TRANSPORTATION RESEARCH RECORD 657

Applications of Interactive Graphics

TRANSPORTATION RESEARCH BOARD

COMMISSION ON SOCIOTECHNICAL SYSTEMS
NATIONAL RESEARCH COUNCIL

NATIONAL ACADEMY OF SCIENCES
WASHINGTON, D.C. 1977

Transportation Research Record 657
Price $3.00
Edited for TRB by Mary McLaughlin

subject area
84 urban transportation systems

Transportation Research Board publications are available by ordering directly from the board. They may also be obtained on a regular basis through organizational or individual supporting membership in the board; members or library subscribers are eligible for substantial discounts. For further information, write to the Transportation Research Board, National Academy of Sciences, 2101 Constitution Avenue, N.W., Washington, D.C. 20418.

Notice
The views expressed in these papers are those of the authors and do not necessarily reflect the views of the committee, the Transportation Research Board, the National Academy of Sciences, or the sponsors of Transportation Research Board activities.

Library of Congress Cataloging in Publication Data
National Research Council. Transportation Research Board.
Applications of interactive graphics.

(Transportation research record; 657)
1. Urban transportation—Data processing—Congresses. 2. Computer graphics—Congresses. I. Title. II. Series.
TE7.H5 no. 657 [TA1205] 380.5'08s [388.4'028'5443]
ISBN 0-309-02686-5 78-25724

Sponsorship of the Papers in This Transportation Research Record

GROUP 1–TRANSPORTATION SYSTEMS PLANNING AND ADMINISTRATION
E. Wilson Campbell, New York State Department of Transportation, chairman

Transportation Forecasting Section
John R. Hamburg, John R. Hamburg and Associates, chairman

Committee on Computer Graphics and Interactive Graphics
Jerry B. Schneider, University of Washington, chairman
William G. Barker, Robert B. Cody, Robert B. Dial, Larry J. Feeser, William A. Fetter, Yehuda Gur, Hubert A. Henry, Robert H. Hinckley, Edgar M. Horwood, Larry L. Howson, Everett S. Joline, Donald C. Kendall, Gene Letendre, Paul Michel Lion, Peter S. Loubal, Patrick E. Mantey, Harold Moellering, Lloyd M. Pernela, Matthias H. Rapp

James A. Scott, Transportation Research Board staff

The organizational units and the officers and members are as of December 31, 1976.

Contents

Interactive Graphics Sketch-Planning Model for Urban Transportation

Alain L. Kornhauser and Jamieson L. Hess, Princeton University

This paper presents a description of an interactive graphics sketch-planning model for urban transportation developed at Princeton University. The model was developed to operate in an IBM 370 VM/CMS environment with functions and subroutines written in both APL and assembler language to improve flexibility and speed in accessing and manipulating large networks and data bases. Benefits of the interactive sketch-planning aspects of the model are described. The sensitivity of sketch planning to spatial aggregation of demand data and remedies for alleviating the inaccuracies resulting from spatial aggregation are discussed.

The need for solutions to the current problems of urban transportation has given a sense of urgency to the transportation planning process. While energy shortages, traffic congestion, and pollution have imperiled the future of the automobile in urban areas, public transportation has been beset by mounting financial difficulties. New, more flexible approaches are required if transportation planners are to accomplish their mission.

In the attempt to improve transportation facilities, attention has been increasingly directed to the use of mathematical models as aids in the planning process. A well-known example is the Urban Transportation Planning System (UTPS) model developed for the Urban Mass Transportation Administration (UMTA). Such models assume that individual decisions by travelers as to when, where, and how to make their trips can be estimated by a set of mathematical equations programmed on a digital computer. Difficulties associated with earlier versions of these transportation planning models were that (a) they were cumbersome to iterate on system design options because output data were difficult to interpret, (b) input data were voluminous and difficult to "digest," and (c) software typically consumed large amounts of time and money.

The transportation planning process is essentially an iterative procedure in which the next move to be made is decided on the basis of the results of the preceding move. This is the approach taken by interactive sketch planning; it allows dialogue between person and computer and enables iterations to be carried out as a continuous process rather than advanced in fits and starts. Moreover, interactive computing is ideally suited to permit the transfer of information between person and computer in easily visualized graphic form rather than in the complicated digital form typified by punched cards.

Interactive sketch planning is not the only valid approach to the transportation planning process, but it does provide a vital link between the coarse types of analysis that indicate basic relations and detailed, microscopically oriented procedures such as real-time simulations. In fact, a sketch-planning model can be used to accept results from a coarse model such as a parametric analysis and then pass on these results as input to a detailed simulation of system operation. One of the most critical tasks that a sketch-planning system can accomplish is the evaluation of a number of network configurations including alternative routings and station locations and the identification of an optimum configuration. This is a difficult problem for which closed-form analytic solutions exist only for the most trivial networks. Given any realistic structure for transportation demand, the analytical problem of specifying the optimum network configuration cannot even be formulated let alone solved because the performance index cannot be made explicit. This is where the planner, the transit system designer, and the policy maker are required to have direct input into the transportation design process. Given data in an easily digestible form, they can weight the various performance measures, determine the impacts of previous iterations, and propose new configurations that would improve the system design. By use of a rapid and inexpensive sketch-planning model that presents results in an easily understood form, designers, planners, and policy makers can experiment with new configurations as a learning process and ultimately fine-tune preferred configurations for final recommendation.

Although several sketch-planning computer models have been developed, it is the premise of this paper that the Princeton sketch-planning model described here, which is more fully documented by Hess and Kornhauser (1), is the only one that begins to meet the interactive graphics user-oriented objectives set forth above. The software itself is rather straightforward. It contains some innovations in that it applies and implements concepts that are only discussed by others. It would be incorrect to imply that the analysis could only be carried out by using an interactive graphics system. In fact, standard UTPS evaluates transit system configurations without use of interactive graphics, and the present interactive graphics model was developed from a batch-process model. But the time-related benefits of the interactive and graphic capabilities are considerable, producing savings in (a) the time necessary to produce the first network and subsequent networks in machine-readable form, (b) the time necessary to ensure the elimination of network coding errors such as tunnel links, and (c) the time to digest and interpret the output of the network analysis process.

Additional benefits include the fact that even a novice user can effectively use the model and learn to appreciate even subtle trade-offs in transit network design. The experienced user can also benefit in that the simplicity of model applications enables the user to probe model performance to greater depths. It was this capability that led to the analysis of the sensitivity of model forecasts of transit ridership to the level of spatial aggregation of base demand data.

DESCRIPTION OF THE SKETCH-PLANNING MODEL

Purpose and Structure

The purpose of the Princeton University project was to design and implement on the computer an interactive transportation planning system. This system was to be fast, inexpensive, and reliable in achieving those answers it set out to achieve. It would combine the facilities necessary for rapid "number-crunching" operations on large data bases with the flexibility required to provide two alternative methods for graphic input, editing, and output of transportation data. That is, it would allow the user to describe, for example, a transportation network in visual terms as he or she sees it on a map. The computer would assume the responsibility for converting graphic constructions and images into their digital representations within the program, analyze the performance of the network, and report its findings by using both graphic and numeric media.

The sketch-planning model was structurally designed so that it could be applied to a wide range of different transportation systems. It was to have the ability to compare the relative merits of two competing transportation modes over a specified geographical area and to allow successive modifications to one mode—the other mode being held constant—in the interest of improving its performance relative to the other mode. This structure would normally be applied to real conditions by holding constant an existing fixed system, such as an urban street network, and varying against it different configurations of a proposed fixed-guideway mass transit system or a new route structure for an existing bus system. However, it would be possible to vary the network originally held constant against a single network from its competitor mode by simply interchanging data bases, if that were desired.

Princeton University Computer System

The program has been developed, and is currently undergoing further development, on Princeton University's IBM 370/158 computing system. The 370 is linked by means of a file spooling system to Princeton's other primary computer, an IBM 360 model 91; these two computers and the various devices attached to them provide the program with great versatility in the areas of input and output. The heart of the program lies in a "workspace" consisting of functions and subroutines written in the APL computer language. APL is particularly well suited to interactive computing because of the ease with which user-oriented input-output structures can be implemented.

In its current version, the workspace operates in the environment of APL/CMS, an implementation of APL written by IBM to run on their VM/370 operating system, which Princeton uses. Although the program is capable of functioning as a self-contained unit within an APL workspace, it presently takes advantage of the external file system provided by the VM/370 system to access large data bases and execute assembled object codes stored outside the workspace for efficiency. This option, which was specially added to the Princeton version of APL, enables the program to overcome the inefficiencies presented by program loops in an interpreted language such as APL (as opposed to an assembled or compiled language) simply through the process of programming the innermost nested loops in assembler language. The program thus benefits concurrently from both the speed of IBM assembler language and the flexibility of a high-level interactive language such as APL.

The model takes advantage of the great variety of input-output devices available at Princeton University. Input to the workspace is accepted primarily from Tektronix model 4013 and 4015 interactive cathode ray tube (CRT) terminals, which use the APL and ASCII character sets and allow both graphic input (via cross-hair cursors) and output (via vector lines) on the terminal screen. In addition, graphic input from maps can be entered on a Tektronix model 4954 digitizer, and card images can be sent from a card reader through the 360/91.

Output too goes direct to the user who sits at a Tektronix CRT, and images on the screen can be hard-copied on an attached dry-silver recording copier. Output may also be sent by means of the 360/91 to a variety of other devices including Calcomp 936 and 565 plotters and IBM line printers or card punches. Thus, input and output options are not at all constricting.

The program as applied to a study of Trenton, New Jersey, for engineering laboratories resides in an "active" workspace of approximately 170 kilobytes of dynamic core. Because the capacity of the workspace has yet to be exceeded, it is unclear how large a network may be analyzed within this structure. Networks containing 120 or more stations have been accommodated without incident. The program itself, stored on disk, occupies approximately 45 kilobytes of functions, subprograms, the symbol table for the APL system, and data internal to the workspace. In addition, external files containing the demand data base and assembled object decks fill about 25 to 30 kilobytes of additional disk space.

The following are the capabilities of the sketch-planning model:

1. The network can be input by means of either a CRT or a digitizer tablet.
2. Network editing in terms of station, link, or intersection deletion, addition, or translation can be accomplished by using the CRT (Figure 1).
3. Flexible "windowing" features allow interactive scanning and focusing on specific parts of the network (Figure 2).
4. Pathfinding and network-assignment routines have been written in assembler language to improve computation speed.
5. Complete transit system analysis routines include a choice of several mode-split analysis techniques and route assignment and economic analysis routines.
6. Various levels of formatted output—graphic supported by numeric—are available.

FUNCTIONS OF THE MODEL

A more detailed description of the methods used by the interactive sketch-planning model is presented below.

Initialization of the Data Base

To analyze the performance of a transit network in a given geographical area, the digitizer or CRT screen must first be associated with the coordinate system of the region to be studied. An initialization function called MAPINIT is available for this purpose. MAPINIT also allows the user to input in graphic form a backdrop of political boundaries, streets, or other physical features that can be displayed or generated on maps in conjunction with the interactively designed transit network.

The next step involves initializing the data base, which normally consists of the location of demand points and two matrixes containing origin-destination (O-D) demand information and disutility for automobile travel between each demand point. In addition, if it is desired to make use of the uniform demand distribution option for some or all traffic assignment zones, two additional pieces of information must be supplied for each zone: the size of the zone and its demand characteristics (whether uniform distribution, point source, or a cross between the two). Two functions, DEMAND and EDGE, can be used to perform these tasks. When all this information has been entered into the computer by the most convenient method, it is compressed by using a bit-packing function and stored principally in external data files on an on-line disk by means of the STORE function.

