traffic assignment requires separate minimum time paths for the SOV and HOV trip tables. Link travel times for non-HOV facilities depend on all vehicles assigned from both SOV and HOV trip tables, whereas HOV link travel times only reflect HOVs. Although it is possible to carry out separate assignments and then combine them, the order of the assignments affects the link volumes. Should the SOV trip table be loaded first and the resulting link times be used to assign the HOV trip table, or vice versa?

Assignment procedures that approximate the behavior of drivers when HOV facilities are present conflict with state-of-the-art multiple path equilibrium traffic assignment. This traffic assignment algorithm converges to the theoretical equilibrium conditions between link travel times and drivers' route choices, and it is now available in most model software packages. These equilibrium conditions are the following. First, all paths traveled between the same origin-destination zone pair have similar travel times. Second, no driver may reduce travel time by transferring to another route (*5*).

The concept of network equilibrium is especially relevant to HOV lane traffic assignment because HOV lanes usually parallel general purpose expressway lanes. Small differences in travel times between the HOV lane and general purpose lanes can determine the link that appears in most zone-to-

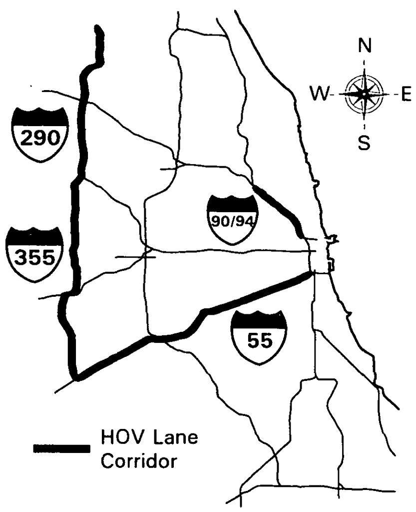

FIGURE 1 Potential northeastern Illinois HOV lane corridors.

zone minimum time paths. Traffic assigned to the HOV lane and general purpose lanes may vary substantially from one traffic assignment to another if the sequence followed to assign trips and calculate link times favors either the HOV or general purpose lanes.

Accurate HOV lane traffic assignments are particularly important for the design of an HOV lane because of the interdependence between an HOV lane's characteristics and its traffic. Some concurrent HOV lanes permit access at many locations. However, HOV lanes separated from general purpose lanes sometimes restrict access to a few interchanges, which effectively eliminates some traffic movements from the HOV lane. Design of an HOV lane is further complicated by the vehicle occupancy permitted to use the facility. The best design for an HOV lane for carpools of two or more people may be quite different from the preferred design for an HOV lane restricted to carpools of three or more people.

Estimated travel times on an HOV lane are as critical as accurate traffic forecasts. If the HOV lane offers a major time savings over general purpose expressway lanes, then shifts in mode choice and vehicle occupancy must be considered. An HOV lane will also reduce travel times for single-occupant vehicles because vehicles are removed from general purpose lanes. This means that the HOV lane may cause transit riders to shift to both SOVs and HOVs.

NETWORK EQUILIBRIUM WITH HOV LANES

Equilibrium traffic assignment with an HOV lane in the network is illustrated by the following example. Figure 2 shows the simplest possible network when only two paths are possible between an origin and destination. One path includes only conventional roadways, whereas the second path travels through an HOV lane at some point. All drivers may use the first path, but only drivers of HOVs may travel the second path. It is assumed that all HOVs will use the HOV path if it offers a time savings.

Trips between the two locations include a fixed number of carpools. The remaining trips are single occupant vehicles and trucks that may not use the HOV facility. The three diagrams in Figure 3 show the possible equilibrium assignments of SOV and HOV traffic onto the two paths, depending on the relative path travel times and traffic volumes for SOV and HOV traffic.

Each diagram shows two relationships between path travel time and the traffic on the conventional roadway and HOV lane paths. The HOV lane path travel time–traffic relationship points to the right and travel time via the HOV lane path is read on the left y-axis. The travel time–traffic relationship for the conventional roadway path points to the left. Travel time via the conventional roadway path is read from the right y-axis.

The full width of the x-axis represents all vehicles traveling between the origin and destination. However, the x-axis is divided into two parts for SOV and HOV traffic. HOVs are shown on the left side of the x-axis and SOVs on the right side. Any point on the x-axis, therefore, indicates the split between HOVs and SOVs in the trip interchange.

Figure 3(*a*) shows the unlikely situation of the HOV traffic on the HOV lane being equal to the equilibrium traffic volumes. The split between HOVs and other vehicles is such that traffic on the HOV lane and conventional roadway paths produce identical travel times. This is the equilibrium condition that would normally be reached if both paths were unrestricted conventional roadways.

Now suppose there are fewer HOVs than the equilibrium volume. This situation is depicted in Figure 3(*b*). Equilibrium

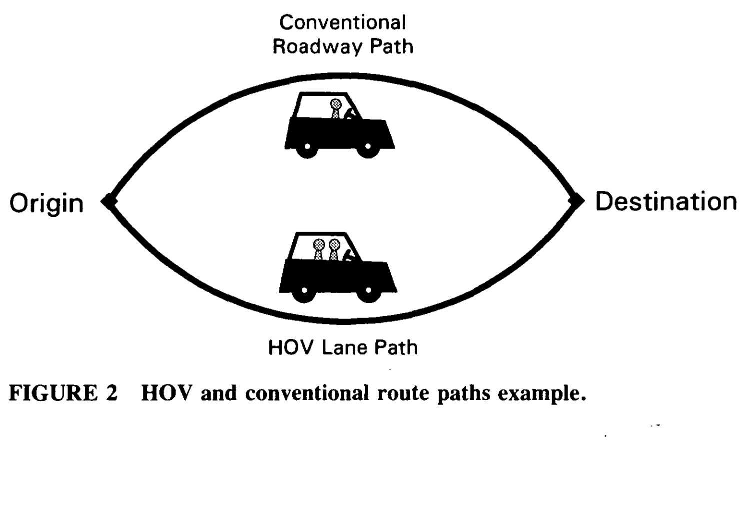

FIGURE 2 HOV and conventional route paths example.

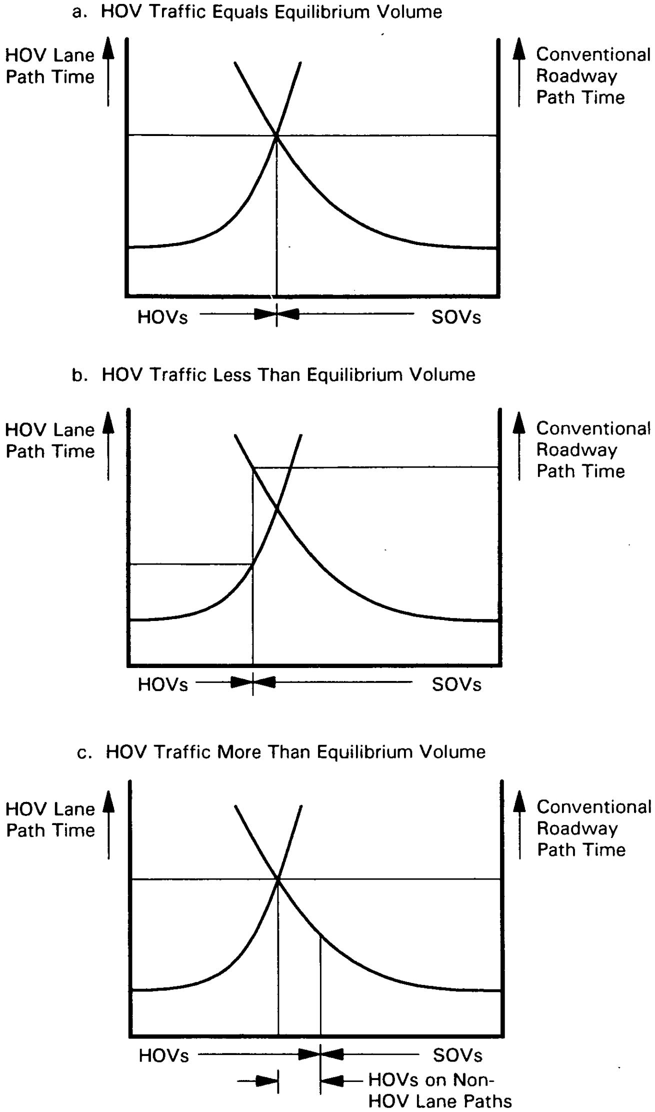

FIGURE 3 Possible equilibrium assignments of SOV and HOV traffic.

traffic conditions cannot be reached because of the restrictions barring SOVs from the HOV lane. Trucks and single occupant vehicles may not freely transfer from the slower conventional roadway path to the faster HOV lane path. Travel time via the HOV lane path remains less than via the conventional roadway path. In practice, this is the desired situation so that the HOV lane provides a higher level of service.

The third case shown in Figure 3(*c*) is the most complex of the three diagrams. There are more HOVs than the equilibrium volume. HOVs spill over into the conventional roadway path since they may transfer between the two paths. Drivers of HOVs make path choices to bring about equilibrium traffic conditions, and travel times via the HOV lane and conventional roadway paths become equal.

The algorithm for equilibrium traffic assignment determines those traffic volumes that minimize the sum of the areas under the travel time–traffic relationship between zero and the assigned traffic for all links. As discussed previously, this would normally produce equal travel times among all paths between the same origin and destination. This algorithm is still valid for traffic assignment if the network contains an HOV lane. Minimizing the sum of the areas under the link travel time–traffic relationships will produce one of the latter two solutions shown in Figure 3, depending on the availability of HOVs. In contrast to usual equilibrium assignment, travel times via HOV lane and conventional roadway paths may differ because too few HOVs are present to reach equilibrium traffic conditions.

THE STEVENSON EXPRESSWAY FEASIBILITY STUDY

The Stevenson Expressway (I-55) HOV lane project featured considerable interaction between the project consultant and CATS. The consultant specified an alternative HOV lane and CATS then coded the design features of the HOV lane into the agency's traffic assignment network. Morning and evening peak-hour traffic assignments were completed by CATS and the results transmitted to the consultant for their evaluation. The traffic assignment for one HOV lane alternative would then be used in developing the next HOV lane alternative to be tested. The recommended HOV lane evolved through four HOV lane alternatives during the course of the project.

Coding of the HOV Lane in the Traffic Assignment Network

Figure 4 depicts network coding for an eastbound concurrent HOV lane alternative next to the Stevenson Expressway general purpose lanes in the CATS regional traffic assignment network. This network covers the six county northeastern Illinois region and includes more than 30,000 one-way links. Movements between general purpose lanes and the HOV lane are permitted at the weaving locations shown in Figure 4. The HOV lane is separated from the general purpose lanes elsewhere.

A separate set of HOV links are coded parallel to the links for the existing Stevenson Expressway general purpose lanes. Connecting links are included where it is possible to move

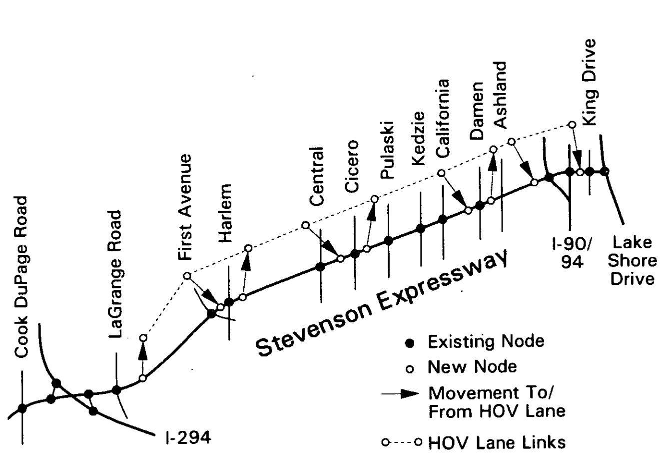

FIGURE 4 Coding of concurrent HOV lane.

between the HOV lane and the general purpose lanes. They are coded without travel time because these connecting links represent only a merge into an adjacent lane.

This network coding is done with the UTPS program HNET. The network is maintained in the UTPS N and Z-file formats. In these formats, each link attribute is tabulated in a vector containing an entry for every link in the network. The HNET-UROAD programs allow multiple HOV lane alternatives and the base network to be combined in one network. HOV lane and HOV–general purpose lane connecting links are coded with a link use code that allows them to be included or excluded from the network during assignment in program UROAD.

Figure 5 shows the network coding at an interchange in more detail. It shows an existing interchange between an arterial street and the eastbound lanes of the Stevenson Expressway. The concurrent HOV lane is coded adjacent to the general purpose Stevenson lanes. The connecting links are represented as links 1-2 and 6-7.

Major weaving movements occur on links 2-3 and 5-6 on the general purpose Stevenson lanes. These links are important in the minimum time paths because considerable travel time may be required to move from the HOV lane across the general purpose lanes to the exit ramp, as well as for the reverse movement from the entrance ramp to the HOV lane. In the Figure 5 diagram, an exit from the HOV lane to the cross-street would be represented as a movement through links 1-2, 2-3, and 3-4, a connecting link, weaving link, and ramp link. A movement from the cross-street to the HOV lane would follow the link 4-5, 5-6, and 6-7 path.

HOV Lane Traffic Assignment Algorithm

The equilibrium assignment procedure available in the UTPS program UROAD was adapted to perform an equilibrium assignment with an HOV lane coded in the network. The HOV lane assignment is not performed within UROAD, but relies on UROAD for path building and assignment.

A short program written at CATS replaces the part of UROAD that merges assignments from separate iterations of the equilibrium assignment algorithm. This program was required because UROAD does not allow different networks to be simultaneously assigned. Development of a separate program appeared to be a better option than changing UROAD because the program is extremely complex and is difficult to reliably alter.

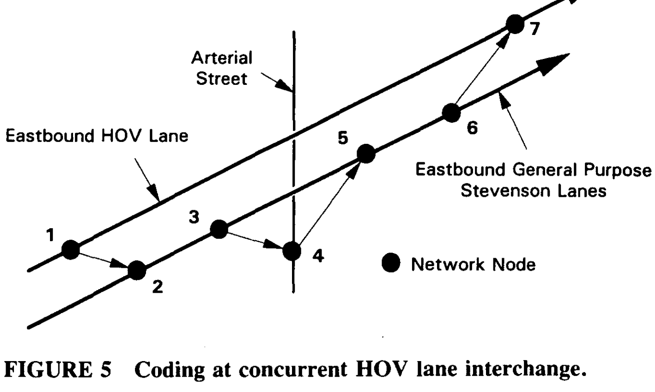

FIGURE 5 Coding at concurrent HOV lane interchange.

The following is a summary of the HOV lane equilibrium assignment algorithm and its implementation with UTPS software:

1. An initial feasible traffic assignment on the network with the HOV lane is completed. This starting point assignment includes link volumes from an assignment of SOV trips plus the link volumes from an assignment of HOV trips (carpools of two or more people or three or more people).
   a. HOV lane links are coded into a base network using HNET. HOV lane links have a use code of 2; all other links have a use code of 1.
   b. The SOV trip table is assigned to all network links with a use code of 1 with the program UROAD. This initial SOV assignment may be an all-or-nothing, incremental, or equilibrium assignment because it is just the starting point for the algorithm.
   c. The HOV trip table is assigned to all links in the network (use codes 1 and 2) using UROAD. This first HOV assignment also may be an all-or-nothing, incremental, or equilibrium assignment. Initial link travel times may either be uncongested link times or the times after the initial assignment of SOVs.
2. Travel times in the network are updated based on the combined SOV and HOV link volumes. Program UMATRIX adds the volumes from the two assignments into total link volumes, then computes new link travel times with the totaled SOV and HOV link volumes.
3. An all-or-nothing SOV and HOV combined assignment is completed. This requires two separate UROAD runs.
   a. The SOV trip table is assigned to its network (links with use code 1) using the updated travel times from Step 2 in an all-or-nothing assignment.
   b. The HOV trip table is assigned to its network (links with use codes 1 and 2) also with the updated travel times from Step 2 in an all-or-nothing assignment.
   c. Link volumes from the HOV and SOV all-or-nothing assignments are added together to obtain total HOV and SOV link volumes using the program UMATRIX.
4. The initial assignment HOV and SOV link volumes from Step 1 are merged with the total HOV and SOV all-or-nothing link volumes from Step 3. A CATS program that is compatible with UTPS computes the optimal linear combination of the initial and all-or-nothing assignments to form the merged assignment. The fractions (usually noted as λ and $1-\lambda$) of the two assignments in the merged assignment minimize the summed areas under the conventional roadway and HOV lane link travel time-traffic relationships.
   a. The λ value is determined by the search procedure in the CATS program.
   b. Program UMATRIX writes out new link volumes computed by

$$(1 - \lambda) * \text{initial HOV and SOV volumes} + \lambda * \text{all-or-nothing HOV and SOV volumes} \quad (1)$$

5. The link travel times are recalculated with the link volumes from Step 4 by UMATRIX.

6. The algorithm returns to Step 3. The link travel times from Step 5 are used for the next SOV and HOV all-or-nothing assignments. The current solution link volumes determined in Step 4 replace the initial assignment in the algorithm. Additional iterations of the algorithm continue until link volumes are approximately equal from one iteration to the next.