Network Design Procedure

The normal network design process involves sketching an initial network on a map of the study region, placing the map on a digitizer, and tracing over the network with an electronic pen or "bullseye." Transit networks are constructed by means of links interconnecting stations at which passengers are permitted to board and depart from the network, and interchanges, at which three or four links are allowed to come together. The initial input procedure for the digitizer is controlled by a function called TAB3. Alternatively, if the user has provided a geographic background that is sufficiently detailed to obviate the digitizer, he or she may sketch the network directly on the CRT screen by using the CRT3 function, which accepts input through the use of the crosshair cursors.

After the initial network loop has been accepted by the computer, the user may then enter the network editing mode, which permits manipulation of the network

Figure 1. Network map and political boundaries produced by NETMAP.

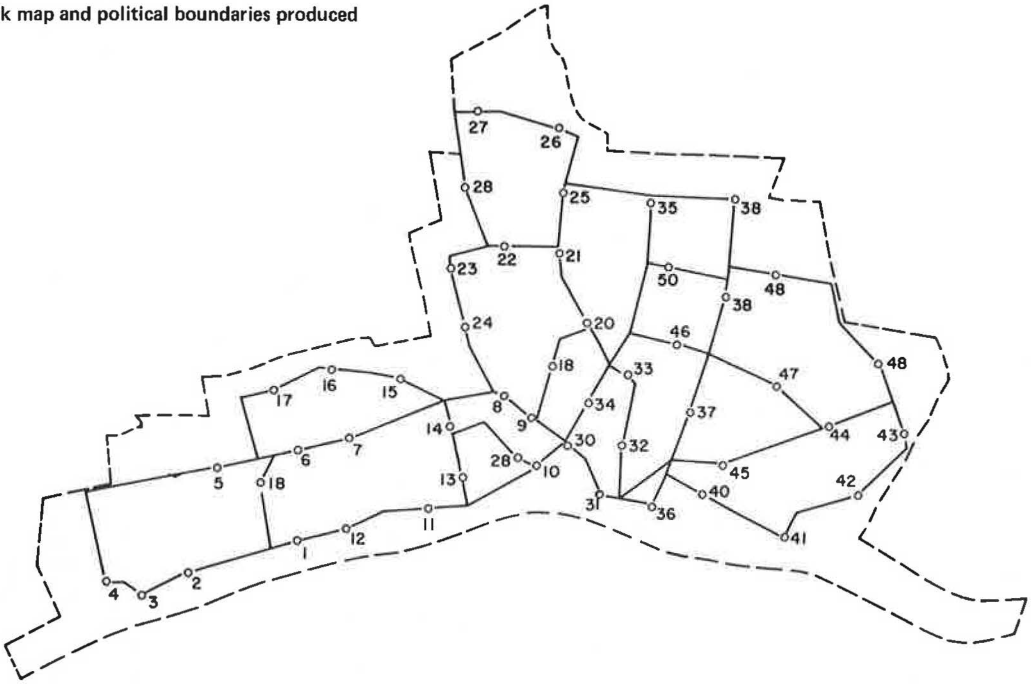
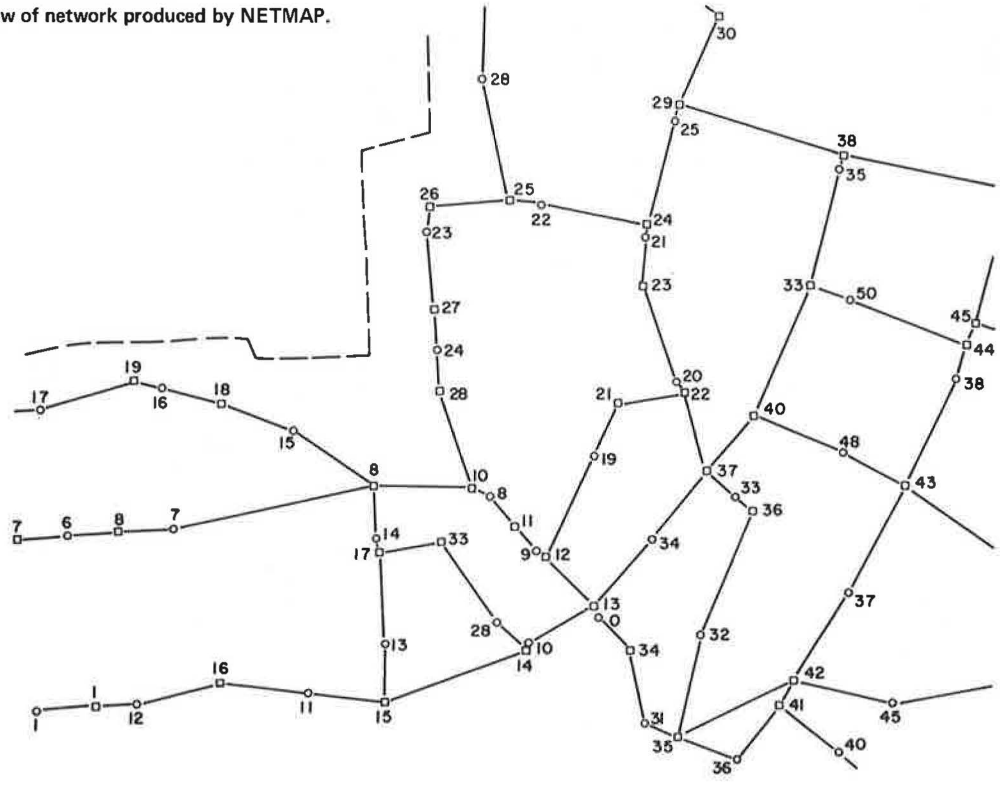


Figure 2. Window of network produced by NETMAP.



by use of the following functions: ADDINT inserts an interchange on an existing link; ADDLINK adds an entirely new link to the network; ADDSTA inserts a station on an existing link; DROPINT deletes an interchange from a link; DROPLINK deletes an entire link from the network; DROPSTA deletes a station from a link; REVERSE changes the direction of flow of a link; and TERMOVE moves a node to a new location.

After having entered a series of modifications, the user can exit from the network editing mode by means of the STREAM function, which updates the network on the basis of the modifications ordered by the user. This function also reassigns node numbers to improve future computational efficiency and creates a buffered set of graphic control characters that contain the image of a network map. These characters may be dumped to the terminal screen at any time along with the NETMAP function to produce a map (Figure 1). NETMAP allows the user to specify exactly which of five sets of graphic information should be displayed in the map: the geographic backdrop, the network itself, station node labels, interchange node labels, or directional arrows on links.

When a map of standard scale produced by NETMAP is on the terminal screen, a selected portion of the map may be expanded to fill the entire screen, a process sometimes referred to as windowing. This process, which is handled by a function called ENLARGE, permits close examination of sections of a compact, detailed network for increased clarity. The process can be repeated in recycled form any number of times until the scale is sufficiently large (Figure 2). Two options for the NETMAP and ENLARGE map-generation functions (referred to as NETMAPC and ENLARGEC) are used to send output to a Calcomp paper-and-ink plotter as well as to the terminal screen.

The next step in the analysis procedure is to compute distances around the interactively designed network for the purpose of determining station-to-station travel time. This is a four-stage process. The first stage involves setting up data on adjacent nodes on which an algorithm for minimum-path calculation can operate. The original approach consisted of using a function called ADJACENT to create an "adjacency matrix" in which all values were infinity except those elements representing direct connections between adjacent nodes. The adjacency matrix was then operated on by the MPATHALL minimum-path function, which took advantage of the extendability of APL operators to multidimensional arrays to achieve unusually efficient execution times for APL. This algorithm also made use of the loop network structure and sequential node numbering to improve efficiency in the generation of a predecessor node matrix. Although this algorithm was designed according to the characteristics of the networks on which it was to operate and is one of the fastest available in APL, when the ability to load and execute assembled object decks from within the APL workspace was added to the Princeton APL system, no APL minimum-path program was able to compete with assembler language. As a result, ADJACINT and MPATHALL were redesigned to make use of an assembler minimum-path program written by Cormen, a Princeton student. For input this program requires, instead of an adjacency matrix, a "link list" that specifies for each link the distance (or travel time) and the number of the node on the downstream end. This link list is indexed with a set of pointers that indicate the identities of the upstream nodes.

When minimum paths among interchanges have been determined, the information must be translated into station-to-station distances. The NERINT function finds the distances between interchanges and adjacent stations and performs this translation.

The final stage in the assembly of travel-time information is the most crucial: the determination of feeder access from travel demand sources to stations on the transportation network. It is at this stage that the effects of aggregation of travel demand manifest themselves in an underestimation of the average access distance between demand point (zone centroid) and transit station.

The original version of NERSTA treated zone centroids as point sources of demand and merely determined the direct walk distance from centroid to station. During the study of aggregation effects, however, this approach was found to be invalid except when it was applied to disaggregate data. The new version of NERSTA makes use of three added features as inputs: (a) the area of each traffic assignment zone; (b) the demand characteristic of each zone (uniform, point, or superposition); and (c) the function CIRDIST, which finds the average distance between the station and the locus of points within a circular zone as a function of the radius of the zone and the distance between its center and the station.

The new NERSTA produces a considerably more realistic approximation of walk access than does the original version. This formula of course cannot hope to replicate exactly the disaggregate characteristics of walk access because of the effects of averaging resulting from aggregation, but it yields access data whose average is much closer to the average across the disaggregate data base.

After the completion of NERSTA, the total transit travel time for every O-D pair can be compiled and a transit disutility calculated. For each trip or group of trips with a common origin and destination, COMPATH examines the 16 possible choices of routings on the transit network represented by the combinations of the four stations nearest to the origin demand point with the four stations nearest to the destination point. After weighting the feeder access (i.e., walk) and ride times by means of the user-specified feeder access speed and transit vehicle speed, the function picks the minimum disutility path of the 16 possible paths and computes a relative disutility value between the transit mode and the automobile mode. It then examines the origin and destination stations to see if they are the same, in which case that trip (like all intrazonal trips) is assigned to the pedestrian model. When all walk trips have been eliminated from further consideration, a logit model is used to calculate the transit ridership between each O-D pair.

In the process of determining mode split, a record is made of transit travel time and feeder access distance for all but the intrazonal trips regardless of the mode to which any trip was assigned. This information is later displayed in a pair of graphs and tables for reference. Meanwhile, now that the function has decided exactly which trips are to be assigned to the user's network, the program loads these trips onto the network so that passenger volumes on links and in stations can be determined.

As it performs these operations, COMPATH accesses two external data files that contain demand data and competing-mode, travel-time disutility and reads them sequentially into the workspace, one record at a time. Thus, only one record of each of these large data files is in the workspace at any time, a feature that results in substantial economies with respect to dynamic core requirements for program execution. Each record, being a single row of a matrix, contains information on all trips that originate at a particular demand point. The records are decored from bit-packed into normal

APL integer format and, when the travel-time differences have been found, APL branches to an assembler program that tabulates time difference, transit travel time, and walk distance and loads transit trips onto the network by tracing each path through the predecessor node matrix. When this process is complete, or if the demand point did not generate any transit trips, the program returns to APL, which reads the next record from each file. Processing continues until the last demand point has been examined, at which time APL outputs the results to the user's terminal.

COMPATH sends information to the terminal on overall mode split based on the logit model; this is followed by a table on travel-time differences for trip ends, differentiated by demand point. This table is perhaps the most important feature because it presents in detail an indication of the level of service that the user's network offers to each demand point. The table not only shows which demand points have the densest concentrations of demand but, more importantly, which demand points exhibit dense concentrations of demand at or near the point of zero utility difference. It is this demand that can frequently be transferred into the transit column by means of minor modifications to the interactively designed network implemented with the aim of improving service to specific demand points.

After the detailed time-difference table has been generated, the time-difference distribution for the entire study area is listed and plotted (Figure 3), to be followed in turn by graphs and tables of transit walk access and travel time and station and link volumes. Once calculated, the link volumes will show up on all enlargements until the network is modified.

Although COMPATH is the "workhorse" of the workspace, it normally does not represent the end of the analysis procedure. Link volumes generated by COMPATH can be used to create a "bandwidth" plot with the function LOADNET in which each link on the network is drawn in a width proportional to the passenger traffic on that link (Figure 4). The option LOADNETC sends a bandwidth diagram to the Calcomp plotter. In addition, MINIPATH, another available function, performs mode-split duties, draws minimum-path trees out of a designated origin demand point, and writes travel-time information for the user's network into an external data file. Finally, a COST model is available to provide a rough economic analysis of the network including total capital cost, annual capital cost (as determined by the capital recovery factor), annual operating and maintenance cost, and total annual cost, both in absolute terms and in terms of expenditures per unit distance of passenger travel.

At this point, the user must make an important decision: (a) decide that he or she has seen enough of the analysis procedure and quit or (b) return to the network editing mode and attempt to improve the network on the basis of what has been learned from the results of the analysis. If the user chooses the second course, he or she enters the iteration process in which the search for the optimum centers on the dual aim of increasing ridership while decreasing average cost. The iteration process continues until the user is satisfied with the network. Typical users working with the Trenton data base have achieved relatively well-optimized networks after, at most, 20 iterations.

By using this program, a network can be created from scratch in as little as 15 min and subsequent iterations can be performed within 5 to 10 min. Because execution time for the current version of the workspace depends primarily on the number of demand points in the data base rather than on the size of the network, execution time for one iteration is relatively constant within the range of 45 to 60 s. In terms of daytime rates on the Princeton 370/158, the typical cost per network iteration on the Trenton data base has run to about $5, including "connect time"; after midnight the charges are half as much, or $2.50. In view of the reasonable cost and the advantages offered by time sharing, it is hoped that this program will see increasingly wide application in the future.

IMPACT OF SPATIAL AGGREGATION

Because the model could be readily executed and modified, several analyses were undertaken to investigate the sensitivity of some of the analytic techniques used in the sketch-planning process. The primary focus of one investigation, which probably would not have been undertaken if it were not for the availability of a flexible and easily implementable sketch-planning model, was to establish the sensitivity of the transit ridership forecasts of the mode-split modeling process to the degree of specificity of O-D location.

It is well-known that the mode-split task of any transit demand forecasting project is the task that is most sensitive and prone to error. Questions arise as to the functional form of the mode-split model and the value of various coefficients. A further question has been raised recently about the impact of spatially aggregating demand data into zones by assuming that all trips within that zone are generated at the zone centroid. Of principal interest here is the spatial aggregation of trip-generation points to the centroids of zones.

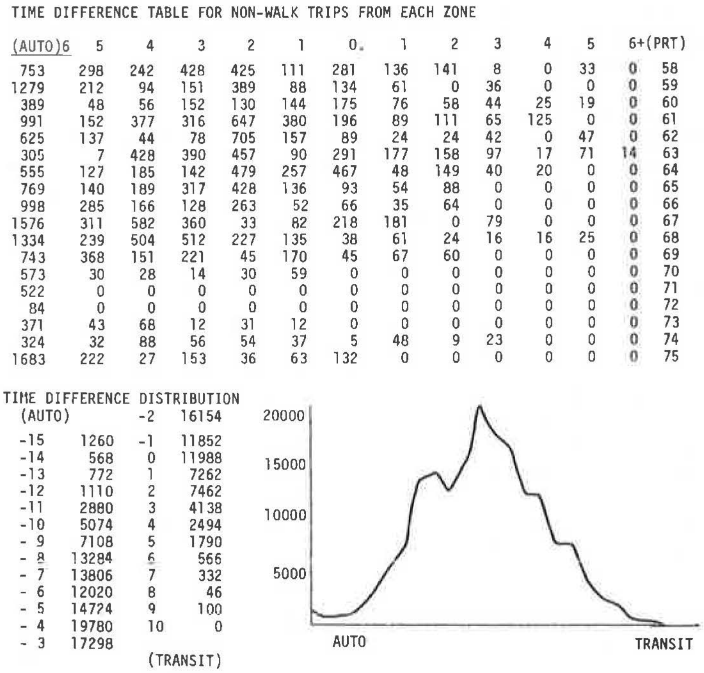
Mode-split modeling could be considered an art if it were not founded on the behavioral consumer choice theory of individual utility maximization. Any application of the mode-split models should be done at the level of the individual traveler unless it can be documented that the aggregation of travelers into any group does not bias the quantitative output of the mode-split model. Nonetheless, the first step normally taken in the urban transportation planning process is to aggregate travel demand into traffic assignment zones finite (and often large) in size. This is usually indispensable because raw data do not exist, computer storage facilities are inadequate, or the repetitive process of analyzing each traveler would require excessive computer time. Mode-split estimates are then made without validation of the spatial aggregation of the data; at least the literature does not contain documentation of that validation.

It may be argued that spatial aggregation is acceptable as long as the disutility collection and distribution elements of the transportation system are insignificant as compared with the line-haul disutility of urban trips. This is true for most automobile trips because walk distance at both trip ends is usually small and the street system is ubiquitous. This is not the case for many transit trips and especially those that must depend on the walk mode for collection and distribution. For these systems, the access modes contribute heavily to total trip disutility to such an extent that the disutility of the line-haul portion of the trip may even become insignificant (2).

Travel Demand Data

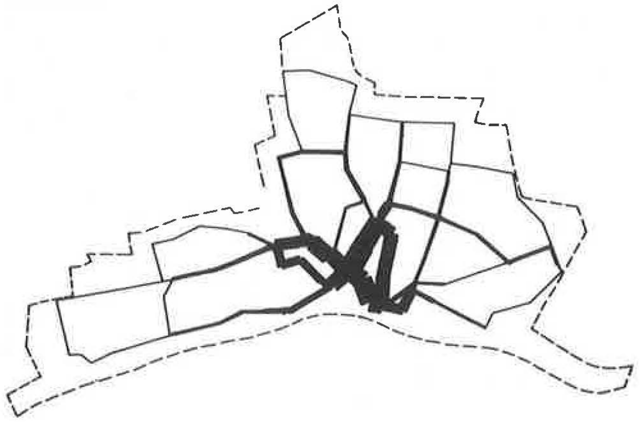
An analysis of the effects of spatial aggregation on the patronage forecasts of the sketch-planning model made extensive use of data collected by the Princeton University Transportation Program during its study of transportation alternatives for the city of Trenton, New Jersey. Of the data available from this study, the most valuable input was a collection of disaggregate travel demand information compiled from a 1972 home inter-

Figure 3. Travel-time differences presented on the terminal screen.



```
TIME DIFFERENCE TABLE FOR NON-WALK TRIPS FROM EACH ZONE

(AUTO)6   5    4    3    2    1    0.   1    2    3    4    5    6+(PRT)

  753   298  242  428  425  111  281  136  141    8    0   33    0  58
 1279   212   94  151  389   88  134   61    0   36    0    0    0  59
  389    48   56  152  130  144  175   76   58   44   25   19    0  60
  991   152  377  316  647  380  196   89  111   65  125    0    0  61
  625   137   44   78  705  157   89   24   24   42    0   47    0  62
  305     7  428  390  457   90  291  177  158   97   17   71   14  63
  555   127  185  142  479  257  467   48  149   40   20    0    0  64
  769   140  189  317  428  136   93   54   88    0    0    0    0  65
  998   285  166  128  263   52   66   35   64    0    0    0    0  66
 1576   311  582  360   33   82  218  181    0   79    0    0    0  67
 1334   239  504  512  227  135   38   61   24   16   16   25    0  68
  743   368  151  221   45  170   45   67   60    0    0    0    0  69
  573    30   28   14   30   59    0    0    0    0    0    0    0  70
  522     0    0    0    0    0    0    0    0    0    0    0    0  71
   84     0    0    0    0    0    0    0    0    0    0    0    0  72
  371    43   68   12   31   12    0    0    0    0    0    0    0  73
  324    32   88   56   54   37    5   48    9   23    0    0    0  74
 1683   222   27  153   36   63  132    0    0    0    0    0    0  75

TIME DIFFERENCE DISTRIBUTION
 (AUTO)          -2   16154

 -15    1260     -1   11852
 -14     568      0   11988
 -13     772      1    7262
 -12    1110      2    7462
 -11    2880      3    4138
 -10    5074      4    2494
 - 9    7108      5    1790
 - 8   13284      6     566
 - 7   13806      7     332
 - 6   12020      8      46
 - 5   14724      9     100
 - 4   19780     10       0
 - 3   17298
                 (TRANSIT)
```

Figure 4. Bandwidth plots of assigned link densities produced by LOADNET.



view survey that covered a 14.7 percent sample of Trenton's population of 105 000. The respondents to this survey reported making approximately 24 800 trips/d, of which slightly more than half, or 12 600, were intracity trips whose origins and destinations were all within the Trenton city limits. The scope of the analysis was limited to those intracity trips and the city of Trenton alone.

O-D geographic locations for these trips were coded in terms of census blocks, which correspond very closely with actual city blocks. Trenton contains 1217 of these census blocks. This implied a geographic resolution of approximately 60.96 m (200 ft). After assembly of the interview data, the trips were "factored"—based on the number of interviews conducted in each section of the city as a proportion of the number of residents in each section—into a total of 99 408 intracity trips estimated to be taken by city residents on a typical day. This information constituted the disaggregate data base.

Accompanying the disaggregate data was an aggregation scheme that clustered the 1217 census blocks into 74 traffic assignment zones for the purpose of simplifying analyses of the various transportation options considered for Trenton during the course of the Princeton University study. Neighboring homogeneous census blocks were grouped to form traffic assignment zones that varied in area from as few as 2 city blocks in the central business district to as many as 33 blocks in residential neighborhoods near the city limits. Average zone size was approximately 0.25 km^2 (0.1 $mile^2$); the geographic location resolution was approximately 243.8 m (800 ft). Each zone in turn was assigned a centroid, which was located at the center of trip-generation activity for the census blocks that comprised each zone. All trips with origins or destinations in one of the census blocks encompassed by a particular zone were then assumed to originate or terminate at the centroid of that zone. Thus, the disaggregate trip data derived from the home interview survey were assembled into a 75-by-75 O-D demand matrix.

Analysis Results

The principal test network—an automated guideway transit (AGT) system with 38.6 km (24 miles) of one-way guideway and 48 stations—was analyzed by using the Trenton demand data first in spatially disaggregate form and then in aggregate zonal form. Overwhelming discrepancies were found between the two analyses. Based on a logit mode-split model that used a weighted travel-time difference between automobile and transit as a measure of utility, the number of daily transit trips assigned by the model rose from 14 811 for the disaggregate case to 31 015 for the aggregate. This represents a difference in the results of 109 percent, attributable solely to the spatial aggregation of demand.

In addition to the disturbingly large shift in mode split resulting from spatial aggregation, there were similar movements in the distributions of transit access distances, travel times, and the time difference between transit and automobile. Compared to the more disaggregate results, the average walk distance involved in a transit trip, as computed by means of the aggregated data, decreased by 45 percent or by nearly half; the average total transit travel time fell by 1.82

min; and the time-difference distribution shifted by an even greater displacement—2.3 min—in favor of the transit system. The discrepancy between the two sets of results in terms of modal split could be eliminated only by holding the logit parameter (α) for the aggregate analysis at the high value of 6 while implementing an extremely drastic reduction in the parameter to 0.198 for the disaggregate:

$$MS = 1/(1 + e^{\alpha \Delta T}) \qquad (1)$$

where

MS = mode split to transit,
α = logit parameter, and
ΔT = travel-time difference (transit minus automobile).