Mode Choice and Vehicle Occupancy Adjustments

Only route choice is dealt with by the HOV lane traffic assignment. However, mode choices and vehicle occupancies will also adjust if the time savings from traveling the HOV lane are substantial. Other modeling steps are needed to introduce mode and vehicle occupancy choices into the HOV lane traffic forecasts. The entire model process used to forecast traffic on the Stevenson HOV lane alternatives is summarized in Figure 6.

Trip tables are developed from several sources. Daily work trip tables are assembled from the 1980 census (the 1990 file is not yet available) journey to work data (*6*), which contains detailed work trip mode choice and vehicle occupancy data. The census work trips are expanded to 1990 using regional employment and households. Base 1990 trip tables from CATS long-range planning supply the daily nonwork automobile, transit person, truck, and external vehicle trip tables for the project. Nonwork trips are further factored by automobile occupancy.

These trip tables were also adjusted for a new rail transit service that will shortly begin operation in the eastern part of the Stevenson corridor. Transit mode shares from the alternatives analysis for this new line were used to factor selected trip interchanges. The net effect is to reduce automobile trips in zones served by the new line.

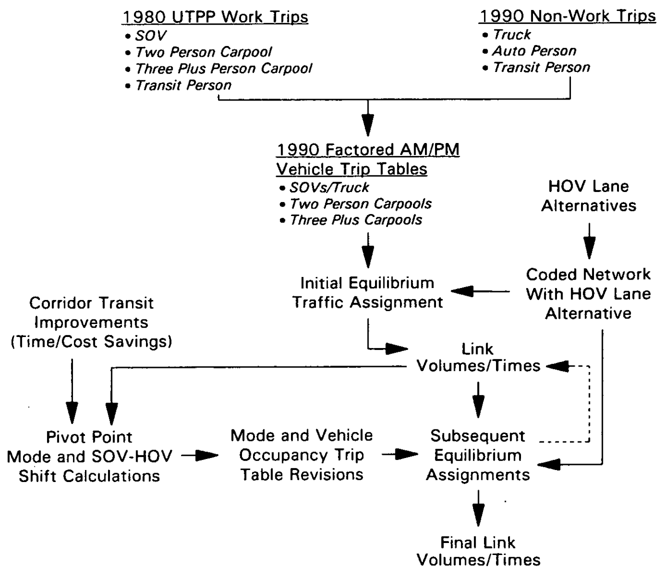

FIGURE 6 Final Stevenson HOV lane traffic forecasts.

As a final step, the daily trip tables are factored to morning (7:00 to 8:00 a.m.) and evening (4:00 to 5:00 p.m.) peak hours. The resulting peak-hour trip tables strongly reflect the work trips reported in the census.

After this processing, three a.m. and p.m. trip tables are available for assignment: SOVs and truck; two-person carpools; and carpools of three or more people. The two-person carpool trip table is included in the SOV or HOV trip table assigned, depending on the carpool occupancy requirements of the HOV lane.

An HOV lane alternative is coded into the network, and the initial trip tables are assigned. This first assignment provides link travel time estimates. The time savings between the HOV lane paths and non-HOV lane paths, as well as any overall reduction in highway travel times caused by the added capacity of the HOV lane, may be estimated from these travel times.

Transit-Automobile Mode Choice Adjustments

The bottom portion of the model process shown in Figure 6 assumes that the HOV time savings are large enough to significantly alter mode and vehicle occupancy choices. To adjust mode choices, automobile and transit trip tables are revised through pivot-point calculations. The pivot-point method approximates the change in mode share that would be estimated by a logit mode choice model given some change in the time and cost characteristics of a mode. The general pivot-point calculation for this trip table adjustment is as follows:

$$\Delta \text{ mode share} = \text{mode share} * (1 - \text{mode share}) * \text{time savings} * \text{model time coefficient} \quad (2)$$

Equation 2 is first used to estimate the transit ridership diverted to HOVs given the change in HOV travel time due to the HOV lane. All trips shifted from transit to automobile are calculated using base automobile and transit mode shares, time savings via HOV lane paths, and the in-vehicle time coefficients from the CATS logit automobile-transit primary mode choice model (*7*). Because only carpool trips can benefit from the HOV lane time savings, the total transit trips diverted must then be factored downward by the ratio of carpool person trips to all automobile person trips.

A similar calculation estimates the change in SOV mode share as a result of the improved operating conditions for SOVs after the HOV lane is opened. Trips shifted from transit to automobile are again calculated, but with the before and after HOV lane travel time savings for SOVs. Because transit ridership shifted to HOVs is already determined, the total transit trips diverted are factored by the ratio of SOV person trips to all automobile person trips.

The CATS mode choice model in-vehicle time coefficients are presented in Table 1. Transit improvements in the corridor may also be introduced into the Figure 6 modeling process through transit cost and time savings and additional automobile-transit pivot-point calculations. All trip table adjustments are carried out with the UTPS program UMATRIX.

TABLE 1 CATS Mode Choice Model In-Vehicle Time Coefficients (minutes)

Category of Trip	Model Coefficient
Work Trip to Central Business District	0.0159
Work Trip to non-Central Business District	0.0186

Vehicle Occupancy Adjustments

To use the pivot-point approach for vehicle occupancies, mode choices are replaced with vehicle occupancy choices. These choices correspond to the project drive alone, two-person carpool, and three-or-more-person carpool trip tables. The vehicle occupancy model structure and coefficients for in-vehicle time savings that affect the choice of vehicle occupancy are from the model developed by COMSIS Corporation for the Maryland National Capital Parks and Planning Commission (MNCPPC) (*8*).

The portion of the MNCPPC model that is used in the project features two nested choices for the level of ride sharing and drive versus shared ride. A logit shared ride model determines the proportions of two-person, three-person, and four-or-more-person carpools. A higher level logit model is linked to the shared ride model and allocates automobile users to drive alone and shared ride options. Coefficients from the MNCPPC model used in the vehicle occupancy pivot-point calculations are presented in Table 2.

The pivot-point calculations depend on the carpool occupancy permitted on the HOV lane. If two or more person carpools are permitted on the HOV lane, then a single pivot-point calculation is used to predict the shift from drive alone to carpooling. Two pivot-point calculations are needed if the minimum HOV lane automobile occupancy is raised to three or more persons. In this case, one pivot-point calculation estimates automobile drivers shifted to carpools of three or more people, and a second pivot-point predicts the two-person carpools that are shifted to three-person carpools.

After completing the pivot-point calculations, the trip tables are revised for additional traffic assignments. In Figure 6, the dashed line shows potential iterations through the pivot-point and traffic assignment steps. The intent is that the trip tables and travel times resulting from the traffic assignment should be brought into near equilibrium. This iterative procedure is far from robust, but the trips shifted from transit and to higher vehicle occupancies will generally not be large enough to greatly affect zone-to-zone travel times because route choices are also adjusting during each iteration.

TABLE 2 MNCPPC Mode Choice Model In-Vehicle Time Coefficients (minutes)

Category of Trip and Auto Occupancy	Model Coefficient
Work Trip Shared Ride Choice (2 Person, 3 Person, 4 or More Person Carpool)	0.1380
Work Trip Drive Alone-Shared Ride Choice (Drive Alone, Carpool)	0.0740

SAMPLE HOV LANE ASSIGNMENT RESULTS

Some results from an HOV lane assignment are presented in this section. The HOV lane alternative is similar to the one shown previously, except that it has a separated HOV lane configuration with direct ramps onto the HOV lane. A single HOV lane is provided in each direction.

Eight iterations of the HOV lane assignment algorithm were completed. The SOV trip table is loaded onto a maximum of nine different paths between each origin-destination zone pair—the paths from the initial feasible carpool assignment plus eight paths built for all-or-nothing assignments during the assignment algorithm.

The λ values determined for the eight iterations of the algorithm are presented in Table 3. The adjacent column shows the contribution of each assignment to the final link volumes. These shares are computed by multiplying each assignment's initial fraction times $(1 - \lambda)$ for subsequent iterations. All λ values are fairly small, which implies the initial assignment was reasonably close to equilibrium conditions. Slightly more than 69 percent of each link's final traffic volume comes from the initial feasible assignment.

Travel times and assigned traffic volume are compared with capacity ratios via the Stevenson general purpose lanes and the parallel HOV lane in Table 4. These quantities are measured over the eastern two-thirds of the HOV facility, from Harlem Avenue to the Dan Ryan Expressway, a distance of 15.0 km (9.3 mi). Uncongested or free-flow travel times are the same, 89 km/hr (55 mph), for the HOV lane and Stevenson general purpose lanes.

Travel times via the HOV lane and general purpose lanes are reasonably stable after eight iterations. The HOV lane is 4.4 min faster over this section than the general purpose lanes on the basis of the travel times calculated from the link volumes after the algorithm's eighth iteration. This is equivalent to a speed differential of 10.1 km/hr (6.3 mph) between the HOV lane and general purpose lanes. There are not enough carpool trips to reach equilibrium volumes and travel times, which corresponds to the desirable situation depicted in Figure 3(*b*).

The average volume to capacity ratios on the HOV lane and Stevenson Expressway general purpose lanes between Harlem Avenue and the Dan Ryan Expressway are also summarized in Table 4. Average volume-to-capacity ratios are obtained by dividing total vehicle-kilometers traveled by the vehicle-kilometers of available capacity. The lane capacity is

TABLE 3 Eight Iteration HOV Lane Assignment Results

Assignment	λ	Contribution to Link Volumes
Initial Feasible		0.6923
First All-or-Nothing	0.0389	0.0280
Second All-or-Nothing	0.0394	0.0295
Third All-or-Nothing	0.0477	0.0376
Fourth All-or-Nothing	0.0517	0.0429
Fifth All-or-Nothing	0.0456	0.0397
Sixth All-or-Nothing	0.0483	0.0442
Seventh All-or-Nothing	0.0474	0.0455
Eighth All-or-Nothing	0.0403	0.0403

TABLE 4 Morning Peak-Hour Assignment Times and Volume-to-Capacity Ratios Between Harlem Avenue and Dan Ryan Expressway

	HOV Lane		General Purpose Lanes	
Algorithm Step	Travel Time (min.)	Volume to Capacity Ratio	Travel Time (min.)	Volume to Capacity Ratio
Initial Feasible	33.9	1.43	20.6	1.19
First Iteration	30.6	1.38	21.0	1.20
Second Iteration	27.7	1.33	20.5	1.19
Third Iteration	24.9	1.28	20.8	1.19
Fourth Iteration	22.4	1.22	20.9	1.20
Fifth Iteration	20.6	1.18	21.0	1.20
Sixth Iteration	19.1	1.14	21.5	1.22
Seventh Iteration	18.1	1.10	21.7	1.22
Eighth Iteration	17.7	1.09	22.1	1.23

1,650 vehicles per hour. After the eighth iteration, the HOV lane has approximately 230 fewer vehicles per hour per lane than the general purpose Stevenson lanes. This difference in lane volumes again indicates that equilibrium conditions will not be reached because of limited HOV trips.

FINAL COMMENTS

Procedures developed at CATS for traffic assignment when the network includes an HOV lane are documented in this paper. The intent was to develop an HOV lane assignment algorithm that was consistent with equilibrium assignment principles and could be carried out with CATS network coding and assignment software. Most of the discussion is directed toward the practitioner who must work with the models and resources available at a metropolitan planning organization.

Many outstanding questions remain about HOV lane assignment practices. Accurately coding the characteristics of the many different HOV lane designs in a traffic assignment network is a challenge for the practitioner. The best approach to prepare the initial assignment of HOV and SOV also entails research. Acceptable equilibrium conditions and the number of assignment algorithm iterations required may vary substantially among HOV lane projects.

Future year traffic assignments were not prepared by CATS for the consultant's use. Project trip tables are based on reasonably current observed travel behavior. Following a similar procedure for future HOV traffic assignments appeared questionable at best. One may readily argue that underlying travel behavior may change substantially in the long-term. The Intermodal Surface Transportation Efficiency Act of 1991 and the requirements of the Clean Air Act Amendments of 1991, plus increasingly congested traffic operating conditions, will likely affect automobile occupancy and traffic peaking. Whether the automobile occupancy and peak-hour factors used to create the 1990 trip tables will remain valid in the future is uncertain.

Although the procedure described in this paper may appear awkward, it allows the analyst to track the progress of the assignment as it iterates. This may be the best way to determine whether volumes on the HOV links are nearing steady-state conditions. Traffic forecasts for HOV lanes will generally require more detailed analysis of assignment results than conventional traffic assignments.

ACKNOWLEDGMENTS

The author acknowledges the contributions of Karen Richter of the Puget Sound Regional Council and Ron Shimizu of the Regional Transportation Authority. Their suggestions and support for the work described in this paper are greatly appreciated.

REFERENCES

1. *Operation GreenLight: 1st Annual Report.* Chicago Area Transportation Study, Chicago, 1990.
2. *High Occupancy Lane Feasibility Study: Stage One Report.* Chicago Area Transportation Study, Chicago, 1991.
3. HNTB Corporation. *Feasibility Report: Interstate Route 55 (Stevenson Expressway) High Occupancy Vehicle Lane Study.* Pre-Final Draft Report. Illinois Department of Transportation, District 1, March 1993.
4. *Urban Transportation Planning System: Reference Manual.* UMTA and FHWA, U.S. Department of Transportation, 1982.
5. J. G. Wardrop. Some Theoretical Aspects of Road Traffic Research. *Proc., Institute of Civil Engineers,* Vol. II, No. 1, 1952, pp. 352–378.
6. *Urban Transportation Planning Package 1980 Census: Technical Documentation for Summary Tape.* Bureau of the Census, Journey-to-Work and Migration Statistics Branch, Population Division, 1983.
7. *Network Sensitive Mode-Choice Models: A Two Stage Approach to Mode-Choice Modeling for Use in Preparing Year 2000 Transportation System Development Plan.* Staff Technical Report. Chicago Area Transportation Study, Chicago, 1975.
8. J. Ryan and C. Y. Yeng. Development of the New MNCPPC Nested Logit Model (with HOV Options, New Survey and Skim Data). Comsis Corporation Memorandum to Michael A. Replogle, Director of Transportation Modeling, Maryland National Capital Parks and Planning Commission, Dec. 1988.

Publication of this paper sponsored by Committee on High-Occupancy Vehicle Systems.

High-Occupancy Vehicle Facility Safety in California

E. C. SULLIVAN AND N. DEVADOSS

The findings of a study on how high-occupancy vehicle (HOV) facility operation affects the safety of selected California freeways are documented. The principal conclusion supported by the findings is that the accident rates on freeways with HOV facilities are sensitive to variations in traffic congestion in much the same manner that accident rates are influenced by congestion on non-HOV facilities. Careful investigation of the patterns of accident characteristics revealed no systematic differences in the lane locations of accidents or other factors that could be attributed directly to the presence of the HOV facilities. Freeways with and without HOV facilities appear to be characterized by the presence of locations where peak-period accidents are clustered, typically because of localized congestion conditions.

California has led the country in efforts to use freeway high-occupancy vehicle (HOV) lanes as tools to increase passenger-mile capacities by encouraging travelers to shift from single-occupant vehicles to carpools, vanpools, and public transit. California Department of Transportation (Caltrans) districts in major urban areas continue to advance programs to integrate HOV facilities into new freeway construction and reconstruction projects.

Despite their popularity among transportation professionals, HOV facilities invite controversy on several fronts. One such area of controversy is the question of HOV safety, especially where HOV lanes and mixed-flow lanes operate as a common facility. This situation provides opportunities for traffic conflicts between the rapidly moving HOV lane and stop-and-go traffic using adjacent mixed-flow lanes. At locations where vehicles begin and end HOV operation, conflicts may also arise from vehicle lane changes in the presence of significant speed differentials between lanes.

It is also true that when drivers are confronted by obviously difficult conditions, they may exercise so much extra care that the overall facility operation may actually be safer. Because of these opposing factors, the only valid way to conclude whether freeways with HOV lanes have special safety problems compared with fully mixed-flow facilities is to measure the traffic safety performance of these facilities in the field.

This study is not the first to attempt to quantify the traffic safety consequences of HOV facilities, but it is the most comprehensive investigation undertaken to date. The purpose of the study is to conduct a detailed analysis of the traffic safety impacts of a number of HOV facilities on California freeways. This will provide quantitative evidence to show if some or all of the selected HOV freeway sections differ in accident experience from control sections on similar non-HOV freeways.

Department of Civil and Environmental Engineering, California Polytechnic State University, San Luis Obispo, Calif. 93407.

PREVIOUS RESEARCH

The majority of previous research on HOV implementation focused on "effectiveness," usually defined as a measure of how many people are transported through the facility. Safety was often mentioned as a concern, but only occasionally was it specifically studied. Betts et al. and others have indicated that traffic conditions on HOV facilities present a higher potential for accidents involving injuries (*1*).

Several studies conducted in California and Texas deal with the safety of implementing a concurrent-flow HOV lane. The results are varied, but appear to follow a general pattern: accidents increase in the first months of operation, presumably due to driver unfamiliarity with the system. After a period of several months, accident rates tend to stabilize at either pre-project or slightly higher levels (*2–7*).

Accident rates were evaluated before and after installation of HOV facilities with and without physical separation. Where the HOV lane was separated from the mixed-traffic lanes, no upward surge in accident rates was discernable (*4*). In several southern California studies of before and after accident rates, increased congestion was named as the major factor contributing to the increase in accident rates after HOV implementation (*5–7*).