Several other networks were designed and subjected to analysis at each of the two spatial aggregation levels. A comparison of the results is given in Table 1 (3).

The inescapable conclusion from these results is that spatially aggregating demand data to point sources introduces significant variability into the bottom-line results of the mode-split procedure. It results in an underestimation of walk access distances to transit routes and a corresponding overestimation of the willingness of travelers to engage in the extra walking necessary to use transit instead of the automobile.

Models for Spatial Distribution of Trips Within Traffic Assignment Zones

It appears that, whenever the number of stations or other access points on the transportation network being modeled is of the same order of magnitude as the number of aggregation zones, a more uniform distribution of trips across the area encompassed by each zone must be assumed. Most urban areas exhibit more of the characteristics of uniform distributions than of a small number of point sources of travel demand; apparently the city of Trenton cannot be modeled as a set of 75 miniature oases in a 19.4-km^2 (7.5-$mile^2$) desert. Although a strictly uniform distribution across each zone was found to produce slightly overstated values for walk distance and somewhat understated predictions of transit travel demand (the principal test network, for example, was assigned only 14 331 daily transit trips based on a strictly uniform model), its output was much closer to the results of the more disaggregate analysis. When aggregation must be done, it appears that the solution that will yield the most satisfactory approximation is to accommodate the two approaches by realizing that most zones will exhibit the characteristics of a uniform demand distribution but some zones will possess centers of activity that result to some extent in the concentration of demand around a centroid. This approach should serve to minimize the negative effects of aggregation, which, in extreme cases and if used without safeguards, could lead to totally meaningless computer output.

The approach taken in this analysis to accommodate spatial aggregation of data was to set up a calculation model for access distance that would allow a flexible accommodation between the uniform demand concept and the point source. As implemented, this flexibility permits the planner to tell the computer, for each aggregation zone, whether its demand distribution should take on the characteristics of a uniform spreading, a point source, or a superposition of the two concepts in which a specified percentage of the travel demand in the zone is assumed to be based at the zone centroid and the other part is allocated uniformly around the centroid in the remainder of the zone. As applied to Trenton, this new model resulted in the categorization of the 75 traffic assignment zones into three groups: 50 zones using the uniform distribution, 6 using a point source, and the remaining 19 modeled according to the superposition approach. The degree of superposition was estimated from the distribution of trip-end density over the census blocks that comprised each zone. The configuration of each zone was assumed to be an equivalent circular zone with a radius equal to half the average distance between the four nearest centroids.

The assumption of any type of area-based distribution of travel demand requires the development of parameters that express the dimensions of the areas involved. One of the simplest models possible, but one that has proven itself to be accurate in the Trenton analysis, views a traffic assignment zone as a circle centered on the zone centroid. If the area of the zone is expressed in terms of πr^2, the parameter of interest in the determination of access distances is the effective zone radius (r). The radius appropriate for each zone can be arrived at in several ways. Two such methods were examined in this study, the fastest of which involves examining the distances from the centroid of the zone of interest to all other zone centroids and to the city limits. In this study, in view of the irregularities with which many of Trenton's traffic assignment zones were created, several different variations on this first method were considered; each was compared in turn with the ideal results as determined by the second method—the direct measurement of the area of each zone.

Because the Trenton city limits existed as a continuum, it was possible to examine only the single point on the city limits closest to each zone centroid. The distance to this point was then multiplied by 2 to take account of the fact that this distance served to estimate the radius of the zone rather than its diameter, which was the parameter estimated by the distances to neighboring zone centroids. This result was then combined with the pool of distances from the centroid of interest to the other 74 centroids to form a set of 75 distances. The study procedure involved taking the three shortest distances from this pool, then the four shortest, and finally the five shortest, averaging them, and multiplying by 0.56 (a circle of a radius of 0.56 units contains the same area as a square of sides 1 unit long).

Table 1. Comparative analysis of three networks at two levels of spatial aggregation of demand data.

	Principal Network		Alternate Network 1		Alternate Network 2	
Item	Aggregate	Disaggregate	Aggregate	Disaggregate	Aggregate	Disaggregate
Transit demand, trips/day	31 015	14 811	28 136	14 491	16 155	9899
Average walk distance, km	0.283	0.511	0.285	0.525	0.430	0.592
Transit travel time, min	8.297	10.117	8.350	10.319	10.027	11.237
Time difference, min (automobile minus personal rapid transit)	1.960	4.263	2.036	4.489	3.730	5.410

Note: 1 km = 0.62 mile.

Table 2. Results of network analysis process using zonal distribution functions.

Item	Principal Network		Alternate Network 1		Alternate Network 2	
	Aggregate[a]	Disaggregate	Aggregate[a]	Disaggregate	Aggregate[a]	Disaggregate
Transit demand, trips/day	15 136	14 811	16 712	14 491	10 345	9899
Average walk distance, km	0.441	0.512	0.449	0.525	0.536	0.592
Transit travel time, min	10.178	10.117	10.112	10.319	11.259	11.237
Time difference, min (automobile minus personal rapid transit)	3.795	4.263	3.767	4.489	4.880	5.410

Note: 1 km = 0.62 mile.

[a] With new access model.

Results of these calculations were compared with data for zone radius derived from measurements of the areas of individual zones by the following formula: $r = \sqrt{\text{area}/\pi}$. As it turned out, estimates based on the three shortest distances were significantly lower than data derived from measurements, and estimates based on the five shortest distances were significantly higher than the measurement data; estimates based on the four shortest distances, however, were quite close, exhibiting an average value of 23.67 units versus 24.17 units for the data derived from measurements. The method that used the four shortest distances was therefore adopted for estimating zone radius. However, because the average discrepancy between the measured and estimated data was found to be 2.88 units, it was decided to test the new station access model on the basis of the measured data on zone radii. This final step yielded the closest approximation to the disaggregate results achieved by using an aggregate model. The results for three different networks designed for Trenton are summarized in Table 2 (3).

Even a flexible access model that provides for both uniform and point-source demand distributions cannot hope to match the fineness of a disaggregate model on the microscopic level. The statistical variance of the disaggregate data is diminished by the process of aggregation. The uniform model surpasses the point-source model, however, in that a more acceptable estimate of average walk access distance for a transit system is obtained. There will, of course, be distortions in the model estimations, and in some aspects it cannot be regarded as a completely satisfactory replacement for disaggregate analysis. But what is important is that the distributions it calculates shift much closer to their true form.

USE OF THE MODEL

In addition to being a valuable research tool for the critical evaluation of various modeling procedures used in transportation sketch planning and network analysis, the interactive sketch-planning model is in itself a valuable planning tool. Almost 100 persons have used the software to evaluate many hundreds of AGT network configurations for Trenton and for Philadelphia's South Jersey commuter shed. Some of the networks were studied for academic reasons; others were used in AGT feasibility studies and a proposal for a downtown people mover submitted to UMTA by the city of Trenton.

Although a few users have been computer experts in transportation planning, most users have been freshman students enrolled in a 4-week, 16-h minicourse, Introduction to Urban Transportation. Many of these students were liberal arts majors, and most had no previous computer experience. During the 4-week period, students were taught a few concepts about urban transportation planning; learned how to use the sketch-planning model; designed initial network configurations from socioeconomic and land-use maps of Trenton; established their own implicit design performance indexes that weighed patronage, area served, capital and operating costs, and subsidy; redesigned their network several times to improve on their own performance index; and wrote a final report. The fact that this can be accomplished tends to prove that this sketch-planning tool achieves most of the objectives set forth in this paper.

ACKNOWLEDGMENTS

This work was done by using the facilities of the Laboratory of Interactive Computing of Princeton University as part of an ongoing development of interactive curriculum material, an activity supported in part by a grant from the National Science Foundation. The advice and support of K. Alexander and M. Glatz of the staff of the laboratory are very much appreciated.

REFERENCES

1. J. L. Hess and A. L. Kornhauser. Demand Modeling for Automated Guideway Transit Systems. Transportation Program, Princeton Univ., Rept. 76-TR-13, March 1976.
2. T. Domencich and G. Kraft. Free Transit. Heath, Lexington, 1971.
3. A. L. Kornhauser, S. Strong, and P. Mottola. Computer-Aided Design and Analysis of PRT Systems. In PRT III, Department of Audio-Visual Aids, Univ. of Minnesota, June 1976, pp. 377-384.

Volvo Approach to Computer-Aided Transportation Planning

Ingmar Andréasson, AB Volvo, Sweden

A comprehensive, computerized public transportation planning system developed by Volvo is presented. Considerations that governed the design of the system are emphasized. The approach is based on cooperation between person and computer so that the ability of the human planner to survey a complex problem, identify critical areas, and intuitively find solutions is stimulated. An interactive graphics system is used to achieve efficient communication between person and computer. Dialogue optimization tools are included for some subproblems in transportation planning. Limiting tasks so that they can be better controlled by the planner makes possible experiments with objectives and restrictions that are often controversial or insufficiently understood. Experience with the program is limited to areas of up to 0.7 million population, but results are encouraging for future development.

Traffic on fixed routes with fixed timetables is the most common form of public transportation. Even if various forms of demand-responsive transportation increase, route traffic will still form the backbone of public transportation. Therefore, it is of vital interest to improve the conditions for route traffic so that the highest possible standard is offered within given resource limits. The operation of route traffic is labor intensive and therefore expensive. Rising costs combined with growing demand for service call for improved planning.

Volvo has begun to involve itself in better planning of public transportation, extending its responsibilities from the delivery of vehicles to ensuring that the total transportation system serves the needs of its present or potential users. The planning services of Volvo are available to transit operators and planning bodies independently of other Volvo products. The Volvo planning methodology presented here is applicable to route traffic by bus, tram, subway, train, and combinations of these.

BASIS OF METHODOLOGY

Local Conditions

The design of new transportation systems as well as improvements to existing systems must be based on the following local conditions:

1. Present and anticipated trip patterns;
2. Required standard of service;
3. Existing infrastructure (streets, roads, tracks);
4. City planning and environmental considerations;
5. Energy consumption and safety;
6. Funds available; and
7. Overall economy of planning, implementation, and operation.

Local conditions vary to such an extent that it is impossible to apply ready-made solutions and each system must be tailor-made.

Planning Tasks

The quality and the cost of route traffic depend on finding satisfactory solutions to the following planning tasks:

Task	Considerations
Routes	Direct connections Opportunities for transfer Costs
Frequencies	Capacity Waiting times Vehicle requirements
Departure times	Regular departures Quick transfers Vehicle requirements
Vehicle scheduling	Effective utilization of vehicles
Driver scheduling	Working-hour regulations Costs

These tasks are traditionally approached in the above sequence. But the solution of one task cannot be fully evaluated until the consequences for the other tasks are known. On the other hand, detailed planning for the evaluation of each alternative network is undesirable.

The Volvo approach evaluates routes and frequencies together. Departure times on different routes are assumed to be independent, and the numbers of vehicles and drivers are calculated as instantaneous requirements without coordination between routes. Special tools have been designed to assist in the detailed planning of the other tasks.

Human Planners and Computers

The first requirement in the design of a good transportation system is a method for telling how good or bad a given solution is. Although this can be done by hand, it is tedious and time consuming. In practice, manual evaluations are most often limited to cost calculations and rough capacity planning for one already chosen alternative.

By use of computers, the evaluation can be done both from the aspect of the user (travel desires and opportunities) and from that of the transit operator (loading, productivity, and costs). Thanks to the speed and capacity of modern computers, it is possible to evaluate several alternatives within minutes.

Nevertheless, the human planner has a unique ability to survey complex situations, identify critical areas, and intuitively find solutions. The Volvo approach is to establish cooperation between humans and computers and between local planners and Volvo planners. The key to cooperation is communication. The local experience of every transit operator is different and valuable. It is essential that this experience be integrated in the planning process. Because it is not easy to communicate experience to someone else, local planners must take an active part in all planning. On the other hand, only the major transit operators have their own planning tools, and there is an obvious advantage in using a consultant who has had experience with similar problems in other locales.

Volvo has identified communication problems in two dimensions: (a) between human planners and computers and (b) between local planners and consultants. If communication between man and computer is to be efficient, it must be adapted to suit human needs rather than those of the computer. Man-computer communications should

1. Stimulate humans to make use of their ability to survey a situation, identify trouble spots, and intuitively find solutions;
2. Stimulate local planners to communicate experience-based restrictions and ideas to other planners; and
3. Document facts in a form that can be understood by the nonprofessional public.

The work is shared between the human planner and the computer in such a way that (a) the planner proposes a network of routes with frequencies; (b) the computer evaluates in relation to travel demand; (c) the planner modifies, possibly by using optimization tools; and (d) the computer evaluates again, and so on.

VOLVO PROGRAM

Network Evaluation

The Volvo program for route network analysis consists basically of the following stages:

1. The route network is modeled in the core memory of the computer. The computer facilitates pathfinding by building the following lists: (a) a list of the stops passed by each route with times and distances, (b) a list of routes passing each stop, and (c) a list of transfer opportunities for each pair of routes (points at which the routes meet or part, unless otherwise specified).
2. A search is made of alternative paths for each trip (trips with a common origin and destination are processed together). The total travel time is determined for each trip. Apart from the estimated walk, wait, and transfer time (half the scheduled interval between runs), an extra (variable) penalty is added for each transfer so that trips with few transfers are given priority.
3. Unacceptable paths are eliminated. A path is considered unacceptable if the trip time remaining after boarding is longer than that by any other alternative with wait time included. In such cases, it would always pay to wait for the other alternative.
4. Of the acceptable transportation alternatives, the traveler is assumed to select either that which leaves first or that which arrives at the destination first. (These represent two alternative strategies, and the choice depends on the type of information available to the passenger.) The probability that a traveler selects a certain route is assumed to be proportional to route frequencies. The journeys are distributed onto alternative routes in proportion to these probabilities. This procedure is adopted at every transfer point.
5. Wait and transfer times are recalculated for every path; the total frequency of the runs of acceptable alternatives is taken into account. Total travel time and distance traveled are also calculated.
6. Boardings and alightings, transfers, zonal relations, histograms, and other statistics are updated.
7. The results of the evaluation are printed out.

Pathfinding, calculation of travel times, distribution on paths, and route assignment take place sequentially for each origin-destination relation so that paths need not be stored and retrieved. A detailed description of network representation, pathfinding, and route assignment is given elsewhere (1).

The following types of results are printed after each evaluation:

1. Routes and standards of service for each trip (this table can be partially or entirely suppressed);
2. Boarding passengers and loading for every route segment;
3. Flow of passengers and vehicles in the network;
4. Number of transfers and where and in which directions they occur for every pair of routes;
5. Number of boarding, alighting, and transferring passengers for every stop;
6. Number of trips and service standards for each area (wait, walk, and ride time and number of transfers)
7. Average service standard for each area for trips starting or terminating in that area;
8. Histogram for ride time and distance;
9. Summary of important characteristics for each route; and
10. Summary of service standards, traffic, loading, productivity, and costs for each mode and for the total network (Figure 1 indicates extreme values).

Interactive Graphics

Because it is difficult for the human planner to get a clear view of long tables and to identify interesting points, tables in the Volvo program are converted into figures in a graphic display. The planner specifies interactively what he or she wishes to see and can change the scales or enlarge any part of the display. The following displays are available:

1. The flow of passengers or vehicles in the network, divided up into directions (Figure 2);
2. The flow of transfers between routes at a particular stop;
3. Vehicle loading on every segment of a route;
4. Number of passengers boarding and alighting at each stop (Figure 3);
5. Number of trips between various areas (Figure 4);
6. Service standards for travel between areas (any specified combination of walk, wait, and ride times and number of transfers); and
7. Geographic reference information.

Optimization

Up to this point, the computer is used only to evaluate proposed networks (existing or planned) in relation to travel demands. It is the planner who designs or modifies solutions based on evaluation results.

In the Volvo approach the whole planning process centers around the human planner. Although, in making the planner's work easier, it is neither possible nor appropriate to fully automate the process of transportation design, a set of subproblems have been identified that can be solved automatically and can be satisfactorily controlled by the planner through the use of objectives and restrictions.

Objectives and restrictions in transportation planning are often undefined or highly controversial. So it is advantageous to have small, inexpensive optimization programs for subproblems instead of a large, overall optimization that is run only once. (An overall optimization program is also much more difficult.) If the right tools are available, the planner can afford to experiment with objectives and restrictions and, when a good solution appears to have been found, go back and analyze the new system in full detail.

Some of the optimization programs in the Volvo system are described below.

Route Connections

At each point in the network where several routes terminate or pass, it is possible to connect routes in different

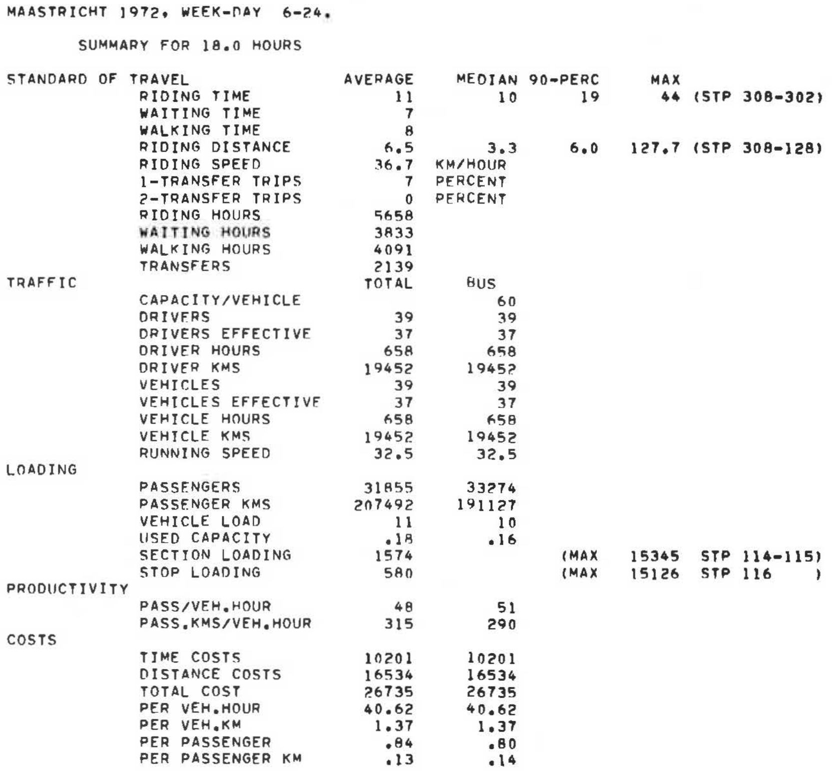


Figure 1. Summary of results from route network evaluations.

```
MAASTRICHT 1972, WEEK-DAY  6-24.

      SUMMARY FOR 18.0 HOURS

STANDARD OF TRAVEL                AVERAGE     MEDIAN 90-PERC      MAX
            RIDING TIME                11         10      19       44 (STP 308-302)
            WAITING TIME                7
            WALKING TIME                8
            RIDING DISTANCE           6.5        3.3     6.0    127.7 (STP 308-128)
            RIDING SPEED             36.7   KM/HOUR
            1-TRANSFER TRIPS            7   PERCENT
            2-TRANSFER TRIPS            0   PERCENT
            RIDING HOURS             5658
            WAITING HOURS            3833
            WALKING HOURS            4091
            TRANSFERS                2139
TRAFFIC                             TOTAL      BUS
            CAPACITY/VEHICLE                    60
            DRIVERS                    39       39
            DRIVERS EFFECTIVE          37       37
            DRIVER HOURS              658      658
            DRIVER KMS              19452    19452
            VEHICLES                   39       39
            VEHICLES EFFECTIVE         37       37
            VEHICLE HOURS             658      658
            VEHICLE KMS             19452    19452
            RUNNING SPEED            32.5     32.5
LOADING
            PASSENGERS              31855    33274
            PASSENGER KMS          207492   191127
            VEHICLE LOAD               11       10
            USED CAPACITY             .18      .16
            SECTION LOADING          1574                 (MAX   15345  STP 114-115)
            STOP LOADING              580                 (MAX   15126  STP 116    )
PRODUCTIVITY
            PASS/VEH.HOUR              48       51
            PASS.KMS/VEH.HOUR         315      290
COSTS
            TIME COSTS              10201    10201
            DISTANCE COSTS          16534    16534
            TOTAL COST              26735    26735
            PER VEH.HOUR            40.62    40.62
            PER VEH.KM               1.37     1.37
            PER PASSENGER             .84      .80
            PER PASSENGER KM          .13      .14
```

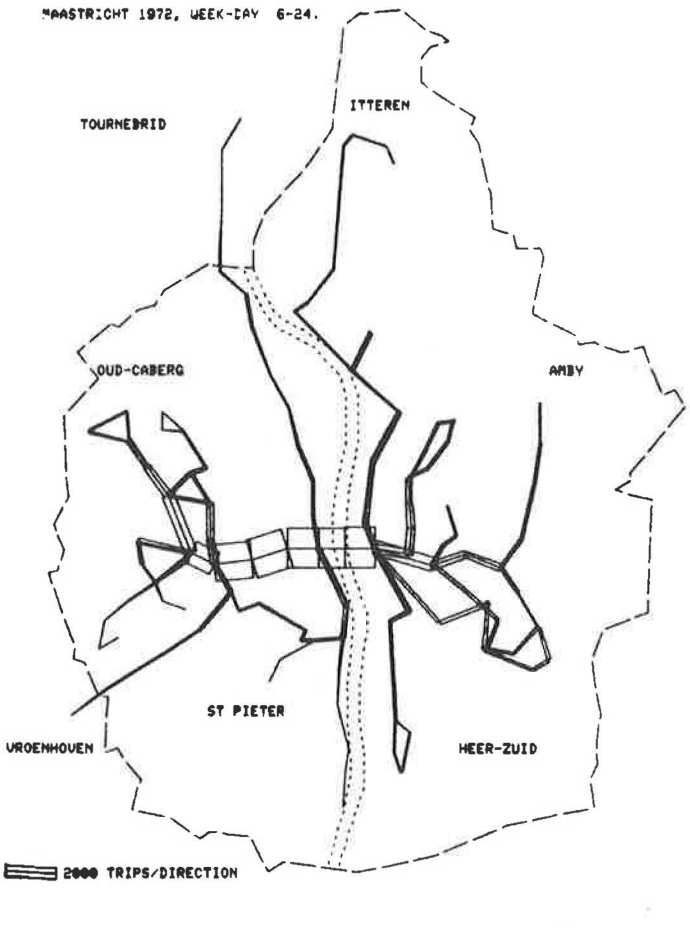


Figure 2. Passenger flows in a route network divided into two directions on each link (right-hand traffic).

Figure 3. Display of loading at stops in which the area of each circle represents the number of passengers using that stop and fractions of boardings and alightings are given by sectors.