Conclusions on the use of data for analysis vary. Some studies conclude that aggregate data (such as annual accidents per million vehicle-miles) should not be used to evaluate safety impacts (*8*). Instead, data stratified according to time of day, type of accident, and other operational characteristics should be used to provide insight into the safety effects of an HOV lane addition. In one study the data were stratified for analysis (*5,9*), and in a parallel investigation the corresponding aggregate data were used (*10*). Very different results were obtained from the two types of analysis. Measures of safety used in accident analysis are a complex topic of continuing research.

STUDY APPROACH

The group of HOV facilities considered in this study vary considerably in design. However all are located in the freeway median and have one of three basic designs:

- Contiguous HOV lanes with unconstrained access along their entire lengths,
- Buffer-separated HOV lanes with controlled ingress and egress at the ends and at numerous intermediate locations via lane changes from adjacent mixed-flow lanes, and

- Barrier-separated HOV lanes with controlled ingress and egress, typically using special ramps at each end and at a limited number of intermediate locations.

The safety impacts of particular freeway sections representative of each of these designs were studied using three interrelated but distinct approaches:

- Statistical analysis of computerized accident records from the Caltrans Traffic Accident Surveillance and Analysis System (TASAS): Accident frequencies and characteristics for the HOV freeway sections during their hours of operation were compared with the same operating periods for non-HOV freeway sections (control sections).
- Examination of selected original accident reports: California Highway Patrol (CHP) personnel provided field reports containing additional detail on the characteristics of certain accidents at locations where accidents tend to occur (or "hot spots"), which provided insights not attainable from the computerized data records alone.
- Visual analysis of HOV and control facility operations: Driver behavior was studied in relation to the traffic conditions observed. The visual analysis was performed using videotapes of peak-period traffic operations at selected locations along both the HOV and control facilities.

The comparisons made between the HOV and control sections generally focused on the heaviest travel directions. It should be noted that the sections vary considerably in their imbalance of directional peak-period volumes.

CHARACTERISTICS OF SELECTED FREEWAYS

HOV sections and control sections were selected on the basis of their similar traffic and geometric characteristics. Some of the selected HOV facilities operate at all times, others during limited hours, outside of which the HOV lanes either revert to mixed-flow use or are closed to traffic.

It should be emphasized that this study addressed cross-sectional comparisons (HOV sections against control non-HOV sections) instead of before and after comparisons relative to HOV facility implementation. Although imperfect, the cross-sectional approach was chosen to avoid the influence of changes in traffic volumes and other underlying conditions over time. Resources did not permit both cross-sectional and before and after comparisons to be combined in the study.

The identities and basic characteristics of the HOV facilities and control sections considered in the study are presented in Table 1.

DATA SOURCES

A variety of data types from Caltrans and CHP were needed to carry out the statistical analysis. The following are data types used in the analysis:

- TASAS accident data,
- Traffic counts,
- Speed data,
- MODCOMP data,
- Vehicle occupancy counts, and
- Video data.

TABLE 1 Characteristics of Selected Study Sections

HOV SECTIONS								
Section	Highway	County	Post-miles	Year Opened	Number of Directional Lanes		Section Type	Operating Hours
					Mixed	HOV		
LA 10	I-10	Los Angeles	LA 101 1.57 - LA 10 29.19	1973	4	1	Barrier and Buffer	24 hr./ 7 days
MRN 101	U.S. 101	Marin	MRN 2.00-17.80	1974	3	1	Conti-guous	6:30-8:30 A.M. SB [1] and 4:30-7:00 P.M. NB [1]
ORA 405	I-405	Orange	ORA 0.23-24.18	1989	4	1	Buffer	24 hr./ 7 days
SCL 101	U.S. 101	Santa Clara	SCL 38.3-SM 1.87	1986	3	1	Conti-guous	5-9 A.M. [1] 3-7 P.M. [1]
SD 15	I-15	San Diego	SD 11.10-M19.47	1988	3	2	Barrier	6-9 A.M. SB [2] 3-6:30 P.M. NB [2]

1 HOV lanes used as mixed flow lanes during other periods
2 Reversible HOV lanes closed during other periods.

ANALYSIS RESULTS

Because of the magnitude of the study and the large quantity of information produced, it is possible to provide only a sample of the results obtained.

The results that follow are associated with two HOV facilities (SCL 101 and MRN 101) in the San Francisco Bay Area in northern California and three HOV facilities (LA 10, ORA 405 and SD 15) in southern California. Results for the SCL 101 section are discussed in greater detail; the results for other HOV sections are discussed briefly. The results presented for the southern California HOV facilities generally reinforce the conclusions reached in northern California. For the complete findings of the study, the reader is referred to the final report (*11*).

SCL 101 HOV Versus SM 101 and CC 80 Control Sections

Comparison of Accident Rates

The time period from January 1, 1989, to September 30, 1990, was selected to make accident comparisons between SCL 101, the HOV facility, and two control sections (SM 101 and CC 80). The number of accidents in TASAS remaining after the elimination of those occurring under atypical conditions (bad weather, bad road conditions, during maintenance activities, etc.) are presented in Table 2.

The larger number of accidents on SCL 101 compared with SM 101 and CC 80 may be due in part to differences in scale and traffic volumes, or other variations in exposure, such as more lanes or a longer section. Unfortunately SCL 101 was one of two study sections for which traffic count coverage was not available for the time periods considered. Consequently, in this case, only geometric features of the sections were used to standardize the accident counts. The accident rates given in Table 3 are standardized on a lane-mile basis.

Northbound (NB) U.S. 101 and westbound (WB) CC 80 serve commuter traffic bound for San Francisco in the morning peak-period hours, whereas southbound (SB) U.S. 101 and eastbound (EB) CC 80 serve the heavy out-bound evening commute.

Generally, morning and evening peak-period accident rates on a lane-mile basis are higher on the SCL 101 HOV facility than on the two control sections. However accident rates are similar during the midday and night hours. The similarity in off-peak rates suggests that there is no obvious design flaw that causes SCL 101 to be uniformly more dangerous than the control sections. Whether the peak-period differences may be attributed directly to the HOV lane operation requires further investigation at the disaggregate level.

Spatial Distribution of Accidents

Accidents tend to be concentrated at a few locations along any freeway. Density traces were developed to identify these "hot-spots." Each density trace shows the number of accidents within 0.05 mi of the indicated postmile locations. Figures 1 and 2 show density traces for accidents on SCL 101 in the heaviest directions for the morning and evening peak periods.

Figure 1 shows two large accident clusters for the morning peak period northbound. They are in the weaving sections between the Trimble Road on-ramp and Montague Expressway off-ramp (Postmile 40.93) and within the Stierlin Road interchange (Postmile 48.60).

Figure 2, for the evening peak period southbound, shows the largest accident clusters at Postmiles 40.68, 41.10, and 50.35. These are also located within major interchanges.

The corresponding congestion diagrams in Figures 3 and 4 show that the four major accident clusters, Postmiles 40.93 and 48.60 northbound and Postmiles 40.68 and 41.10 southbound, coincide with locations of localized congestion. The largest northbound clusters and many of the smaller accident clusters coincide with the congested transition section before the beginning of the higher capacity section containing the NB HOV lane. The congestion is caused largely by the bottleneck directly downstream of the Trimble Road on-ramp. Similarly, many of the southbound accident clusters coincide with the localized congestion that occurs upstream of a far right lane drop at the end of the SB study section. Based on the observations and discussions with Caltrans officials, it was learned that congestion occurs on a recurring basis, especially during peak periods.

In general most accident clusters occur close to on- or off-ramps and in locations of recurrent congestion. These accidents may be related to merging and weaving of traffic, especially at times of congestion. From the limited available

TABLE 2 Total Number of Accidents (SCL 101 and Control Sections)

SECTION	DIR	TIME OF DAY				TOTAL
		A.M.	P.M.	MIDDAY	NIGHT	
SCL 101 - HOV	NB	201	257	82	71	611
	SB	99	352	74	74	599
SM 101 - CONTROL	NB	54	138	95	78	365
	SB	71	147	72	80	370
CC 80 - CONTROL	EB	12	119	39	52	222
	WB	74	32	33	22	161

Note: A.M. = 5:00 a.m. - 9:00 a.m.;
P.M. = 3:00 p.m. - 7:00 p.m.;
MIDDAY = 9:01 a.m. - 2:59 p.m.;
NIGHT = 7:01 p.m. - 4:59 a.m.

TABLE 3 Accident Rates (Total Accidents/Lane-Mile)

SECTION	DIR	TIME OF DAY			
		A.M.	P.M.	MIDDAY	NIGHT
SCL 101 - HOV	NB	3.39	4.34	1.38	1.20
	SB	1.64	5.84	1.23	1.23
SM 101 - CONTROL	NB	0.96	2.44	1.68	1.38
	SB	1.26	2.60	1.28	1.42
CC 80 - CONTROL	EB	0.41	4.09	1.34	1.79
	WB	2.54	1.10	1.13	0.76

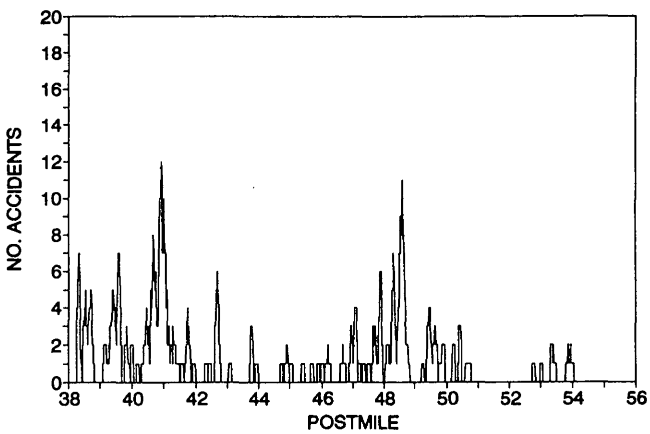

FIGURE 1 Morning accident frequency and locations for SCL 101-NB.

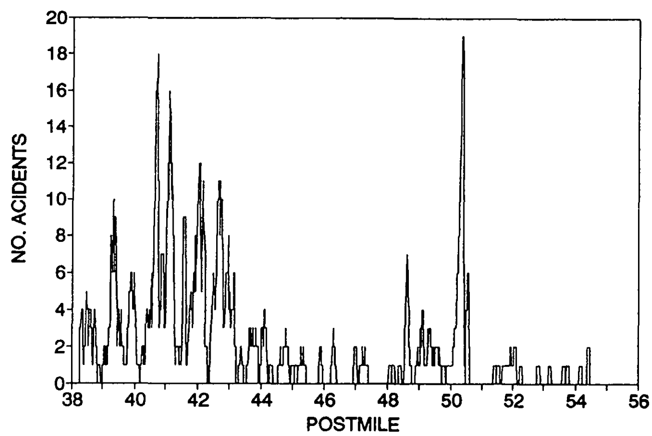

FIGURE 2 Evening accident frequency and locations for SCL 101-SB.

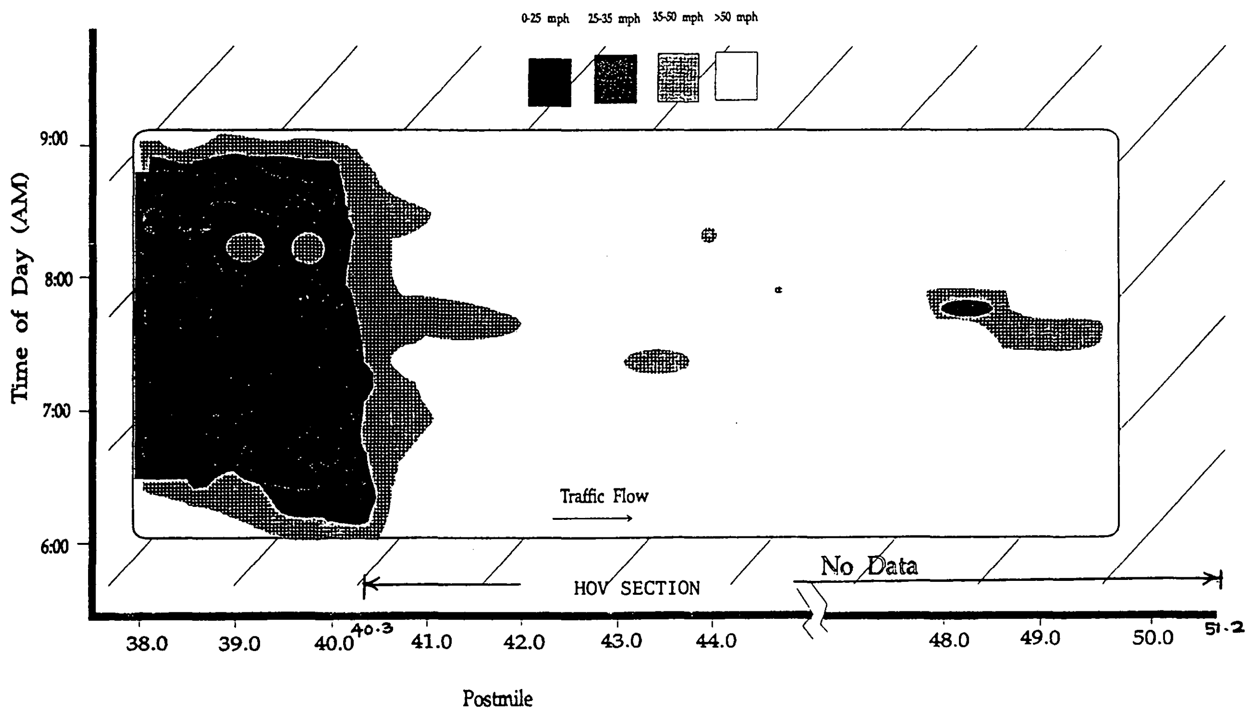

FIGURE 3 Morning congestion diagram for SCL 101-NB.

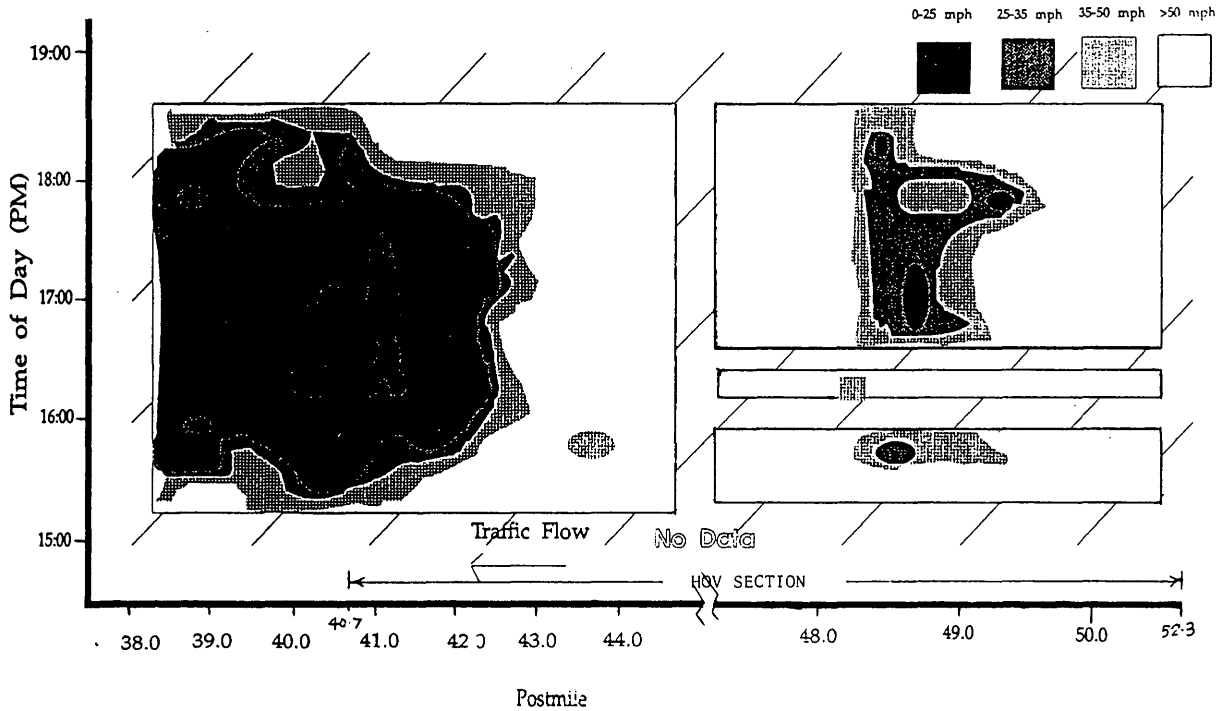

FIGURE 4 Evening congestion diagram for SCL 101-SB.

data, it was not possible to identify whether these weaving maneuvers are a result of drivers trying to use HOV lanes. The disaggregate analysis of the accidents with respect to this issue is discussed later.

Accidents Resulting in Fatalities or Injuries

Separate tabulations and density traces were developed for accidents resulting in fatalities or injuries to investigate whether their patterns might be different from those of total accidents, of which the majority involve property damage only. The accident rates and density traces for accidents resulting in fatalities or injuries generally reflect the patterns observed for total accidents. As in the case of total accidents, accidents involving fatalities or injuries coincide with the congested regions and occur close to the major on- and off-ramps.

Interaction Between Accidents and Speed

Numerous studies have identified a strong relationship between accidents and congestion because congestion tends to increase conflicts and driver frustration and often creates conflicts where high-speed vehicles do not have sufficient distance to stop.