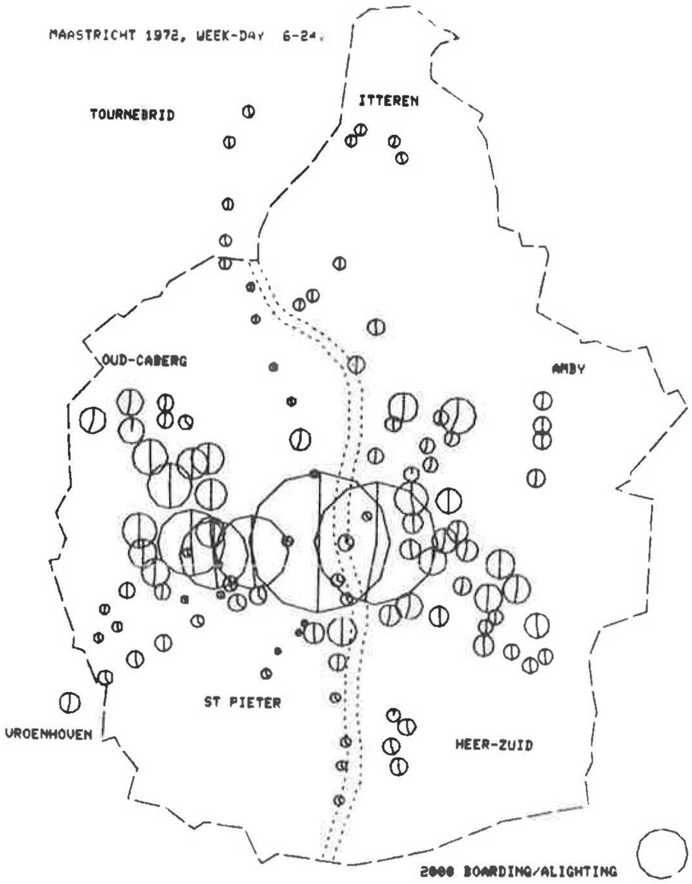


Figure 4. Display of number of trips between districts showing "desire lines".

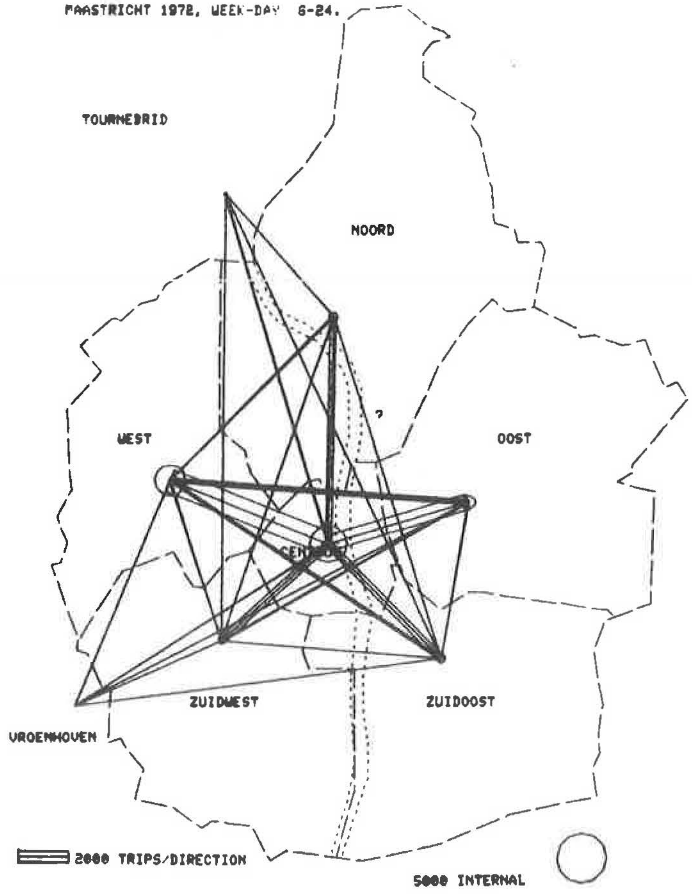
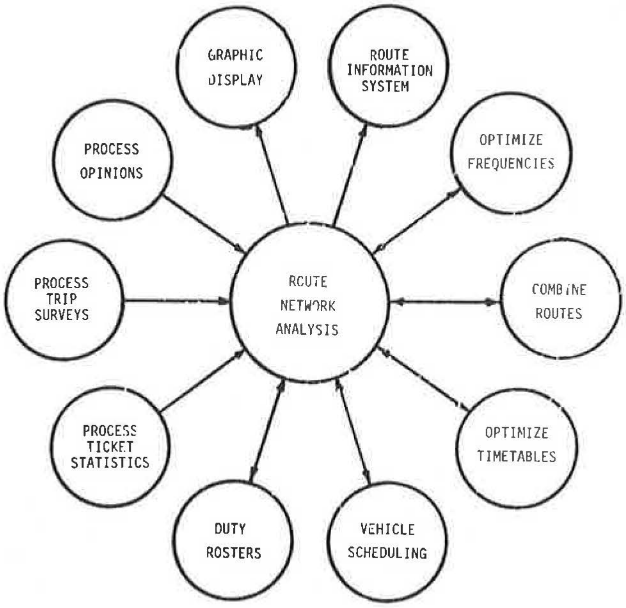


Figure 5. Dialogue optimization of frequencies and vehicle sizes on given routes.

```
BORAS 1974, VARDAG 6-9.

(0 FOR RESTRICTION IGNORED)
NO. OF VEHICLES                ? 50
NO. OF VEHICLE SIZES           ? 2
PASSENGER CAPACITIES           ? 40 80
KILOMETER COSTS                ? 2.50 4.00
VEHICLE HOUR COST              ? 35
MAX. TRAFFIC COSTS             ? 0
PASSENGER HOUR VALUE           ? 0
MAX. HEADWAY                   ? 0

RECOMMENDATIONS
ROUTE  HEADWAY  VEHICLES/SIZE  MAX.LOAD COSTS
                40    80
 52       30     .9                 34     241
  2       20    3.8                 27     856
  2X      20    3.4                 29     775
  3       15    5.3                 38    1133
  4       15          6.0           54    1604
  5       20    3.0                 40     600
  6       20    3.0                 21     623
  7      180           .2           56     143
  9       30    3.4                 40     825
  9X      30    1.5                 29     371
 11       30    2.2                 31     521
 12       60     .6                 11     204
 14       30    2.0                 14     392
 15       30    2.2                 15     540
 15X      60     .5                 11     139
 17       60     .3                 10     128
 26       30    1.7                 15     445
 30      180           .4           48     233
 31      180           .4           45     183

NEEDS           40    9                   9966

 3 MORE VEHICLES ON ROUTE  4  COST  5.42/PASS.HOUR
```

Figure 6. Volvo system of coordinated planning tools.



ways. This primarily affects the number of transfers but also the number of vehicles and the wait times. In preparation, a route network analysis is performed in which all routes are cut off at the point of interest. This is not necessary, but there would otherwise be a strong tendency to keep the existing connections because of large flows on through routes. All transfer flows are calculated from this evaluation. Then, in a dialogue manner, the planner answers questions about hourly driver cost, vehicle-kilometer operating cost, vehicle cost, passenger-hour value, transfer cost, and maximum wait time. Depending on the answers to these questions, different optimization problems are solved by minimization of total costs.

The result is a recommended connection of routes at the point of interest accompanied by data on vehicles needed, operating cost, wait time, and transfers. In case two routes with different frequencies are connected, two possible solutions are compared: Either the lower frequency portion is adjusted to equal the other, or the higher frequency portion is divided into a terminating route and a through route. The optimization algorithm, which is based on "implicit enumeration," is described elsewhere (2).

Frequencies

To solve another subproblem, the frequencies on given routes are determined so that total passenger wait time is minimized. Number of passengers and round-trip times for each route are taken from the route network analysis. The number of passengers is momentarily assumed to be independent of the frequency on each route. This approximation can be checked by rerunning the analysis with the new frequencies.

Again, the planner answers questions about available resources, costs, and time values. If more than one vehicle size is specified, the choice of vehicle type is also optimized (Figure 5). The algorithm starts from minimum frequency with the largest vehicles and improves until a resource limit is reached or until further improvements cost more than they are worth in terms of passenger time saved. As soon as loading permits, a smaller vehicle is recommended. This planning is done separately for different time periods. Vehicle use may justify using larger vehicles than necessary in off-peak hours.

Departure Times

Given routes and frequencies, the departure times on each route must be determined. The solution primarily affects transfer times. Operating costs are affected only if vehicles and drivers alternate between routes. Because all transfer flows are given by the route network analysis, it is possible to determine departure times from both terminals so that total transfer time is minimized. The idea is, of course, to synchronize departures at important transfer points. An optimization program based on heuristic search has been developed for this purpose.

Planning Tools

Volvo has developed a system of coordinated planning tools to process survey data, analyze proposed solutions, and suggest improved solutions to subproblems (Figure 6). The planner controls the application of all tools, sets objectives and restrictions, and judges the results. Every tool is not used in every application; often, in fact, computers are not used at all. The important thing is that Volvo has the resources to solve transportation planning problems—sometimes using common sense and sometimes using available planning tools or, when necessary, developing new methods.

Equipment

The planning tools described in this paper are currently running on CDC computers. The route network analysis was coded in Simula and can be converted to IBM or DEC computers. Interactive graphics are based on Tektronix display terminals and PLOT-10 software. Most of the optimization tools were programmed in interactive FORTRAN. Running times for route network analysis have typically been about 1 min in CPU time; the interactive programs give immediate results.

APPLICATION OF VOLVO SYSTEM

As of the summer of 1976, Volvo had applied its system of route network analysis for some 15 different transit operators in Europe, most of them in Sweden. The largest network to use the system is that of the Gothenburg region, which has a population of 0.7 million and 11 different transit operators. One of the problems in Gothenburg was the distribution of costs and revenues among these operators and the distribution of deficits among the municipalities concerned.

In most cases, Volvo has participated in planning origin-destination surveys on public transit. In one application, optimization of route connections reduced the number of transfers by 24 percent and at the same time reduced wait time by 12 percent and vehicle requirements by 11 percent. Optimization of departure times in another city reduced existing transfer times by 20 percent. In that city, Volvo also printed the departure times from each stop, which were to be posted at the stop.

REFERENCES

1. I. Andréasson. A Method for the Analysis of Transit Networks. Presented at the Second European Congress on Operations Research, Stockholm, Nov. 29 to Dec. 1, 1976.
2. D. Hasselström. Connecting Bus-Routes at a Point of Intersection. Presented at the Second European Congress on Operations Research, Stockholm, Nov. 29 to Dec. 1, 1976.

Computer-Animated Simulation of Taxi-Dispatching Strategies

Ezra Hauer, Ronald M. Baecker, and Paul D. Bunt, University of Toronto

A conversational, animated modeling environment implemented on a minicomputer is used to build and experiment with models of taxi-dispatching systems and strategies. The implications of each strategy are portrayed by the production and playback of an animated movie that depicts the simulated system as if it were real. The modeler can observe the system, modify the strategy, and immediately see the effects of the change. Use of the technique to assist in verifying the correctness of the simulation program, to enhance human intuition and aid strategy formulation and testing, and to facilitate technical communication about the model is described.

The primary task of the taxi dispatcher is to match taxicabs and passengers. The focal point of the dispatching process is the decision as to which taxi is to serve which call. The dispatcher's choice is made in a more or less systematic manner, governed by a set of rules that can be called a dispatching strategy. Obviously, many such strategies exist. They may be designed to (a) be simple, (b) ensure an equal share of business among taxis, (c) minimize the taxi travel distance, or (d) provide a high level of service to passengers.