On the basis of the expected times and locations of different speeds under recurrent conditions, the accidents in the peak-period direction were grouped to obtain the numbers of accidents by speed category. Each group of accidents was normalized by dividing by the lane-mile-hours associated with the speed range. The analysis shows somewhat higher accident rates for lower speeds. Although the relationship is not as strong as had been anticipated, the result does support the expected finding of a positive correlation between accident rates and increased congestion.

Types of Collisions

Major collision types are classified in TASAS data.

For the present analysis Figures 5 and 6 show comparisons of collision type on a percentage basis for the morning and evening peak-period directions. By far, the most frequent type of collision for all three sections is the rear-end collision.

One might hypothesize that any peculiar influence of HOV operations on accidents might be manifested in a shift in the relative proportions of rear-end and sideswipe accidents. The distribution of accidents by collision type is similar for all three study sections. The distributions for the other HOV study sections and their control sections also proved similar. Clearly, the hypothesis that the operation of HOV lanes during congested periods encourages particular accident types is not supported by these findings.

Accident Lane Location

Accident data were analyzed on the basis of the lane location of the accident. The collision location is coded in the TASAS data as follows:

- A. Beyond median or stripe—left;
- B. Beyond shoulder, driver's left;
- C. Left shoulder area;
- D. Left lane;

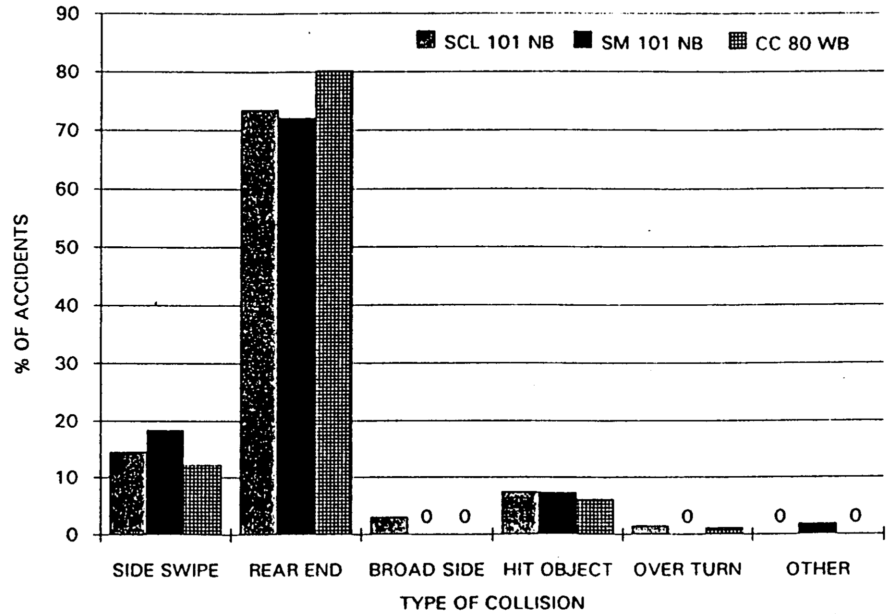

FIGURE 5 Morning collision type comparison.

- E. Interior lanes;
- F. Right lane;
- G. Right shoulder area;
- H. Beyond shoulder, driver's right;
- I. Gore area;
- <. Not stated; and
- –. Does not apply.

It should be noted that a value of accident location is coded for each vehicle involved, so the total number of accident locations exceeds the total number of accidents.

The categories listed above were grouped for analytical purposes. Categories A and B were combined as the Beyond Left Shoulder (BL) category, and were combined with Category C: Left Shoulder (LS). When the left lane operates as an HOV lane, the adjacent interior lane acts as the left lane for mixed-flow traffic, much as the left lane on control sections. Therefore to keep comparisons consistent between study sections and controls, Categories D: Left Lane (LL) and E: Interior Lanes (IL) were also combined. Category F was considered separately as Right Lane (RL), and Categories G, H, and I were combined as Right of Right Lane (RRL). The remaining categories were combined in a category labelled "Other."

The accident location summaries for SCL 101 are shown in Figures 7 and 8. In both figures, the highest accident per-

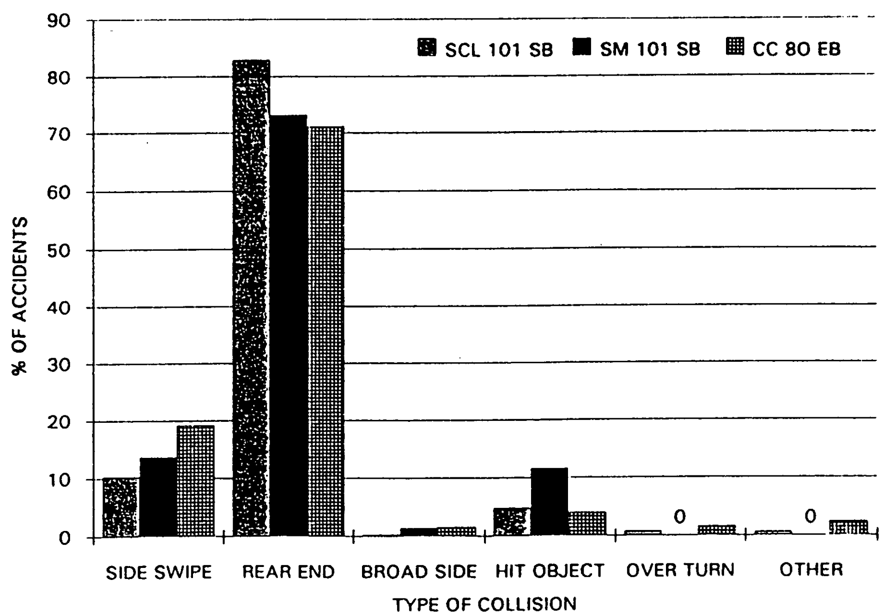

FIGURE 6 Evening collision type comparison.

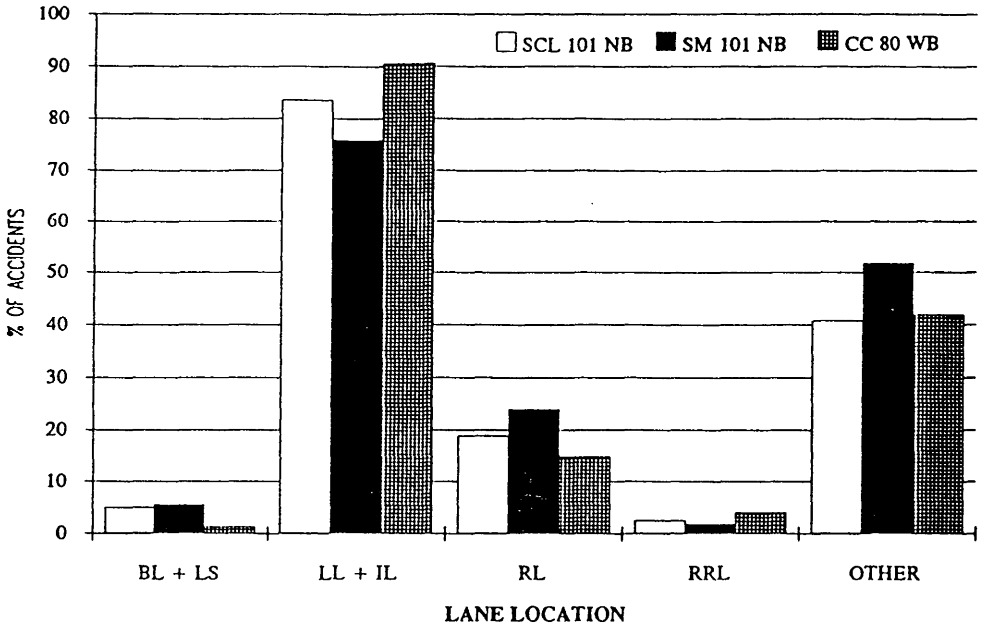

FIGURE 7 Comparison of morning accident lane location.

centages occurred in the left and interior lanes. If the HOV lane were an inherent contributing factor to accidents on SCL 101, one would expect to find consistently higher percentages in the LL + IL category compared with the control sections. Although the LL + IL percentage is higher for the evening peak period, no such effect is evident in the morning. Similar results were found for the other HOV study sections.

In conclusion, the comparison does not support the hypothesis that accident location is consistently different in the presence of HOV lanes.

Movement Preceding Collision

The categories for the analysis discussed in this paper are indicated in the TASAS data. As with lane locations, each observation corresponds to an involved vehicle. The percentages of vehicles in each classification for peak-period accidents in the study sections are shown in Figures 9 and 10.

It may be hypothesized that contiguous HOV lanes might lead to an increase in the number of accidents that occur when drivers are slowing or stopping or changing lanes. From the

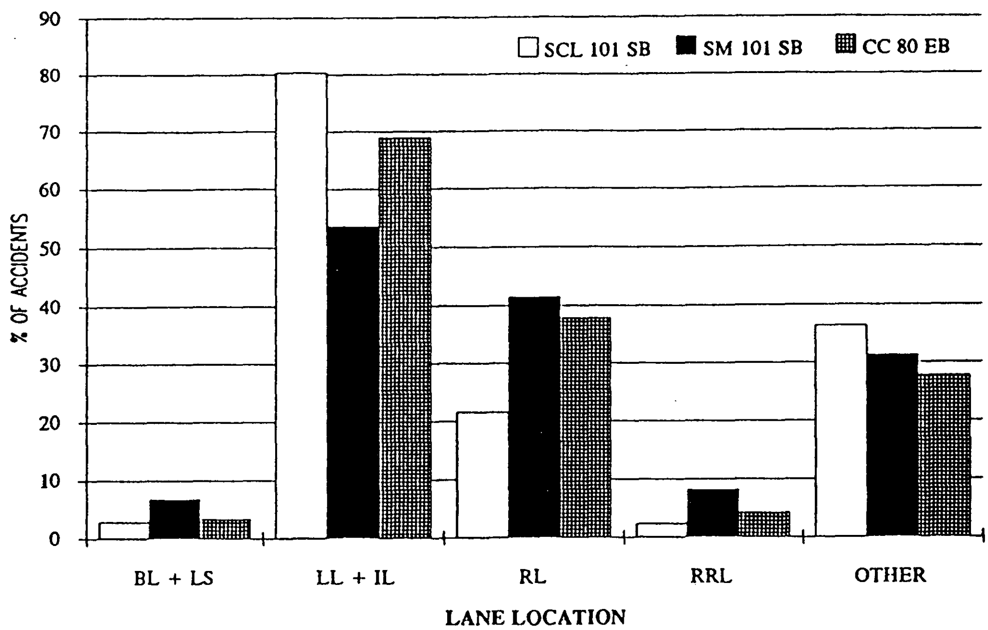

FIGURE 8 Comparison of evening accident lane location.

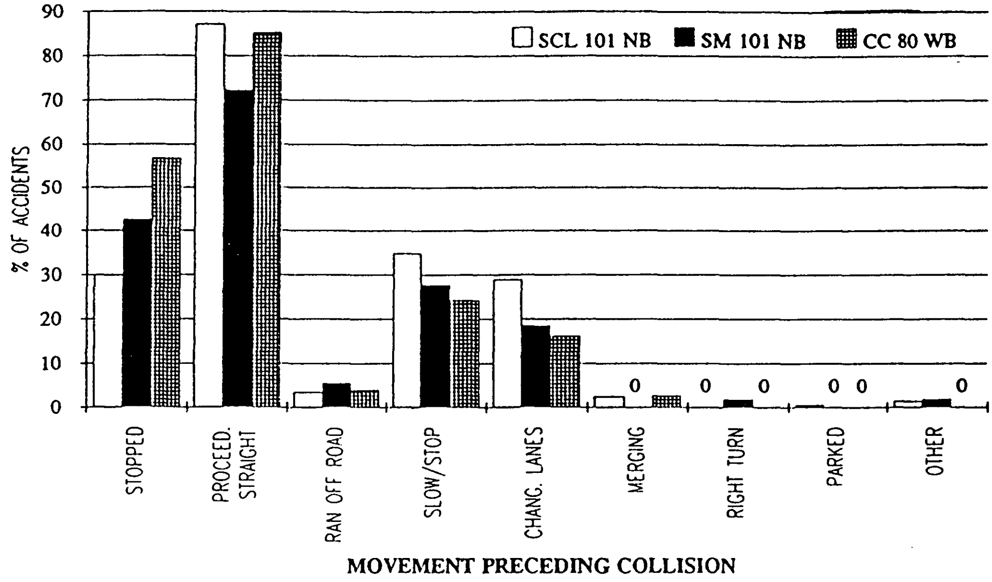

FIGURE 9 Comparison of movement preceding collision, morning.

present findings, the percentage of slowing or stopping accidents on SCL 101-NB is generally the same or less than on the control sections. However the percentage of accidents that occur when drivers change lanes on SCL 101 is always higher than on control sections, although the difference may be regarded as insignificant for the evening peak period.

It may be noted that accident percentages are higher in the "proceeding straight" category on SCL 101 in peak-period hours, possibly because of the influence of congestion during peak-period hours. However except possibly for the "change lanes" category, there is no clear pattern of difference that might be attributable to the presence of HOV lanes.

Results of the Review of CHP Individual Accident Reports

To further investigate accident characteristics at the "hot spots," the original written accident reports were reviewed. This investigation revealed congestion and resulting sudden slowdowns as the main causes of accidents. Of the accidents examined, the lane next to the HOV lane had the majority of accidents.

Most accidents at the hot spots were coded as rear-end accidents, indicating the contribution of speed differentials. Nearly all of the remaining accidents, coded as sideswipe

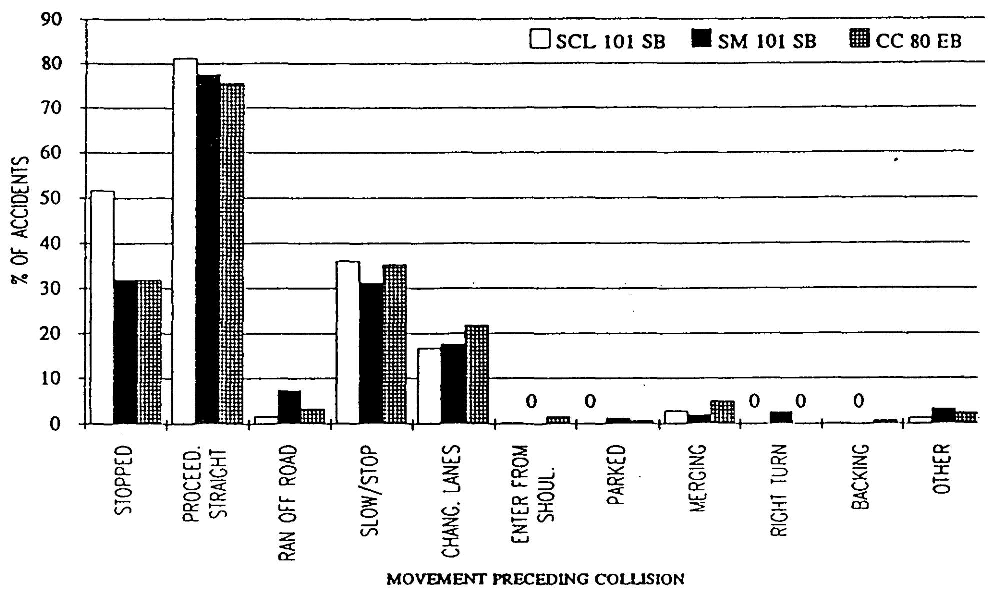

FIGURE 10 Comparison of movement preceding collision, evening.

accidents, appeared from the narratives and collision diagrams to be the results of driver efforts to avoid rear-end collisions. In some cases, drivers attempting to avoid rear-end collisions in the leftmost mixed-flow lane caused a sideswipe or rear-end accident in adjacent lanes. All accidents in the reports for SCL 101 belonged to one of these two categories. Congestion is clearly the major contributing cause of these accidents.

Video Data Analysis

The video data analysis was conducted to determine whether operational conditions on a per-lane basis in the mixed-flow lanes (densities, speeds, lane changes, etc.) are affected by the presence of an adjacent HOV lane, relative to the control sections.

Traffic parameters were manually extracted from the video recordings. These parameters, in combination with visual interpretation of vehicle motions in the field and video tapes, yielded better understanding of the possible causes of traffic accidents.

Data extracted from the video tapes were useful in understanding congestion patterns and how operating conditions varied among different sections. The video data were useful, for example, in revealing unusually high variations in values of flow and density between adjacent lanes in the vicinity of documented accident hot spots. These conditions, which appear unrelated to the HOV lane, clearly make drivers in the left mixed-flow lane especially vulnerable to becoming involved in accidents due to downstream congestion. The causes and consequences of these extreme lane-flow differentials warrant further study.

Summary of the Analyses

Accident rates on SCL 101 are systematically higher than on SM 101 and CC 80 during morning and evening peak-period hours, and are similar during the midday and night hours. It appears that any differences or similarities in accident rates between the SCL 101 HOV facility and the two control sections are largely a result of the differences in their congestion patterns and not of anything inherent in the geometric or operational characteristics of the HOV lanes themselves. The sole exception is that the proportions of accidents following lane change maneuvers are always greater in the presence of the contiguous HOV lane, although this effect is not characterized by amounts that can be considered significant.

MRN 101 HOV Versus CC 80 Control Section

The Marin 101 (MRN 101) HOV facility is unique in its operational and geometric character in that the HOV lane is contiguous and is in operation during the morning peak period (6:30–8:30 a.m.) only in the SB direction toward San Francisco and in the evening peak period (4:30–7:00 p.m.) only in the NB direction away from San Francisco. The facility consists of two HOV lane subsections in each direction separated by a 4- to 5-mi mixed-flow section. CC 80, on the other side of San Francisco Bay, was used as the control section because of its similar flow characteristics.

Analyses of accident rates based on lane-miles, vehicle-miles, and person-miles were compared. Accidents for lane-miles, vehicle-miles, and person-miles were consistently less than on CC 80 during the morning peak period, but higher for the evening peak period in the NB direction. The higher accident rates are associated with accident clusters at a few locations, especially in the gap between the HOV sections, which can generally be associated with localized congestion problems. Disaggregate analysis showed no noteworthy differences due to the presence of HOV lanes even during the evening peak period.

LA 10 HOV Versus LA 210 Control Section

An analysis was performed of accidents along the LA 10 barrier-separated HOV facility (El Monte Busway), parallel mixed-flow lanes, and terminal transition sections on I-10 and U.S. 101, in comparison with the LA 210 (Foothill Freeway) control section. Accident rates for lane-miles, vehicle-miles, and person-miles were calculated. It was found that accident rates on LA 10 are slightly higher than on the LA 210 control section. This may be attributed to the relatively heavy congestion on LA 10.

From the disaggregate analysis it was found that there is no evidence that the patterns of accidents recorded on LA 10 in the presence of the HOV facility is systematically different than on LA 210. Because the LA 10 HOV lanes are barrier-separated and operate as an isolated highway, except at transition sections, these findings are not particularly surprising.

ORA 405 HOV Versus LA 405 Control Section

Another analysis was performed of accident characteristics along the Orange County (ORA 405) buffer-separated HOV facility as compared with the LA 405 control section. The analysis for ORA 405 covered only 1 year because the HOV facility opened in May 1990. The LA 405 control section is of special interest because the standard median shoulders on that facility were reconstructed for use as mixed-flow lanes.

The accident rates for lane-miles, vehicle-miles, and person-miles are generally lower on ORA 405 than on the LA 405 control section, except for the morning peak period in the SB direction. However, fatal and injury accident rates are generally equal during most time periods.

From the disaggregate analysis it was found that there are no systematic differences between the ORA 405 HOV facility and the control sections during both the morning and evening peak periods.

Further investigations of the original accident reports revealed no evidence that the HOV facility specifically contributes to the incidence of accidents.

SD 15 HOV Versus LA 210 Control Sections

The reversible HOV facility north of San Diego on Interstate 15 is unique in several respects. It is a two-lane, barrier-separated facility that operates in the SB direction toward San Diego between 6:00 and 9:00 a.m. and in the NB direction between 3:00 and 6:30 p.m. The HOV facility is closed during off-peak hours and weekends.

The analysis revealed that accident rates on SD 15 in all time periods for corresponding directions are lower than on the LA 210 control section.

Further disaggregate analysis with respect to types of collisions and movements preceding collisions suggests that there are no systematic differences in the presence of the HOV facility. However, the analysis did reveal that accidents located on the left shoulder and beyond are proportionally higher on SD 15 than on LA 210. If time had permitted, it would have been interesting to investigate this issue further. However the number of accidents of this type is so small that any possible effect would be of minor consequence.

CONCLUSIONS AND RECOMMENDATIONS

An extensive study was performed of accident rates and accident characteristics on several California HOV facilities, in comparison with a group of selected control sections. The assembled evidence suggests that there are no inherent operational features of HOV facilities that per se appear to affect either the frequency or the nature of accidents to a significant degree. As reported in previous studies of this type, the influence of congestion on accident rates and locations clearly overwhelms any other operational factors.

It may be concluded that the addition of an HOV lane may lead to more accidents, to fewer accidents, or be neutral with respect to accidents depending on how the additional capacity fits within the overall urban network. This implies the need to examine the short-term and long-term congestion mitigation and migration aspects of proposed capacity enhancement projects and their staging, whether or not they include HOV facilities.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the sponsorship of Caltrans and, in particular, Michael Auslam, of the Division of Traffic Operations, Sacramento, for his assistance and support. The authors wish to acknowledge other staff members of Caltrans districts and CHP for their contributions.

The authors would also like to acknowledge James Daly, Alypios Chatziioanou, and other members of the research team for their efforts on this project.

REFERENCES

1. S. M. Betts, L. N. Jacobson, and T. D. Rickman. *HOV High Occupancy Vehicle Lanes.* Washington State Department of Transportation, District No. 1, Traffic Systems Management Center, 1983.
2. J. W. Billheimer, R. J. Bullemer, and C. Fratessa. *The Santa Monica Freeway Diamond Lanes, Volume I: Summary.* Systan Incorporated, Los Altos, California, 1977.
3. D. L. Christiansen and D. E. Morriss. *The Status and Effectiveness of the Houston Transitway System.* Report FHWA/TX-89/3-1146-2. Texas Transportation Institute, Texas A&M University, College Station; FHWA, U.S. Department of Transportation, 1989.
4. J. W. Billheimer, J. McNally, and R. Trexler. *TSM Project Violation Rates, Final Report.* California Department of Transportation, 1981.
5. Institute of Transportation Studies. *An Analysis of Traffic Safety Relative to the Commuter Lane Projects on SR-91 and SR-55 in Orange and Los Angeles Counties, Final Report.* University of California, Irvine, Oct. 1987.
6. T. F. Golob and W. W. Recker. *Safety Impacts Associated with Installation of HOV Lanes.* Institute of Transportation Studies, University of California, Irvine, 1988.
7. T. F. Golob, W. W. Recker, and D. W. Levine. *Safety of Freeway Median High Occupancy Vehicle Lanes: A Comparison of Aggregate and Disaggregate Analysis.* Institute of Transportation Studies, University of California, Irvine, 1989.
8. H. Hautzinger. Regression Analysis of Aggregate Accident Analysis Data: Some Methodological Considerations and Practical Experiences. *Journal of Accident Analysis and Prevention,* Vol. 18, No. 2, 1986, pp. 95–102.
9. Institute of Transportation Studies. *An Analysis of Traffic Safety Relative to the Commuter Lane Projects on SR-91 and SR-55 in Orange and Los Angeles Counties. Criticisms of the Study: A Reply.* University of California, Irvine, Feb. 1989.
10. L. P. Berriman. *A Critical Review and Analysis of the Institute of Transportation Studies' Report on Safety Aspects Of Car-Pool Lane Projects on SR-91 and SR-55.* Drivers for Highway Safety, Jan. 1988.
11. E. C. Sullivan, N. Devadoss, J. Daly, and A. Chatziioanou. *High Occupancy Vehicle Lane Safety,* Applied Research and Development Facilities and Activities (ARDFA), California Polytechnic State University, San Luis Obispo, Sept. 1992.

Publication of this paper sponsored by Committee on High-Occupancy Vehicle Systems.

Alternatives for Providing Priority to High-Occupancy Vehicles in the Suburban Arterial Environment

KERN L. JACOBSON, LARRY INGALLS, AND ETHAN H. MELONE

During the past 20 years, restricted lanes for high-occupancy vehicles (HOVs) have become a familiar feature of the freeway environment in many areas of the country. HOV lanes allow buses and carpools to bypass delay during congested peak periods, increasing the attractiveness of alternatives to travel by single-occupant vehicle. As continued suburban growth in major metropolitan areas limits mobility on arterial roadways, the need for a similar solution for suburban arterials is becoming increasingly apparent, but HOV facilities remain rare in the suburban arterial environment. Suburban arterials are complex in their function and design, making the simple application of the basic freeway HOV lane concept difficult. The alternatives for providing HOV priority in the arterial environment studied in Snohomish County, Washington, a suburban county in the Seattle metropolitan area, are discussed. All of the treatment options that have been used to provide priority to HOVs were considered. The advantages and disadvantages of treatments that show some potential for success are discussed. An important finding is that suburban arterial HOV treatments must be focused on reducing delay for HOVs at signalized intersections since congestion emanates from the signalized intersection in this environment.

Community Transit completed its Arterial System HOV Study in March 1993. The study examines the potential for high-occupancy vehicle (HOV) facilities, which allow buses, vanpools, and carpools to bypass congestion, to provide options for travel mobility in the suburban arterial environment of Snohomish County, Washington. The focus of this paper is on the first phase of the study, which involved the identification of a comprehensive set of alternatives for HOV priority. This identification of alternatives was followed by the development of an analysis methodology and the analysis of 100 mi of suburban arterial segments in key travel corridors. In the alternatives identification stage, the advantages and disadvantages of the alternatives were discussed in relation to the general characteristics of the Snohomish County arterial environment. "Alternatives" was broadly defined and included physical treatments and distinct options for their implementation and operation.

K. L. Jacobson, Parsons Brinckerhoff Quade & Douglas, Inc., 999 Third Avenue, Suite 801, Seattle, Wash. 98104. L. Ingalls, Community Transit, 133-164th Street, S.W., Suite 200, Lynnwood, Wash. 98037. E. H. Melone, Parsons Brinckerhoff Quade & Douglas, Inc., 999 Third Avenue, Suite 801, Seattle, Wash. 98104. Current affiliation: City of Seattle, Office of Management and Budget, 300 Municipal Building, Seattle, Wash. 98104.

KEY DISTINCTIONS BETWEEN THE ARTERIAL AND FREEWAY ENVIRONMENTS

The objective of arterial HOV treatments is essentially the same as that of freeway HOV treatments: to bypass congestion. Bypassing congestion provides a travel time advantage to HOVs, which is the key to achieving several basic goals commonly associated with HOV treatments, such as increasing transit ridership, increasing the efficiency and passenger-carrying capacity of a facility, and maintaining mobility in the face of severe congestion. There are, however, some clear distinctions between HOV considerations for the suburban arterial environment and those for the freeway environment. These distinctions include the sources of suburban arterial congestion, the geometric and operational characteristics of the arterial, and the function of arterial roadways.

On the suburban arterial, congestion emanates primarily from the signalized intersection, where queues develop and system delays occur. A primary focus in the development of alternatives for HOV priority treatments in an arterial environment must be the signalized intersection because bypassing congestion is the primary objective of HOV treatments.

The geometric and operational characteristics of arterial roadways are quite different from freeways: lane widths may be narrower, speed limits are lower, access restrictions are fewer, and the roadways are signalized to accommodate at-grade intersections. Arterial roadways are also complicated by the variety of activities they serve. Motor vehicle traffic on arterials must interact safely with bicycle and pedestrian movements. Through traffic on arterials competes for roadway space with traffic that is turning to the right or the left or entering the roadway from side streets or driveways. HOVs taking advantage of priority treatments face turning conflicts that may cause safety and operational problems. Bus stops and bus reentry into the traffic flow may cause additional conflicts. Enforcement activities are hampered by the complex movements of the arterial environment and the limited space available for enforcement vehicles. All these differences limit the applicability of HOV experience developed in the freeway environment.

Arterials also serve a different function than that of freeways: they generally provide access within local areas instead of linking more distant areas. Although trip distances on arterials tend to increase as freeways become increasingly congested, arterials continue to serve their local access function, even as they become long-haul alternatives. Therefore the different functions of the arterial roadway in the transportation system must also be recognized in the development of HOV treatments for the arterial environment.

VIABLE ALTERNATIVES

Some alternatives were found to have no potential to meet several basic criteria for effectiveness and were eliminated from further analysis. These macro-level screening criteria were financial viability, geometric feasibility, functional adequacy, and public acceptability. In this section, the alternatives that passed the fatal flaw screening are identified and described. Past experiences with these alternatives are noted, and advantages and disadvantages are discussed. Some general findings from the corridor-specific analysis of alternatives are discussed. Among the alternatives discussed are signal priority treatments, continuous lane treatments, design alternatives for priority at signalized intersections, and support measures and facilities.

Signal Priority Treatments

A major focus of the study was the analysis of the potential benefits of providing priority to buses or other HOVs at signalized intersections by altering the timing of traffic signals to favor such vehicles. The traffic control philosophy guiding signal priority treatments is that traffic signals should be operated to minimize total person delay. This is a natural evolution from current signal control strategies, which strive to minimize total vehicle delay.

Preferential treatment of an HOV at a traffic signal requires two functions: the identification of the vehicle, involving an automatic vehicle identification (AVI) technology; and the modification of the signal timing in response to that vehicle, an alternative signal control strategy (see Figure 1).

A variety of AVI technologies and alternative signal control strategies was analyzed in the study. AVI technologies include radio frequency transmission, microwave transmission, optical or infrared identification, and surface acoustical wave technology. The advantages and disadvantages of each of these technologies are presented in Table 1. Alternative signal control strategies include traditional preemption, traditional priority, specialized phasing, noncycle-based signal control systems (e.g., OPAC-RT with HOV Preemption), and noncycle-based systems with person optimization instead of vehicle optimization (e.g., HOV-Weighted OPAC-RT). The advantages and disadvantages of these signal control alternatives are presented in Table 2. A traditional priority strategy, simulated using TRAF-NETSIM, was shown to provide total delay savings of about one-third and to eliminate three of four stops at signals by buses.

Continuous Right-Side HOV Lanes

Continuous right-side HOV lanes reserve the outside lane for bus or bus and carpool use along a continuous section of an arterial. A slight variation on continuous right-side HOV lanes is the designation of the right lane as a "local access only" and "right-turn-only" lane for all vehicles except HOVs. All general-purpose vehicles in the lane would be required to turn right at a driveway or at the next available intersection. This

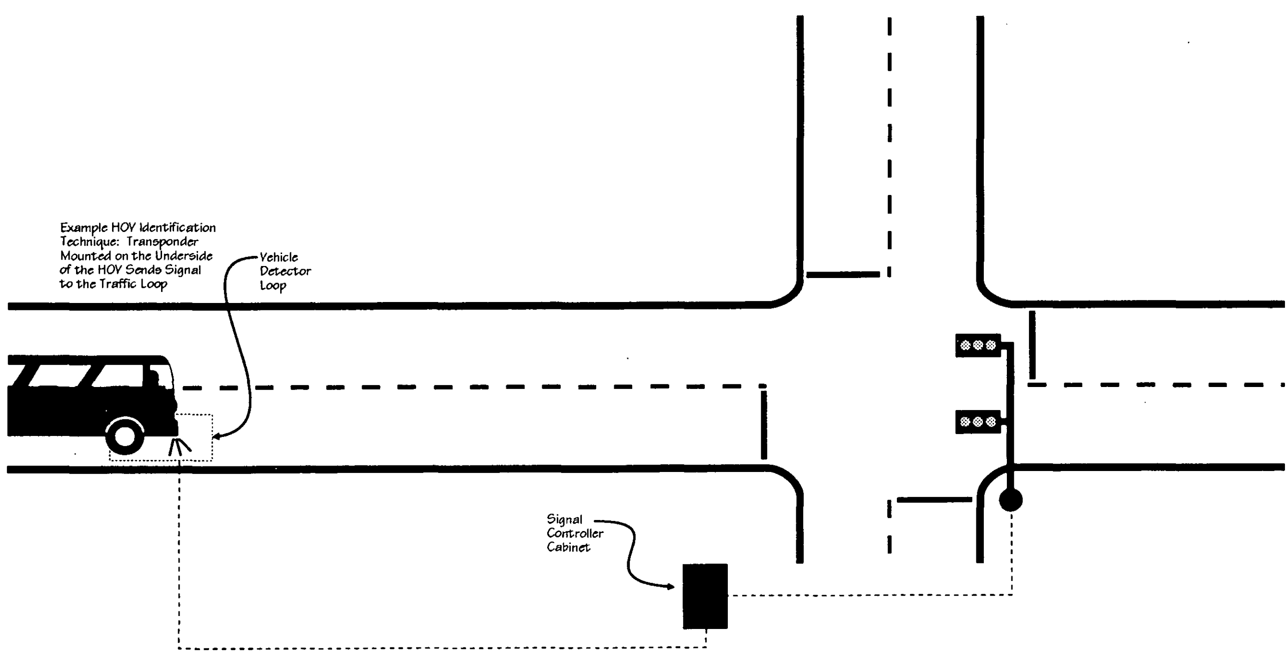

FIGURE 1 Signal priority alternatives.

TABLE 1 Automatic Vehicle Identification Systems

Technology	Configuration	Functions Available	Compatibility with Carpools	Advantages	Disadvantages
Radio Frequency Transmission (RF)	Tags and readers or other roadside or in-pavement antenna; compatible with loop detectors.	ID only; two way communication; voice; transmission of information	Compatible with the use of tags	The most applicable equipment available; compatible with simple tags and more sophisticated systems; used for two-way communication; compatible with roadside or in-pavement antenna.	The amount of data which can be transmitted with a loop configuration is limited.
Microwave	Tags and roadside readers; requires line-of-sight	Same as Radio Frequency Transmission	Compatible with the use of tags	Compatible with tags and two-way communication; transmission is at higher rates than RF.	Line-of-sight transmission, therefore signal can be screened by intervening vehicle; required power levels are high.
Optical/Infrared	Tags or bar-code tags; roadside readers; requires line-of-sight and good visibility.	ID only.	Compatible with the use of tags and bar codes	Compatible with tags/strict mounting requirements for tags and reader; can use bar codes.	Same as for microwave; requires good visibility; susceptible to dirt.
Surface Acoustical Waves (SAW)	Tags and roadside readers.	ID only.	Compatible with the use of tags	Same as for Optical except for use of bar codes.	Insufficient accuracy.

treatment may have essentially the same benefits as a continuous right-side HOV lane and also enhance access to local business. A possible striping and signing concept for this treatment is shown in Figure 2.