Present taxi-dispatching practice has evolved over years of experience. Each company has custom tailored its procedures to a specific area and fleet size. Two elements, however, are common to most: exchange of information between the dispatcher and taxi drivers by means of radio communication and division of the service area into zones for referencing vehicle position. By using only a pencil and paper, a skilled dispatcher can oversee the operation of 100 or more taxis.

Still, it is natural to ask whether the efficiency or the service provided by the taxi industry could be improved by developing better strategies. The question cannot be answered in general terms; the answer must be in terms of the costs of implementing a dispatching strategy and expected benefits such as fewer vehicle kilometers of travel, more equitable distribution of fares, and less waiting time for passengers.

The only practical way of providing quantitative evaluations of proposed dispatching strategies is through computer simulation modeling. By using this technique, the essential features of a taxi system and the rules of a dispatching strategy are expressed as a computer program. The computer then derives the implications of adopting the strategy. There are, however, several major problems with this technique:

1. Simulation models are often so complex that there is little assurance the computer is doing what it was intended to do.
2. Simulation models are only as good as the assumptions on which they are based so that, even if the model is doing what was intended, it may have little relevance to reality.
3. Developing good strategies is difficult. The taxi dispatcher may be thought of as playing a "game" that even beginners can play. The beginner may be content to plan one move ahead and only consider a few passengers and taxis at a time, but this is not very good game-playing. To play the dispatching game well, one must think a few moves ahead, considering all pieces in the game. Not only must immediate mistakes be avoided; the pieces must also be well placed for future developments.
4. In most simulation environments, implementing and testing new strategies are difficult. Cards must be punched, complicated and error-prone job-control conventions overcome, and slow batch turnaround endured.
5. Finally, simulation modeling is a mystery to many transportation planners and practioners—a technological black art—partly because explaining its content and convincingly demonstrating its results to the uninitiated are so diffcult.

This paper describes the application of a relatively new simulation technique—conversational, animated simulation modeling—to the development of taxi-dispatching strategies. The technique helps to verify the correctness of a simulation program by visually tracing the process and thus assuring the user that the computer is doing what it was intended to do.

Animation provides the user with the ability to identify situations in which the existing dispatching strategy leads to poor-quality decisions. The strategy can then be modified and its performance tested again. In other words, the animated taxi-dispatching "game" is an efficient tool for the enhancement of intuitive understanding and for the improvement of dispatching strategies. The implementation and testing of these strategies are aided by the conversational modeling environment, which also makes possible technical communication about models among researchers and with practitioners.

CONVERSATIONAL, ANIMATED SIMULATION MODELING

The use of computer animation in simulation studies is called dynamic modeling. A dynamic modeling environment consists of a computer language for expressing simulation models and their animations and a programming system in which these programs can be written, verified for correctness, and executed. The initial dynamic modeling environment is described elsewhere (1). In the project that is the subject of this paper, a new and superior environment was used—a conversational implementation of the process-oriented simulation and animation language SLOGO (2).

Because the implementation of SLOGO is conversational, the modeler can input, revise, and test simulations directly at a computer terminal. Because the simulation language is process-oriented, models can be expressed in a natural manner, the entities in the program mirroring the entities in the situation being modeled. Thus, with a few keystrokes, the modeler can alter a dispatching strategy or change a model parameter. Seconds later, a new simulation begins to execute and the results of the changes can be observed.

SLOGO contains animation capabilities, which means that the execution of the models is visualized as an animated movie that depicts the evolution of the simulated system. Passenger origins and destinations are represented by arrows and diamonds that appear on the screen when a call is received and disappear when the passenger has been delivered to the destination. Taxis are represented by squares with flags, which are upright when the cab is free, sideways when the cab is en route to pick up a passenger, and down when the cab is delivering a passenger to a destination. As the model executes, the taxis move to their various fare origins and destinations. The movie, which is produced as a by-product of the model execution, can be viewed at the display terminal as it is produced and can later be reviewed at any speed as often as desired. Frames from the movie may be recorded on an electrostatic printer-plotter. An example of this graphical form of model documentation is shown in Figure 1 (the detailed legend shown in this figure is applicable to the other figures as well).

Simulation Verification

Programs that simulate real systems are invariably complex, and the process of developing a complex program is highly idiosyncratic. No matter how competent

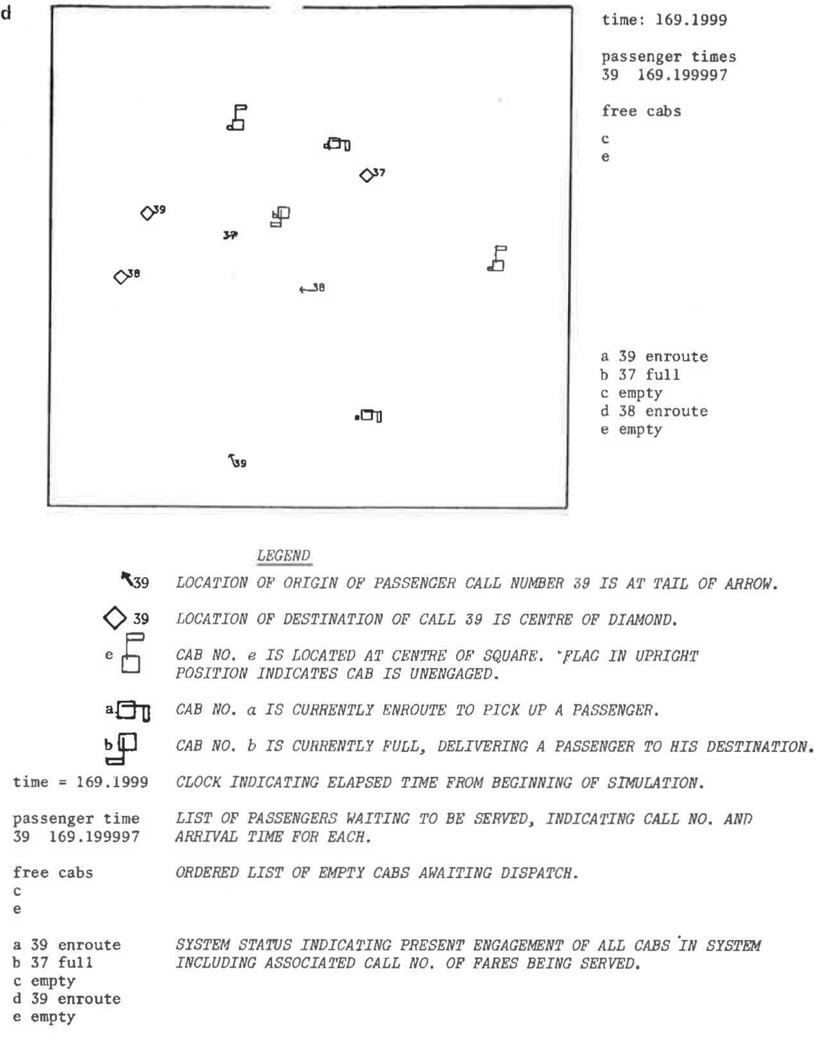


Figure 1. Electrostatic printer-plot animated model and legend.

LEGEND

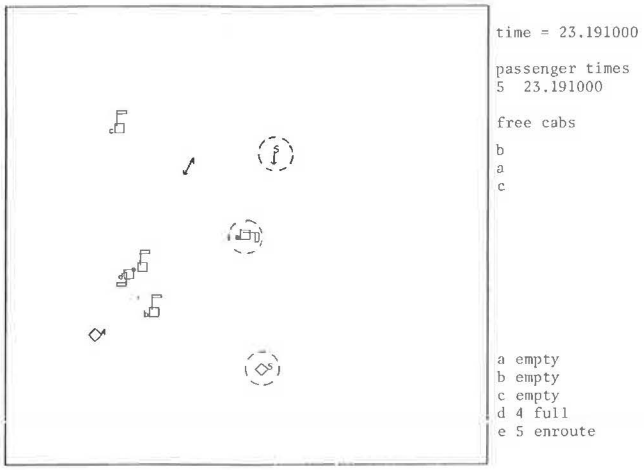
↖39	*LOCATION OF ORIGIN OF PASSENGER CALL NUMBER 39 IS AT TAIL OF ARROW.*
◇ 39	*LOCATION OF DESTINATION OF CALL 39 IS CENTRE OF DIAMOND.*
e	*CAB NO. e IS LOCATED AT CENTRE OF SQUARE. FLAG IN UPRIGHT POSITION INDICATES CAB IS UNENGAGED.*
a	*CAB NO. a IS CURRENTLY ENROUTE TO PICK UP A PASSENGER.*
b	*CAB NO. b IS CURRENTLY FULL, DELIVERING A PASSENGER TO HIS DESTINATION.*
time = 169.1999	*CLOCK INDICATING ELAPSED TIME FROM BEGINNING OF SIMULATION.*
passenger time 39 169.199997	*LIST OF PASSENGERS WAITING TO BE SERVED, INDICATING CALL NO. AND ARRIVAL TIME FOR EACH.*
free cabs c e	*ORDERED LIST OF EMPTY CABS AWAITING DISPATCH.*
a 39 enroute b 37 full c empty d 39 enroute e empty	*SYSTEM STATUS INDICATING PRESENT ENGAGEMENT OF ALL CABS IN SYSTEM INCLUDING ASSOCIATED CALL NO. OF FARES BEING SERVED.*

the programmer may be, when programs exceed a certain complexity it is difficult for the human mind to exercise strict control. Even after the program appears to be working properly, there is always a lurking suspicion that some possibilities may have been overlooked. Yet the machine will continue to churn out numbers, some of which may be wrong. They may not be completely out of the range of reason but sufficiently so to lead to erroneous conclusions. Uncertainty about the correctness of results obviously hampers the use of simulation as an analysis tool.

A variety of measures may be used to increase confidence in the correctness of the output of a simulation program. One technique is to print out intermediate results for the first few events and to examine the correctness of each step. This tedious procedure reduces the change of error, but it is time consuming, cumbersome, and does not ensure that the results will be any more correct than those that have simply been double-checked.

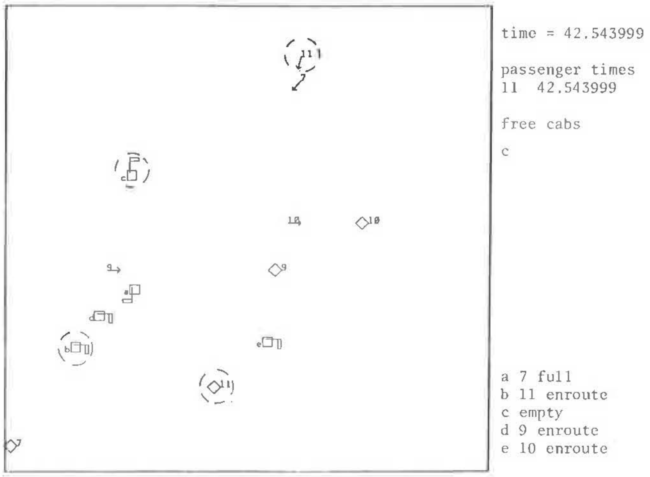
A second method is to replicate the simulation with two different programs in two different languages and then run the two models with the same sequences of random numbers. The two programs should produce identical results at each step. This was done in this project by using FORTRAN and SLOGO. In two instances, discrepancies were detected. Neither discrepancy proved to be the result of an error. Both were caused by slight differences in the assumptions and approaches of the two formulations. Uncovering and clarifying these differences led to a clearer understanding of the programs and greater confidence in the validity of the results.