Past Experience

Right-side HOV lanes operate successfully in several areas of the country. Examples include the San Tomas and Montague Expressways in San Jose, California, and North Washington Street in Alexandria, Virginia. The San Tomas Expressway HOV lane is an 8-mi facility that operates during the morning and peak periods and is designated for use by buses and carpools of two or more people. The facility was opened in 1982. The Montague Expressway, a 7-mi facility with the same occupancy designation and hours of operation, opened in 1985. The North Washington Street HOV lane, which opened in 1984, is a 3-mi facility that operates during peak periods and is designated for use by buses and carpools of three or more people.

More common than right-side HOV lanes on suburban arterials are right-side bus-only lanes, which have been used for years in downtown areas. At least 95 such projects have been

TABLE 2 Alternative Signal Control Strategies

Strategy	Configuration	Function	AVI Technology	Advantages	Disadvantages
Traditional Premption	Local preemptor connected to controller; may be under system control.	Strict preemption.	Opticom; tag with roadside reader; loop detector with transponder on underside of vehicle.	Simple configuration; inexpensive.	No flexibility of control; possible safety problems with short intervals; disruption of general purpose traffic can be severe; legislative prohibition.
Traditional Priority	Requires traffic control system modification.	Flexible priority treatment.	All of the above.	Very flexible control options; simple concept.	Requires customized equipment.
Specialized Phasing	HOV lane at signal.	Provides priority to HOV lane.	Standard loop detection.	High service level to HOVs.	Directly impacts general traffic movements; requires HOV lane.
OPAC-RT with HOV Preemption	OPAC coordinator unit on standard controller with advanced detection (25 seconds) implements OPAC.	Strict preemption with facilitated recovery.	Same as traditional preemption.	OPAC provides control efficiency to minimize negative preemption impacts.	New technology; disadvantages of preemption.
HOV-Weighted OPAC-RT	Same as above.	Minimizes person delay and stops/maximizes throughput.	Same as traditional preemption.	Maximizes people movement efficiency.	New technology; disadvantages of preemption.

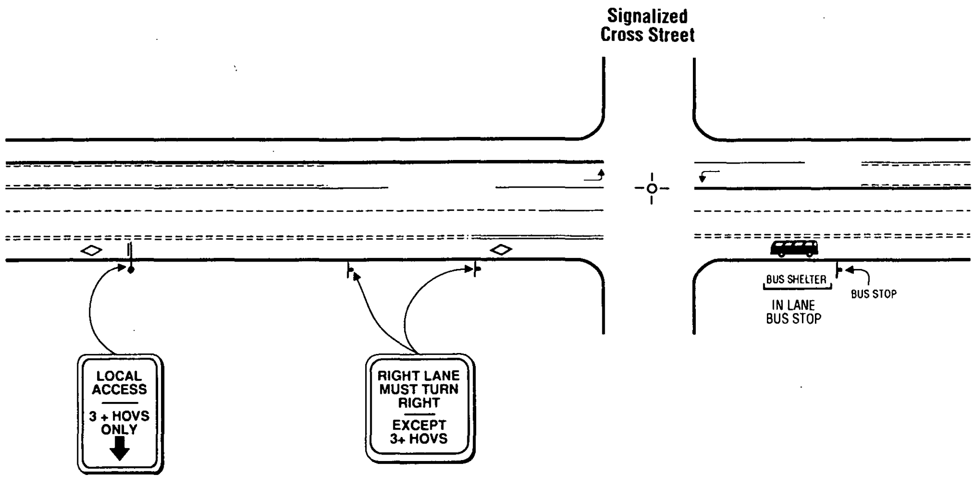

FIGURE 2 Typical striping and signing concept for 3+ HOV emphasis lane.

implemented since 1956, with varying degrees of success. Case-specific factors have produced widely varying impacts on person throughput, transit use, service reliability, and safety. A significant number of these lanes have been suspended because of low use, safety concerns, or enforcement problems, but the majority continue to operate, and data available for many of the operational lanes indicate that they are successful.

In the Puget Sound region, two arterial roadways currently contain right-side HOV lanes. A continuous right-side HOV lane on SR 99 northbound begins at NE 115th Street and ends 1.5 mi to the north at NE 145th Street. The SR 99 HOV lane has a 3+ HOV designation and operates with restrictions 24 hr/day.

On SR 522, a 3.3-mi southbound bus-only lane and a 0.9-mi northbound bus-only lane operate during the (directional) peak hours. During the off-peak periods, the lane reverts to shoulder use in each case.

Advantages

This treatment may be implemented at low cost when parking lanes or shoulders are available for conversion. Right-side HOV lanes provide good access to bus stops. In some cases, it may be the only geometrically feasible option for a continuous HOV treatment.

Disadvantages

Depending on their destinations, some potential users of right-side HOV lanes may need to make a left turn as they leave the HOV corridor. A right-side HOV lane requires them to weave into slower moving traffic to reach a lane from which they can turn left. This weaving movement may cause safety problems, reduce speeds, and reduce the perceived advantage of the HOV lane.

A safety review of HOV lanes in the Puget Sound region of Washington revealed the primary operational difficulties of right-side HOV lanes: HOV traffic comes into conflict with right-turn movements, pedestrian crossing movements, and weaving movements of vehicles entering and exiting driveways (Senn 1990, unpublished data).

In-lane bus stops impede HOV travel in right-side HOV lanes, causing safety problems and reducing travel times for carpools and vanpools forced to wait behind stopped buses. Additional right-of-way would be required to resolve this problem by providing bus turnouts.

Past experience with right-side arterial HOV lanes has shown that they are perceived as short-haul facilities, which makes the lanes less attractive to commuters and limits usage.

Continuous Left-Side HOV Lanes

Continuous left-side HOV lanes may be provided in the middle of two-way, multilane arterial streets where sufficient right-of-way exists. Inside or median HOV lanes would typically operate as concurrent flow lanes.

Past Experience

Median-strip reserved bus lanes have most commonly been used in U.S. cities where street cars previously had been operated in the center median. For example, a two-way busway operates in the median of Canal Street in downtown New Orleans, Louisiana.

Advantages

Operating an HOV facility in the center of an arterial eliminates the various traffic conflicts in the curb lanes, such as conflicts with ingress and egress from driveways and right-turn movements. These advantages lead to higher speeds and reduced stops and delays for HOVs. Left-side HOV lanes are also perceived as being safer than right-side lanes. The left-side treatment may also be perceived as a long-haul treatment because of the reduced conflicts with movements required for local access.

Disadvantages

Inside HOV lanes require bus stops in the center of the roadway. This placement of bus stops forces passengers to cross busy streets to board and exit from buses, increasing conflicts between pedestrians and vehicles. Left-turn movements replace right-turn movements as a potential operational problem. This problem requires restriction of left turns in some cases and special signal phasing in others.

Reversible HOV Lanes with Signal Control

Reversible HOV lanes provide a lane for HOV travel in the peak direction of travel, reversing direction of operation as the peak-travel direction reverses, typically from the morning home-to-work direction to the evening work-to-home direction. Among the various design options, control with signals is a low-capital option that may be suitable for the arterial environment. Lane control consists of overhead signals and variable-message signs. Cones or other manually moveable barriers also may be used in conjunction with a reversible HOV lane.

Past Experience

Many examples of reversible arterials exist, including in the cities of Houston, Memphis, St. Louis, and Charlotte. However no known examples of reversible HOV lanes on arterials exist.

Advantages

Reversible lanes are less expensive to construct than two-way treatments and require less right-of-way. They provide added capacity in both directions, allowing use of lanes that might otherwise be underused in the off-peak direction.

Lane control with signals is the reversible treatment that is probably most suited to the geometric context of the suburban arterial. It allows operation in the peak direction without requiring barriers, which restrict mobility and use roadway space. Lane control with signals may be implemented and operated at a lower cost than more capital-intensive reversible lane options.

Disadvantages

Without careful design, traffic control for reversible lanes may be confusing and dangerous because of the possibility of wrong-way movement. The lanes may be perceived as dangerous even if adequate safety measures are provided. To be effective, the treatment requires a strong directional split, which is not common in most arterial settings. Although the treatment is not capital intensive, it may be labor intensive if it requires daily cone movements.

Signal Queue Jump

A spot application of HOV priority treatment involves allowing access to right-turn only drop lanes at traffic signals for through movements by buses, carpools, or both. An additional merge lane downstream of the intersection is provided to allow HOVs using the right-turn drop lane for queue bypass to re-enter the through traffic flow. A special phase for the HOV movement may be provided with this treatment. See Figure 3.

Past Experience

A 1,000-ft curbside HOV lane has been implemented on NE Pacific Street in Seattle's University District, with signal priority at the intersection of NE Pacific Street and Montlake Boulevard. A signal queue jump currently operates at NE 4th Street in downtown Bellevue, Washington.

Advantages

This treatment reduces HOV delays and improves transit reliability. The costs of the treatment are relatively low, requiring widening for a distance downstream of the intersection. A standard for the length of this widening is 12 ft/sec of average green time. The length of widening upstream of the intersection is a function of the maximum normal length of the queue of general-purpose vehicles waiting at the signal.

Disadvantages

Additional merging is required downstream of the intersection, which would increase traffic conflicts and potentially produce a higher accident rate. These impacts could be mitigated by providing an HOV "early release" control strategy at the signal, but the early release would in turn increase delay and reduce capacity for general-purpose traffic. Therefore, such an early release strategy is most appropriate at nonbottleneck intersections. Conditions would have to include relatively minor queuing for peak-hour right turns, significant queuing upstream of the intersection, and freedom from bottlenecks for a considerable distance downstream of the intersection to allow easy re-entry into the through traffic flow. If any of these conditions were to change significantly after implementation of the treatment, many of the benefits of the treatment would be lost. The treatment would be geometrically feasible only where existing right-of-way allows for the downstream merge lane.

Special Access

Special access for HOVs, such as HOV-only freeway ramps or HOV-only connecting streets, provide a time-saving shortcut to HOVs at key points in their travel routes. Such treatments could be used as links to other HOV treatments, such

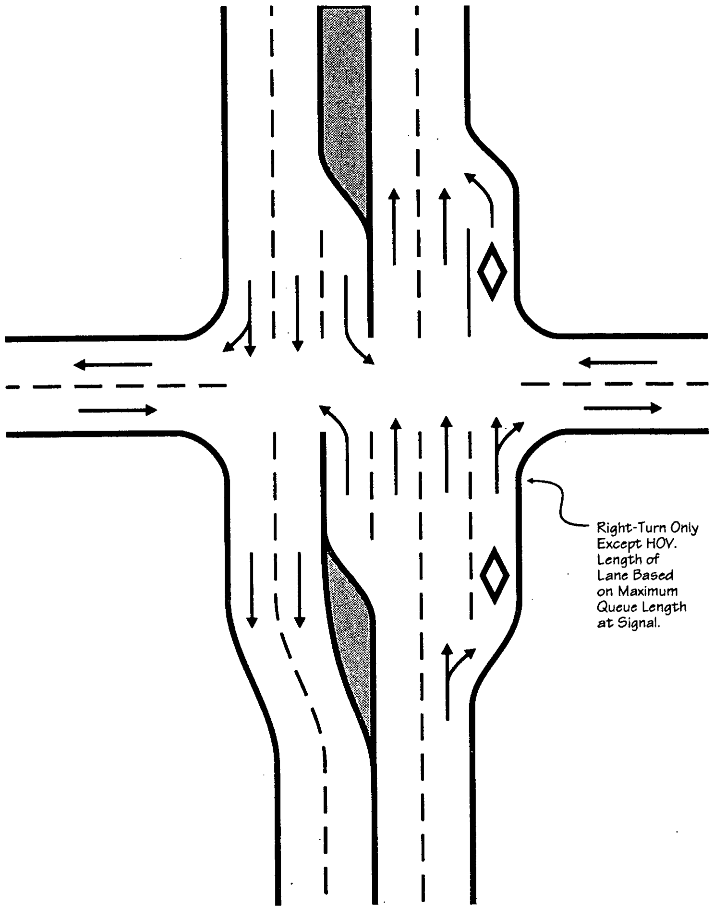

FIGURE 3 Signal queue jump.

as HOV lanes, ramp bypasses, or park-and-ride lots. In some cases they may be effective as stand-alone treatments.

Advantages

Special-access facilities could improve the connectivity of the arterial system to existing HOV facilities and could have a positive impact on mode shift by making HOV modes more desirable at key locations.

Disadvantages

Special-access facilities do not have any general disadvantages. In specific cases the facility that would be required to provide special access may be excessively costly.

CONCLUSION

Of the various arterial HOV alternatives that have been implemented in the United States and abroad, all have advantages and disadvantages that require evaluation of the particular application. The characteristics of the arterial environment are diverse and complex enough to preclude general conclusions regarding any particular alternative. Signal priority treatments that use advanced technologies to minimize person delay at intersections appear to have the most universal potential to achieve the goal of bypassing congestion without unacceptable impacts to general-purpose traffic, but such treatments are nevertheless limited by intersection capacity constraints.

Publication of this paper sponsored by Committee on High-Occupancy Vehicle Systems.

Evaluation of Seattle's South I-5 Interim High-Occupancy Vehicle Lanes

GARY FARNSWORTH AND CYRUS G. ULBERG

The Washington State Transportation Center is developing a high-occupancy vehicle (HOV) monitoring and evaluation program for highways in the Seattle area. This paper represents a part of the transportation center's work in developing the program by evaluating one particular HOV facility on Interstate 5 in South King County. The study is both an evaluation of this facility's operation and a means of testing various measures that have been established and used in other research endeavors. Each of the following measures provided helpful information for evaluation of the facility: HOV lane volumes, congestion impacts, safety impacts, travel-time savings and reliability, average vehicle occupancy, violation rates, and public perception. The findings of this study give reason to believe that this facility has been relatively ineffective as a short-term application. Geometric improvements should be made in the facility to alleviate congestion and safety impacts caused by the HOV lanes. Following geometric improvements, the facility's occupancy requirement should be reduced from three to two persons per vehicle.

High-occupancy vehicle (HOV) lanes have become an integral part of regional transportation planning throughout the United States. Particularly in a city like Seattle, Washington, where there are no rapid transit and commuter rail lines, HOV lanes can play a valuable role in efforts to manage ever-increasing congestion. With new emphasis on transportation systems management (TSM), officials at the Washington State Department of Transportation (WSDOT) are interested in gaining a better understanding how effective its HOV facilities are. To gain this understanding, WSDOT has commissioned the Washington State Transportation Center (TRAC) to study the state's existing HOV facilities and to develop an HOV monitoring and evaluation program for facilities in the Puget Sound Region.

By evaluating the operation of each HOV facility against standard effectiveness measures, the research team can assess any apparent deficiencies and recommend improvements for more effective facility operation. Also, by using and testing the standard effectiveness measures, the research team can use the findings from each facility to help develop the long-term monitoring and evaluation program. Research findings from one HOV facility, on Interstate 5 in south King County, Washington, are discussed in this paper.

G. Farnsworth, Washington State Department of Transportation, Traffic Office, Traffic Building Room 1C11, P.O. Box 47344, Olympia, Wash. 98504-7344. C. G. Ulberg, Graduate School of Public Affairs, Washington State Transportation Center, University of Washington, JE-10, 4507 University Way N.E., #204, Seattle, Wash. 98105.

BACKGROUND

The state of Washington regards priority treatment of HOVs as an effective means of maximizing highway travel capacity. WSDOT officials believe that HOV treatments can support efforts to reduce traffic congestion, dependency on fossil fuels, and automobile pollution. WSDOT takes an active role in promoting and coordinating the development of HOV facilities, transportation demand activities, and related TSM programs. This general policy approach to HOV facilities is intended to provide incentives for people to shift from single-occupant vehicles to ridesharing modes. Such efforts support the state's ultimate goal of protecting the environment while maintaining or improving each person's ability to travel (*1*).

The facility studied in this research is referred to as the South Corridor, and it consists of a pair of lanes on the Interstate 5 freeway in the southern part of King County (see Figures 1 and 2). There is one left lane in each direction, and each lane runs approximately 6.5 km (4 mi), adjacent to the freeway general-purpose lanes. Only vehicles carrying three or more people (3+ occupancy definition) are permitted on the HOV lanes. The South Corridor represents Phase I of the ultimate HOV facility, which will extend from downtown Seattle to south of Tacoma. This facility is part of the regionwide HOV system recommended by Washington's Interagency HOV Task Force.

Although originally planned for opening to traffic in the late 1990s, the South Corridor facility was opened in the summer of 1991. The accelerated construction of this facility was the WSDOT's short-term response to requests from the State Transportation Commission. The commission was acting on a request from Pierce County's Transit Board of Directors to install an HOV facility somewhere along I-5 between Federal Way and I-405. (Pierce County is south of King County, where I-5 continues to Tacoma). Pierce County's request resulted from many factors, including efforts to improve bus service from Pierce County to Seattle and reaction to a petition from one activist group demanding HOV lanes on I-5 (*2*).

METHOD

The overall method of this study includes the collection, analysis, and comparison of all data from within and across the South Corridor. The data were collected by field observations, detector systems, and questionnaire surveys. The Washington State Highway Patrol, WSDOT, the Municipality of Metropolitan Seattle (Metro), and Pierce Transit also provided relevant information.

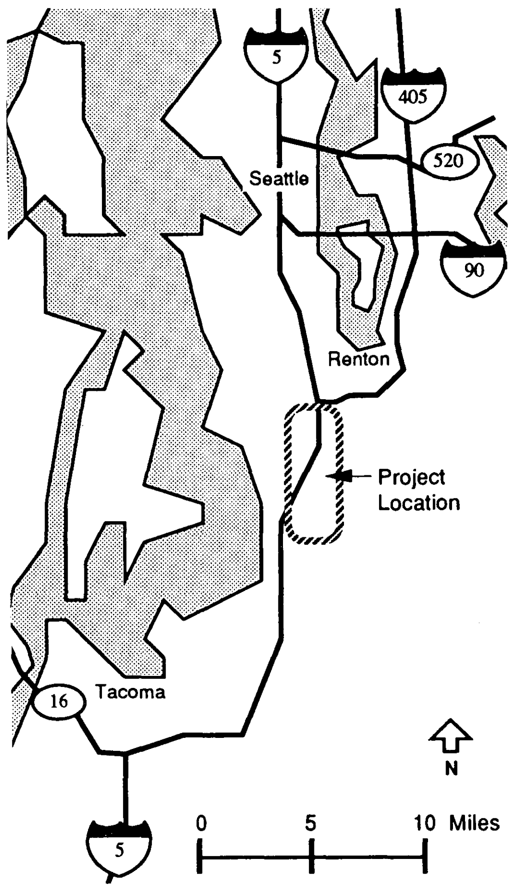

FIGURE 1 Location plan.

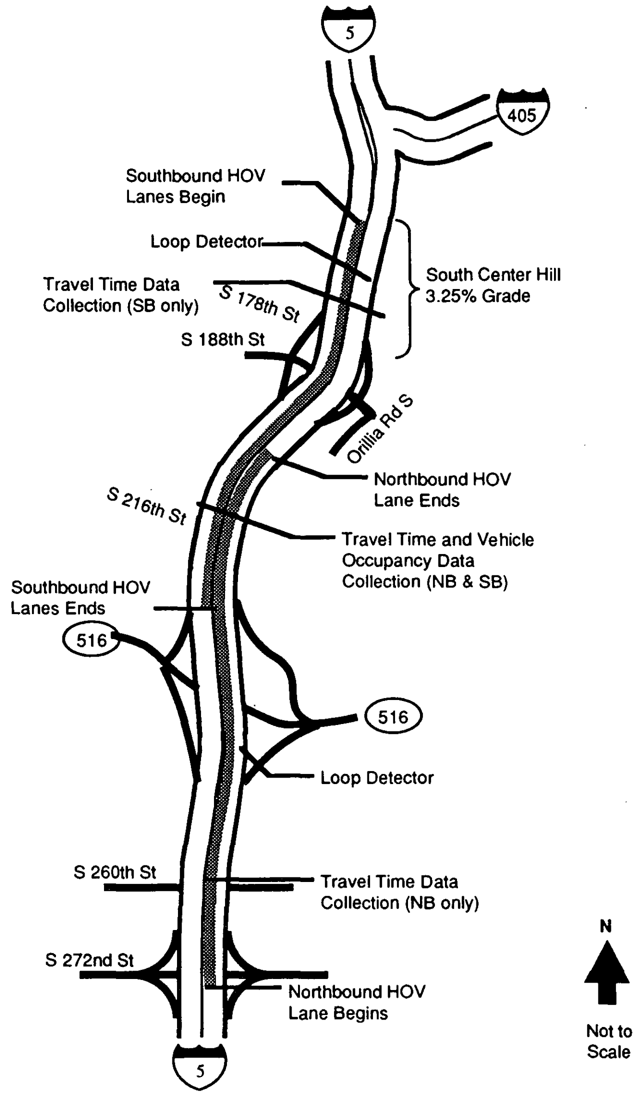

FIGURE 2 South Corridor plan.

Effectiveness Measures

The analysis included an application of the various data to a group of effectiveness measures. The effectiveness measures that the research team is testing for the evaluation and monitoring program have been developed by other research efforts and generally accepted as appropriate for evaluating HOV facilities. Accompanied by supporting objective statements, the following is a list of those measures, essentially derived from work by Turnbull (*3*), that were specifically chosen for the South Corridor evaluation:

- HOV lane volumes: the number of HOVs using the HOV lanes. The HOV facility should increase the per lane efficiency of the total freeway corridor. Of equal importance, the facility should maintain public support through a perception of adequate use.
- Congestion impacts: the impact of HOV lanes on the overall highway section. The HOV facility should not unduly impact the operation of the freeway mixed-flow lanes.
- Safety impacts: the effects of HOV lanes on highway safety. The HOV facility should be safe and should not unduly impact the safety of the mixed-flow lanes.
- Travel-time savings and reliability: the advantages provided by the HOV lanes. The HOV facility should provide travel-time savings and a more reliable trip time to HOVs using the facility. The facility should also increase the operating efficiency of bus service in the freeway corridor.
- Average vehicle occupancy (AVO): the average number of occupants per vehicle in all lanes of each highway section. The HOV facility should improve the capability of a congested freeway corridor to move more people by increasing the number of persons per vehicle.
- HOV enforcement: the number of vehicles violating the 3+ restriction. An HOV facility relies on enforcement to maintain the integrity of HOV travel benefits. Enforcement should also assist in maintaining public support by a perception of proper use.
- Public perception: public opinions, including traveler support of the HOV facility. It is imperative that the HOV facility have public support.

Each of these measures may also be applied to the following objectives provided by the *Washington State Freeway HOV System Policy* (*1*):

- Improve the capability of congested freeway corridors to move more people by increasing the number of persons per vehicle;

- Provide travel-time savings and a more reliable trip time to HOVs that use the facilities; and
- Provide safe travel options for HOVs without unduly affecting the safety of the freeway general-purpose mainlines.

Traffic Data

The following types of traffic data, which represent the peak periods of 6:00 to 9:00 a.m. (northbound) and 3:00 to 6:00 p.m. (southbound), were collected from the South Corridor:

- Lane volumes: number of vehicles traveling in each lane, applied to the measures of HOV lane volumes and congestion impacts;
- Accidents: number recorded by State Patrol before and after HOV lanes opened, applied to the measures of safety impacts and congestion impacts;
- Travel times: time for vehicles traveling in the corridor (HOV lanes versus regular lanes), applied to the measure of travel-time savings and reliability; and
- Persons per vehicle: number for each lane in the highway section, applied to the measures of AVO and HOV enforcement.

Lane volume data were collected both by roadway induction loop detectors and by visual field counts. Also, data on travel time and persons per vehicle were collected by traffic observers recording visual observations from field sites. For travel times, observers recorded (into laptop computers) the license plate numbers of vehicles in the appropriate lane at two sites within the corridor. The time difference and distance between the two recordings of each license plate were used to calculate travel times. For persons per vehicle, observers monitored and recorded the type of vehicle (or number of persons in each vehicle) into a laptop computer.

Because many unique traffic conditions or causes cannot be understood by traffic data analysis alone, many important discoveries about vehicle interaction were made in this study from the visual observations during data collection. The major limitation of this project, as with most HOV studies, is the lack of available data (AVO data are the exception), such as travel times or lane volumes, for any time period before the HOV lanes were opened to traffic in the summer of 1991.

Traveler Opinion Data

The research also includes responses to questionnaire surveys, distributed to travelers observed in the South Corridor. The travelers were distinguished by those who were observed using the HOV lanes versus those who were observed using the mixed-flow lanes. The questionnaire asked travelers for perceptions and opinions about HOV lane effects on safety, effects on travelers using both HOV and mixed-flow lanes, ways to improve each respective corridor, and HOV lane enforcement. In addition, the survey also asked for general demographic (personal) information, including typical mode choice for commuting, level of education, household size, and number of vehicles owned. The survey data were applied to the measure of public perception.

RESULTS AND DISCUSSION

HOV Lane Volumes

In general, analysis of the peak-period volume data indicates that the HOV lanes in the South Corridor were underused, in comparison with minimum volumes that are typically considered necessary for HOV lane effectiveness (see Table 1). A common criterion for minimum HOV lane volumes is based on developing and maintaining public perception that the facility is being adequately used. Fuhs provides the following insights on this issue: (*4*) "The perception of non-HOV drivers, and often the general public, is that the facility is not adequately used when they do not see vehicles in it. This may create what practitioners commonly refer to as "empty lane syndrome." This perception, whether correct or not, may create pressure to roll back occupancy restrictions . . . or terminate the project altogether. Such perceptions have been the single most critical issue in upholding the long-term viability of project operations in some locales . . . ."

Analysis of the HOV volume data also indicates a 33 percent drop in the volume of vehicles traveling in each HOV lane, within about 4 to 5 km (2 to 3 mi) downstream. It is unclear why there was a drop in HOV lane volumes. However one factor in the southbound direction could be that the southbound lane began at the base of the Southcenter Hill, which has a long, steep grade (see Figure 2). Buses entering the southbound HOV lanes would have to climb slowly because of the steep grade of the hill. To pass buses, drivers of HOV automobiles would often leave the HOV lane and travel in the general-purpose lanes, but only if traffic in the general-purpose lanes was traveling faster than the buses.

Perhaps more important, it was also discovered that carpools and buses tended to merge out of the HOV lanes well

TABLE 1 HOV Lane Peak-Hour Volumes

Direction/Data Collection Site	Peak Period Volumes (Vehicles per hour)
Northbound at Highway 516 (detector)	423
Northbound at S. 216th St. (manual)	257
Northbound Mean (combined data)	335
Southbound at S. 170th St. (detector)	404
Southbound at S. 216th/178th St. (man.)	261
Southbound Mean (combined data)	326
Literature Source	**Recommended Minimum (Vehicles per hour)**
Cal. Air Resources Board (5)	800
Nuworsoo & May (6)	1000
WSDOT (2)	200*
Parsons Brinckerhoff, et al. (7)	700**
Cechini (8)	800
Fuhs (4)	400

*For Seattle, otherwise 500
**For Orange/Riverside Counties, Cal.

in advance of the downstream lane endings to avoid bottlenecks created at the point where many other HOVs had to merge back into the regular traffic lanes at the end of the HOV lanes. There are also several exit ramps within the corridor, and some of the HOVs appeared to merge out of the HOV lanes to reach the appropriate exit ramp.

Congestion and Safety Impacts

The merge point bottleneck problem was particularly noticeable in the southbound direction at the south end of the HOV lanes. One symptom of this was an increase in accidents in the left-side general-purpose lanes since the installation of the HOV lanes (see Figure 3). The average number of accidents in the left lanes of the facility increased from 2 per month to 7 per month after the HOV lanes were installed. This increase in accidents is important for further investigation into safety and congestion impacts of the HOV lanes. Several survey respondents complained about added congestion near the downstream ends of the HOV lanes.

Two survey questions addressed concerns about the width of the HOV lanes and that those lanes use a part of the original highway shoulder. However the survey responses indicate that there was only marginal concern among travelers about such safety problems.

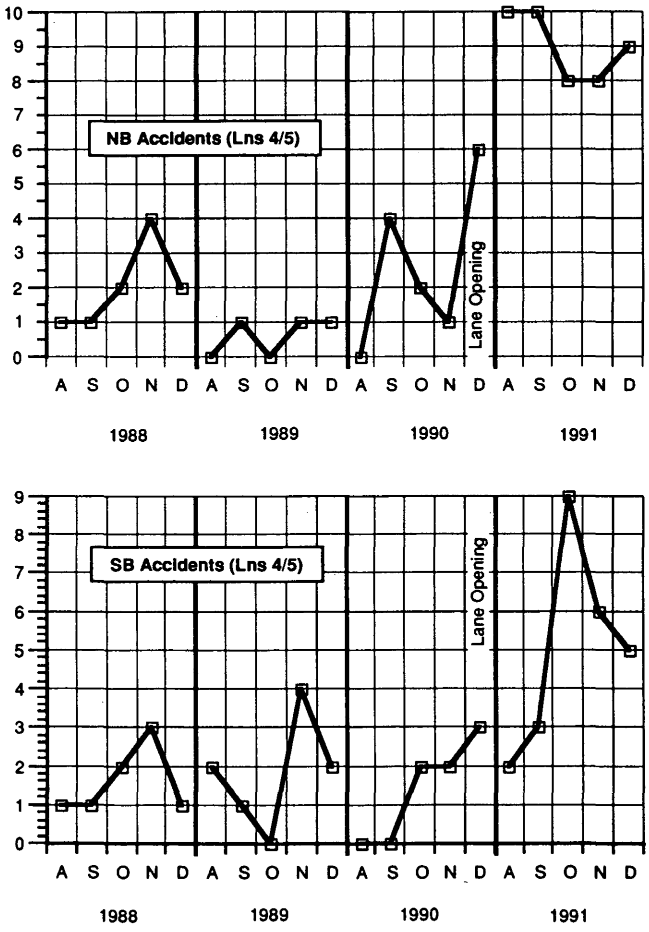

FIGURE 3 Monthly accidents, August–December, 1988–1991.

Because these HOV lanes were apparently contributing to increased highway congestion and accidents, they were not only ineffective in attracting HOVs, but they were also working against goals of increased mobility and welfare for all travelers (*9*).

Travel-Time Savings and Reliability

The travel-time comparisons, presented in Table 2, indicate that there was a limited amount of travel-time savings provided to travelers who used the HOV lanes, in relation to a 1-min/mi level of travel-time savings that is typically considered necessary for HOV lane effectiveness (*6,8*). This is a key point addressed by a majority of the literature searched for this report. Travel-time savings are important in encouraging travelers to carpool. If the HOV lanes on a roadway do not provide such an advantage, the incentive for current carpoolers to use the lanes or for other travelers to begin carpooling is limited.

Through informal discussions with the researchers, managing representatives from Metro and Pierce Transit have expressed satisfaction with bus route operations on the South Corridor. They cited time savings and reliability as the assets provided by the HOV lanes. They have adjusted route schedules for shorter trip times because the lanes allow for faster and more reliable service. However, through a comparison of the travel-time data with the bus schedule changes, it is difficult to verify that the transit schedule changes actually reflect travel-time savings that may have been provided by the HOV lanes.

The travel-time analysis also indicates that, with the exception of the Friday afternoon peak, there is little advantage of day-to-day travel-time reliability (or consistency) provided by use of the HOV lanes (see Table 2). Although travel-time reliability is considered a valuable measure of HOV facilities, and even if higher levels of reliability were reflected in the data, the findings of this research do not provide a quantitative level for which travel-time reliability may be considered an effective incentive for mode shift. (This indicates the need for further study.)

Average Vehicle Occupancy

Before and after comparisons of persons per vehicle indicate there was only a marginal increase of AVO in the South

TABLE 2 Travel-Time Summaries

Overall am & pm						
Direction	Lane	Day	Time of Day	Time Mean	St.Dev.	mph Mean
PM/SB	HOV (n=151)	Mon — Fri	3:00-5:15	3:42	0:47	46
PM/SB	SOV(n=1186)	Mon — Fri	3:00-5:15	3:58	1:18	43
	(t = -2.41, 2-tail p = 0.16)		Difference	0:16	0:31	-3
AM/NB	SOV (n=28)	Wednesday	7:30-8:15	3:17	1:05	45
AM/NB	HOV(n=204)	Wednesday	7:30-8:15	3:32	1:08	43
	(t = -1.07, 2-tail p = 0.29)		Difference	0:15	0:03	-2

Corridor since the HOV lanes were installed (see Table 3). A minimal change (1 to 2 percent) of AVO in the South Corridor implies that the HOV lanes were relatively ineffective in providing an incentive for an increasing number of travelers to carpool. "A related question is how much the existence of an HOV lane will trigger and sustain mode shift in favor of carpool formation and/or transit ridership" (*6*). The actual mode shift impact caused by HOV lanes may further be diminished by the fact that new policies for HOV incentives have been implemented for the Seattle area within the past few years.

The vehicle occupancy analysis shows that carpools of 3 or more persons increased by more than 100 vehicles per hour in the morning peak period. However the increase in carpools is only about 50 vehicles per hour in the afternoon peak period. The analysis also indicates that only 50 to 60 percent of the vehicles that qualified to use the South Corridor HOV lanes were actually found to be traveling in those lanes (see Table 3). Therefore the analysis indicates that some new carpools were attracted to the facility (either by change in mode or change is route), but the actual attractiveness of the HOV lanes is questionable because of the low percentage of carpools that were actually found in those lanes.

Note that in Table 3 each transit bus occupancy was counted as 40. This is an average consistent with transit buses throughout the region. Also note that at the time of this study, Metro and Pierce Transit reported essentially no increase in ridership attributed to opening of the HOV lanes. Also, the state's HOV system policy indicates that HOV lanes are appropriate improvements when "evidence exists that during peak hours of operation, the HOV lane will move more people than the per lane average of the adjacent general purpose lanes" (*1*). The vehicle occupancy and lane (vehicle) volume data for this corridor show that the HOV lanes were each moving approximately 1,900 persons per hour during the peak, versus an average of 2,500 persons per hour in the adjacent lanes.

HOV Enforcement

Further vehicle occupancy analysis indicates that the HOV lane violation rate is 26 percent in the South Corridor, which is a high rate in comparison with rates typically considered acceptable for an effective HOV facility. "A recommended goal in enforcing HOV lane restrictions is to keep violations at or below 10 percent" (*6*). Enforcement is quite difficult in this corridor because the highway shoulders are narrow and there is little access for police to observe and ticket violators. It is also arguable that the high violation rates are another symptom of the general ineffectiveness of this facility. The combination of underuse and a high violation rate could easily stimulate a negative public attitude toward HOV facilities, which clearly presents a concern in the South Corridor. Responses from the traveler surveys demonstrate much concern about the high number of violators seen traveling in the HOV lanes.

Traveler Perception of HOV Lanes

Several of the survey questions addressed traveler support for HOV lanes. A summary of the responses is shown in Figure 4. The results indicate that even most of the respondents who said they normally drive alone are supportive of HOV lanes.

The survey also asked for respondents to choose among alternative facility improvements, which would make the HOV lanes more attractive for carpooling and bus riding in the corridor. The most popular choice was that the HOV lanes should be extended farther in each direction. Eighty-four per-

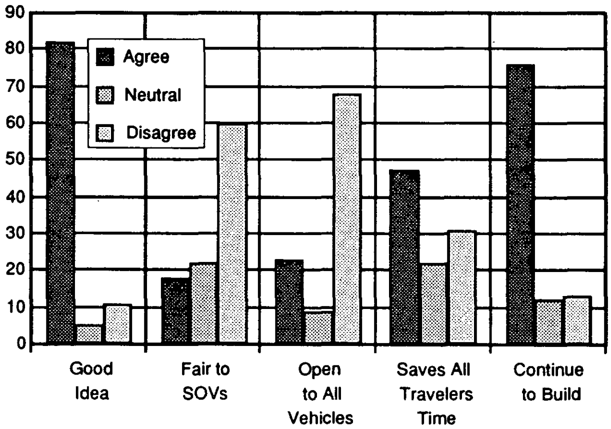

FIGURE 4 General traveler support for HOV lanes.

TABLE 3 Vehicle Occupancy Summary

South Corridor				Averages for each category over all lanes												%	%3+	%3+	%2+
Dir	Year	Start	End	1	2	3	4	Van	Tbus	Bus	Strk	Btrk	MC	Sum	AVO	Viol	in HOV	Ovral	Ovral
AM	92	7:00	9:00	6424	780	156	40	8	28	8	116	248	12	7820	**1.32**	**25**	**49**	**4**	14
				82%	10%	**2%**	**1%**	.10%	.36%	.10%	1%	3%	0%	100%					
AM	90	6:00	9:00	6202	746	66	18	8	26	2	104	272	38	7482	**1.29**	NA	NA	**2**	12
				83%	10%	**1%**	**0%**	.11%	.35%	.03%	1%	4%	1%	100%					
PM	92	3:00	5:00	4416	871	144	58	18	16	4	56	156	8	5747	**1.38**	**27**	**59**	4	19
				77%	15%	**3%**	1%	.31%	.28%	.07%	1%	3%	0%	100%					
PM	90	3:00	6:00	5368	1071	136	50	14	18	6	70	190	16	6939	**1.35**	NA	NA	4	19
				77%	15%	**2%**	1%	.20%	.26%	.09%	1%	3%	0%	100%					

Ln — Lane Number
1 — Single Occupant Vehicle(counted as 1 for AVO)
2 — 2 Occupant Vehicle(counted as 2 for AVO)
3 — 3 Occupant Vehicle(counted as 3 for AVO)
4 — 4 Occupant Vehicle(counted as 4.1 for AVO)
Van — Designated Vanpool(Counted as 10 for AVO)
Tbus — Transit Bus(counted as 40 for AVO)
Bus — all other buses
Strk — Double Axle Truck
Btrk — Triple Axle Truck
MC — Motorcycle

cent of the respondents chose this alternative. The results are consistent with another survey question that asks for agreement on whether or not more people would carpool if the lanes were extended farther (43 percent agreed, 21 percent disagreed). The relevance of this traveler support is also consistent with a general approach to HOV lane applications around the country. "The continuity of HOV lanes along a given corridor and/or connecting with other corridors is a significant factor contributing to the overall effectiveness of a high-occupancy vehicle lane system" (*8*).

Left Versus Right Side of Freeway

The survey listed another alternative improvement: moving the HOV lanes from the left to the right side of the highway in each direction (see Figure 2). The responses indicate low support for such a change. This lack of support tends to demonstrate that travelers perceive the use of HOV lanes on the right side as a hindrance to traffic, particularly when there are a high number of on- and off-ramps within each corridor (see Figure 2). This lack of support is also an indication that the HOV lanes would be even less attractive to travelers if they were located on the right side if each direction of the South Corridor.

Park-and-Ride Lots

Another alternative is for more park-and-ride lots to be located in the vicinity of the South Corridor. The responses show lack of support for such an improvement, which indicates that added park-and-ride lots would not particularly improve existing attractiveness of the HOV lanes. The research team found that park-and-ride lots were in good strategic locations for support of the HOV facility.

Occupancy Definition

The survey responses show an extremely strong preference (more than 75 percent agree) for an occupancy definition of two or more people per vehicle versus three or more. There are probably two main reasons for this response. First, it is much easier for travelers to form two-person carpools than three-person carpools. Second, travelers understand that if the definition is changed to two or more people per vehicle, it is likely that more vehicles will use the HOV lanes. In general, if travelers see more cars in the HOV lanes, then they perceive that the lanes are better used.

Because there is such strong support for a 2+ occupancy definition for the HOV lanes, it is reasonable to believe that the existing 3+ requirement is the cause for underuse in the South Corridor. However, because there is already new congestion on the highway at the merging ends of the HOV lanes, a change to the 2+ definition would likely create too much of an increase in HOV lane volumes, and an even larger bottleneck could occur at the merge locations. Therefore geometrics at the lane terminations must be addressed before a definition change would be feasible. It is also arguable that a change to the 2+ occupancy requirement as a short-term solution may work against the long-term goals of the state (*1,3*).

CONCLUSIONS AND RECOMMENDATIONS

Conclusions

Installation of the South Corridor HOV lanes was accelerated to provide a short-term solution to highway congestion-related problems in southern King County. However the findings of this research indicate that the HOV lanes were not an effective short-term application. An apparent lack of travel-time savings and low traveler use of HOV lanes were the most noticeable deficiencies. The merge alignment at the ends of each HOV lane created additional highway congestion, which also magnifies the facility's short length and contributes to the lack of traveler incentive to use the lanes.

By the application of various traffic and public opinion data to the measures established for this study, the South Corridor showed deficiencies in several other areas: significant congestion and safety impacts to the mixed-flow traffic, a low level of improvement in the peak-period person movement on the basis of vehicle occupancy counts, and a relatively high HOV lane violation rate. The vehicle occupancy data also indicated that, of the vehicles traveling in the corridor, only about half of the 3+ qualifiers were found to be using the HOV lanes.

Traveler responses to applicable survey questions indicated that nearly all travelers, independent of mode, were supportive of HOV facilities. However the survey responses, traffic data, and field observations also indicate a need for improvement of the facility. This point is particularly important, because effective performance and healthy public perception are critical to long-term acceptance and development of HOV facilities.

A positive aspect of the research is that representatives from Metro and Pierce Transit expressed satisfaction with the facility in providing improved service for those buses that travel through the South Corridor. The park-and-ride lots in the South Corridor also provided an advantage to HOV travelers, particularly for those who use the transit buses.

As this facility is developed toward long-term application from Tacoma to Seattle, performance should improve with permanent construction and extension of the lanes and with WSDOT (and local agencies) providing future HOV support programs and other connecting HOV facilities. Improved performance will also depend on how well the regional community will adapt to a general HOV approach to transportation.

Recommendations

Geometric Improvements

It is recommended that appropriate geometric changes be made in the facility to alleviate the bottlenecks for the particular intention of improving effectiveness and public perception of the HOV lanes.

First, the lanes should be extended downstream in each direction. The northbound lanes should be extended to I-405.

This recommendation is supported by WSDOT's final report for the South Corridor: "In order to avoid this potential bottleneck, the HOV lane should be extended past these ramps to Southcenter and I-405. Therefore, the section between S. 200th Street and I-405 becomes a top priority" (*2*).

Second, geometric improvement should also be made to the HOV lanes. The lane striping should be modified at the HOV lane endings in each direction, so that the HOV lanes extend and become mixed-flow lanes (see Figure 5). For WSDOT to maintain the same number of mixed-flow lanes downstream, the far right lane would have to be changed to an exit-only lane at the appropriate exit.

This type of lane configuration was proposed as an alternative (Plan B), as presented in a White Paper from Parsons Brinckerhoff for WSDOT (*10*). By this plan, it was recommended that the operation of the southbound HOV and mixed-flow lanes would be improved for both safety and congestion impacts. Current traffic volume exiting to SR 516 in the southbound is sufficient (~1,200 vehicles per hour in the peak) to justify an exit-only lane, with no negative impact to mainline traffic. The traffic volume exiting at South 200th in the northbound is not as easily justifiable for an exit-only lane (~400 vehicles per hour in the peak). Therefore in the context of current HOV volumes under the 3+ definition, it is recommended that the exit-only configuration be developed in the northbound lanes and that the HOV be extended lane to I-405.

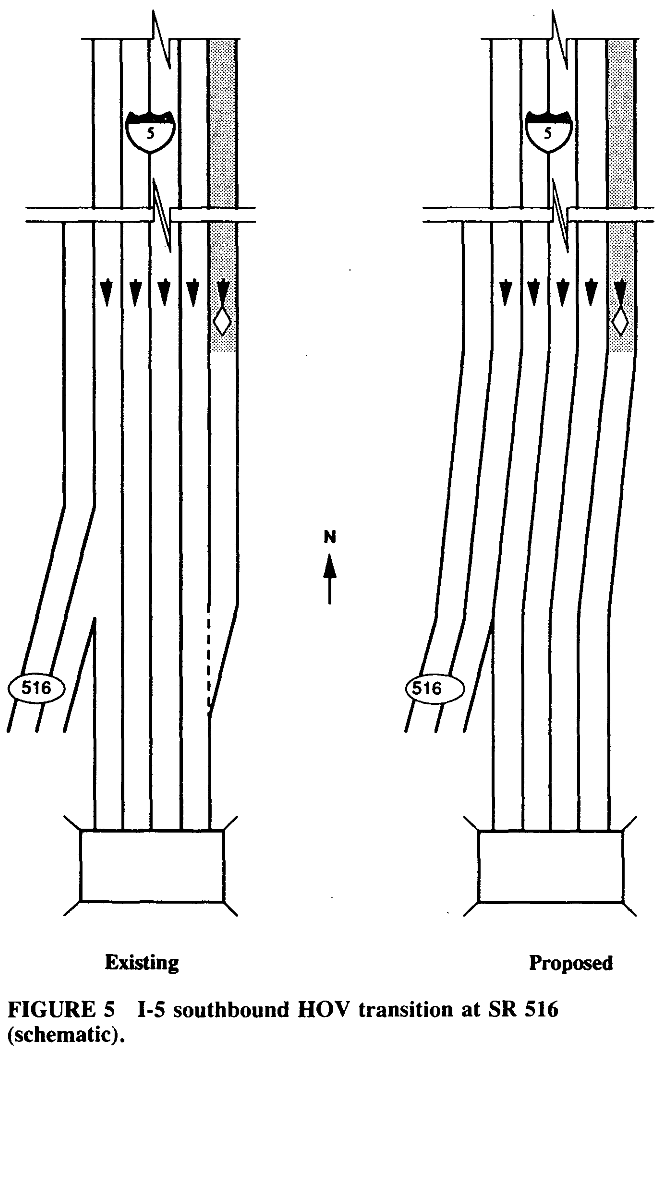

FIGURE 5 I-5 southbound HOV transition at SR 516 (schematic).

2+ Demonstration

A new evaluation should be done for this facility after the recommended geometric improvements are made. This evaluation should reflect operational improvements in the corridor on the basis of the measures presented here. With speculation by the researchers that HOV use will be satisfactory for efficient use, it is also recommended that, following the appropriate geometric changes, a demonstration should be conducted for the feasibility of a 2+ occupancy definition. The 2+ demonstration should be conducted similar to the 2+ demonstration conducted in another HOV facility on I-5 in north King County (*11*).

Because bottleneck problems exist at the HOV termination points under the 3+ definition, excessive impacts to the mixed-flow traffic is likely with an increase in volumes under a 2+ definition. Therefore it is essential that a change to 2+ be made as a demonstration only, and only after the recommended (minimum) geometric improvements have been made and evaluated.

SUGGESTIONS FOR FURTHER RESEARCH

It appears that each of the HOV effectiveness measures provided helpful information for evaluation of this HOV facility. However the research team discovered that there is a relative rank of productiveness provided by the measures, according to available resources. The most effective evaluation measurements came from the opinion surveys and the vehicle occupancy data. Analysis of both of these data types revealed how travelers in the corridor perceived and reacted to existence of the HOV lanes, perhaps the most critical elements in the evaluation of HOV facilities.

The least effective measurements came from travel-time data. With the efforts of the research team to collect comprehensive data, many limitations were discovered. For travel-time data to be completely effective, it should be comprehensive, representing all days over the entire length of the facility and over the entire peak period. This is an expensive and time-consuming process, particularly with the incorporation of traffic observers recording license plates. However if other techniques, such as automobile detector or video applications could be made, the feasibility daily peak-period collection could improve.

Travel-time data would also become much more beneficial if a reasonable means were developed for determining the quantitative (threshold) significance of travel-time reliability to HOV travelers. WSDOT has developed a new policy on travel-time reliability that should help to address this issue in future evaluation studies. It is also important that all types of data be collected from highway corridors for which HOV lanes have been planned before the HOV lanes are opened.

Frequent observation of vehicle interaction, beyond simply collecting and analyzing data, cannot be sufficiently emphasized. The process of training and incorporating observers is expensive and time-consuming, but because each highway corridor is unique, particularly in short-term application, field observations are probably the most important part of the HOV monitoring and evaluation process.

New evaluation should begin immediately, particularly following each of the suggested geometric and operational improvements. Whether or not the improvements are made, the data limitations discussed here should be addressed by data collection that represents all seasons and accounts for the entire peak period. The opinion surveys should include questions related to mode shift and perhaps be directed to travelers who changed modes after the HOV lanes opened. Data collection could also include automobile occupancy at entrance and exit ramps within the corridor. Methods should also be developed for more observation and data collection at the HOV lane termination points.

Perhaps the most beneficial research for this corridor will occur over time as the region begins to realize more impacts from transportation and land use policy changes. Evaluation of this corridor should be continuous over time and should be incorporated into the regionwide HOV monitoring program. This should include more concentrated evaluation efforts by researchers as the South Corridor improvements and extensions are made toward the ultimate facility plan of Seattle to Tacoma.

ACKNOWLEDGMENTS

The author would like to thank the following individuals and organizations involved in this project: U.S. Department of Transportation Region 10 Research Center, Transportation Northwest at the University of Washington; University of Washington professors Scott Rutherford, Fred Mannering, and classmate Maria Escude; TRAC staff members Ron Porter, Elaine Donovan, Duane Wright, Mark Hallenbeck, Roschell Farnsworth, Barbara Miller, and Susan Seferian; TRAC traffic observers; Eldon Jacobson, Les Jacobson, and Ken Knutson of WSDOT; Pierce Transit; and the Seattle Metro.

REFERENCES

1. *Washington State Freeway HOV System Policy.* Executive Summary. WSDOT, Olympia, Wash., 1991.
2. *Final Recommendations Report: I-5 South Interim HOV Project.* WSDOT, Olympia, Wash., July 1991.
3. K. F. Turnbull, et al. *Suggested Procedures for Evaluating the Effectiveness of Freeway HOV Facilities.* Technical Report 925-2. Texas Transportation Institute, College Station, Tex., Feb. 1991.
4. C. A. Fuhs. *High-Occupancy Vehicle Facilities, Current Planning, Operation, and Design Practices.* Monograph 5. Parsons Brinckerhoff, New York, N.Y., Oct. 1990.
5. California Air Resources Board. *High Occupancy Vehicle System Plans As Air Pollution Control Measures.* Sacramento, Calif., May 1991.
6. C. K. Nuworsoo and A. D. May. *Planning HOV Lanes on Freeways: Site Selection and Modal Shift Prediction.* Institute of Transportation Studies, University of California, Berkeley, Feb. 1988.
7. Parsons Brinckerhoff Quade & Douglas, Inc. *HOV Development Plan for State Route 91.* Calif., Dec. 1989.
8. F. Cechini. Operational Considerations in HOV Facility Implementations: Making Sense of It All. FHWA, Sacramento, Calif., Oct. 1988.
9. F. Mannering and M. Hamed. Commuter Welfare Approach to High Occupancy Vehicle Lane Evaluation: An Exploratory Analysis. *Transportation Research*, Vol. 2A, No. 5, 1990, pp. 371-380.
10. *Interstate 5 South Interim HOV Lanes, Peak Hour Analysis of 2+ and 3+ Operation.* White Paper by Parsons Brinckerhoff. WSDOT, Olympia, Wash., Aug. 1991.
11. C. Ulberg. I-5 North HOV Lane 2+ Occupancy Requirement Demonstration Evaluation. WSDOT, Olympia, Wash., Feb. 1992.

Publication of this paper sponsored by Committee on High-Occupancy Vehicle Systems.