Figure 2. Dispatching strategy S1.



```
time = 23.191000

passenger times
5  23.191000

free cabs

b
a
c

a empty
b empty
c empty
d 4 full
e 5 enroute
```

Figure 3. Shortcoming of strategy S1.



```
time = 42.543999

passenger times
11  42.543999

free cabs

c

a 7 full
b 11 enroute
c empty
d 9 enroute
e 10 enroute
```

A third technique is to write down in precise mathematical terms the assumptions embodied in the program and the properties of the result it is supposed to compute and then to prove that the result of the program does satisfy those properties. This technique of proving programs correct is in its infancy and has apparently never been applied to a simulation model or a program of comparable complexity.

The fourth verification method is visual process tracing, a capability unique to model animation. In principle, simulation translates the main features of a real process into computation commands and then collects and outputs numerical information about the process. Animation of this computation reproduces an image of the real process. Thus, it is now possible to confirm, through visual clues, that the real process and the simulation procedure are homologous.

In the simulation of a taxi system, the program is supposed to follow a certain dispatching rule. For instance, taxis are listed in order of priority, and calls are assigned to the first taxi on the list. When the passenger is delivered and the taxi is free again, that taxi is placed at the bottom of the list. It is essential that the simulation program follow this strategy exactly; otherwise, its results are useless. In the writing of the program, however, many things can go wrong. The typing of one incorrect letter may mean that calls are assigned to the last and not the first taxi on the list. If a command is neglected, taxis may join the list in the incorrect position. Errors of this nature are difficult to detect because the program will not abort in execution. In an animated process, such errors are immediately evident.

In a conversational, animated modeling environment, visual verification is a constant and continuous process that happens every time the researcher sits in front of the display and tries to examine the performance of a dispatching strategy.

Strategy Development

Use of the conversational, animated modeling environment enhances human intuition about the process being simulated and therefore contributes to the development of new dispatching strategies. This can be illustrated by examining the evolution of some simple dispatching strategies that use the animated model.

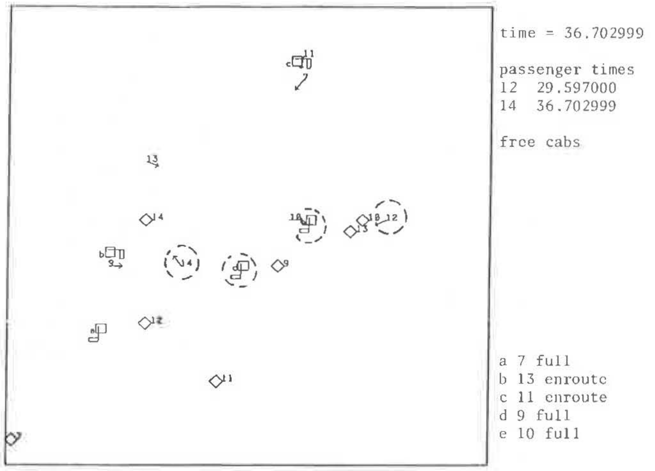
One common element in current taxi-dispatching practice is reliance on a "priority list." When a customer calls, the dispatcher assigns the call to the taxi at the top of the list. The same taxi is placed at the bottom of the list when it delivers its fare. Thus, a natural rotation and a demonstrably equitable distribution of fares among taxis are established. This strategy (S1) was selected to be animated first to serve as a benchmark against which other strategies could be compared.

Once the program was debugged and its correctness was verified, the performance of strategy S1 could be examined. Figure 2 shows the state of affairs at simulation time = 23.2 min. The list of free taxis appears in order of priority in the upper right corner. Thus, the next call to come in will be assigned to taxi b. At this point taxi d is carrying passenger 4 from origin (arrow labeled 4) to destination (diamond labeled 4). (As explained in the legend of Figure 1, the table in the lower right corner lists the status of each taxi and the passenger to whom it is assigned.) Passenger 5 has just called in. His or her origin and destination are given by the arrow and diamond labeled 5. Taxi e, which was on top of the priority list, has just been assigned to the call and its flag lowered to indicate that it is now en route. This assignment seems to be appropriate because taxi e happened to be the free taxi closest to the origin of passenger 5.

It takes only a few more frames to identify a situation in which adherence to the priority list is not in the best interest of either the taxi company or the traveling public. Figure 3 shows the situation at the instant at which passenger 11 calls in. At this time taxicabs b and c were both available, and b had priority. The call has thus been assigned to b, which is now en route to pick up the fare. But taxicab c was obviously in a better position to serve the call. Assignment of taxi c to passenger 11 would have resulted in shorter travel distance for the taxi and less waiting time for the passenger.

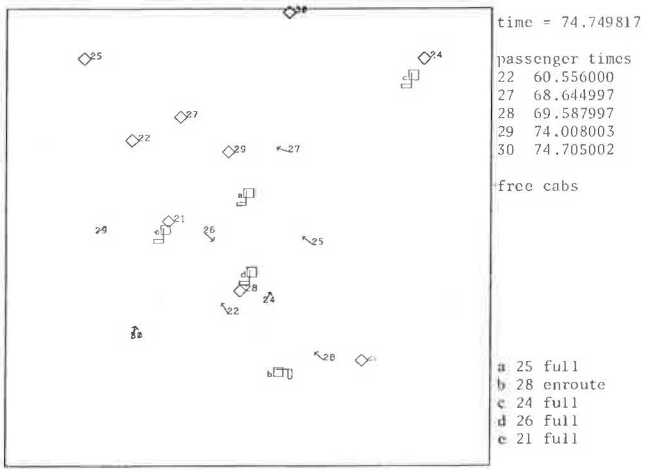
The identified deficiency of priority dispatching occurs only if more than one taxicab is available for service when a call comes in. Thus, significant improvements may be hoped for only when the demand is less than capacity. If only two taxis are free when the call comes in, the simple priority strategy will make the correct assignment in half of the cases; if three taxis are free, the right assignment will be made in one-third of the cases. Thus, the larger the number of idle taxis is at a given instant, the less desirable strategy S1 appears to be.

Figure 4. Shortcoming of strategy S2.



```
time = 36.702999

passenger times
12  29.597000
14  36.702999

free cabs

a 7 full
b 13 enroute
c 11 enroute
d 9 full
e 10 full
```

Figure 5. Shortcoming of strategy S3.



```
time = 74.749817

passenger times
22  60.556000
27  68.644997
28  69.587997
29  74.008003
30  74.705002

free cabs

a 25 full
b 28 enroute
c 24 full
d 26 full
e 21 full
```

In the conversational, animated modeling environment, modification of dispatching strategies is often a simple task. Changing only a few lines in the dispatcher module of the program results in the derivation of S2, a strategy that always assigns the closest free taxi to a call. Moments later, the new model begins to execute. Now taxi c and not b will pick up fare 11. All goes well with S2 until a temporary shortage of free taxis occurs, that is, until the time at which more than one passenger is waiting when a taxi becomes free.

If strategy S2 is used, the taxi that just became free is ordinarily assigned to pick up the passenger who called in first. Thus, in Figure 4, taxi d is about to deliver passenger 9 and become free. At that point in time, passengers 12 and 14 will be waiting. Because passenger 12 called in some 7 min before passenger 14, taxi d would be assigned to serve passenger 12. But the origin of passenger 14 is closer to the location at which taxi d will become free. Furthermore, taxi e is also about to become free and could be assigned to passenger 12. So it seems beneficial to assign taxi d to the service of passenger 14 and taxi e to pick up passenger 12. This switch will reduce the distance traveled en route for both taxis. It will somewhat increase the waiting time of passenger 12 but reduce waiting time for passenger 14.

Strategy S2, therefore, needs to be improved to deal more effectively with congestion. This results in the derivation of strategy S3, which will do for passengers what S2 did for taxicabs, that is, consider which of the waiting passengers is closest to the taxi that is about to become available.

Strategy S3 completely abandons the concept of a priority list. It is based on the following two rules:

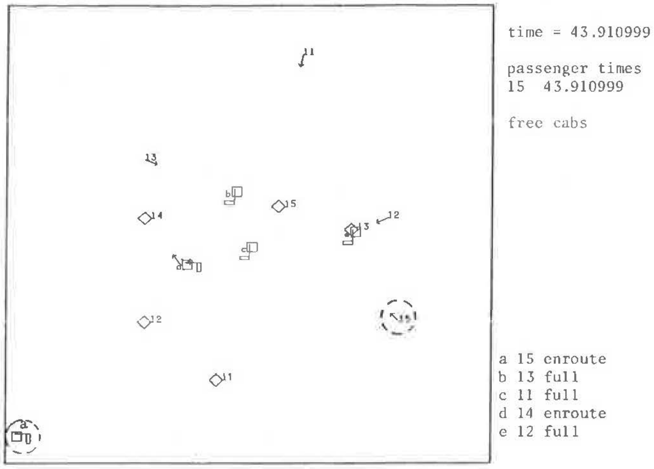
1. If more than one taxi is free when a call comes in, assign the call to the taxi closest to it.
2. If more than one passenger is waiting to be assigned when a taxi becomes free, assign the taxi to the closest call.

Strategy S3, however, is not perfect either. One of its deficiencies can be vividly demonstrated by means of animation. In Figure 5, passenger 22, who called in 14 min earlier, is still waiting to be assigned to a taxi in spite of the fact that five of the passengers who called in later have already been assigned and some have even been delivered to their destinations. This inequity is a result of the use of the closest taxi-closest passenger rule. Calls 23 to 26 effectively screened passenger 22 from being assigned to taxicabs as they became free. The screening process may happen in central as well as outlying areas.

Several devices may be used to increase the equity of strategy S3 for both passengers and taxi operators. For example, one can limit maximum waiting time or the number of deviations by using a background priority list. Strategy S3 also suffers from other shortcomings. For example, only the free taxis are considered in making the assignments. Anticipating the time at which additional taxis will become free could probably eliminate some inefficiencies.

An example of this is shown in Figure 6, where taxicab a was the only free taxi and was thus assigned to a new call from passenger 15. It would probably have been better to dispatch taxi b to pick up passenger 15 after it had delivered passenger 13. Figure 6 also shows another deficiency of strategy S3. Taxicab a is clearly in no position to serve future calls. Its next dispatch order should thus be formulated so that the overall constellation of taxicabs is improved with respect to anticipated demand. One way in which this problem is alleviated in practice is by dividing a city into zones or areas and circulating taxis only within these areas. Dispatching systems based on zones could also be portrayed and examined by using the animation technique.

Figure 6. Failure to anticipate future developments (strategy S3).



The intuitive concepts gained by use of animation need to be tested quantitatively in order to evaluate the relative merits of different dispatching strategies. Statistics were collected for performance measures such as passenger waiting time, taxi revenue, and taxi load sharing under a variety of demand rates. Deficiencies and strengths of each strategy that were observed in the animation are supported statistically (Tables 1, 2, and 3). For example, at low demand rates, the mean passenger waiting time is reduced from 7.5 to 4.5 min by switching from strategy S1 to strategy S2.

Communication About Models

The animation of a simulation helps the model builder to communicate with those who have an interest in the model's operation. Although conventional simulation models produce pages and pages of individual and aggregate data, it is difficult to explain how the models actually work in terms of the output they produce. Through visual tracing, animation provides a simple yet effective method of demonstrating the internal workings of the model. Animation can thus become a vehicle for effective communication with those who are totally unfamiliar with simulation models. In this project, for example, a taxi broker and one of his dispatchers were invited in for consultation. The visual display made it very easy to communicate about the model, and the resultant exchange of ideas had several beneficial effects. The practitioners were given some ideas about how current dispatching strategies might be improved. The modeler obtained further insight into the nature of the dispatching game. A valuable communication channel was opened by means of which undocumented real-world expertise was made available.

The relative ease with which the model could be revised also proved beneficial in the demonstration of the model for the practitioners. Many inevitable "what if" questions can be answered quickly by making simple modifications to the model. Conversational, animated simulation modeling is also a valuable tool for improv-

Table 1. Statistical results of three model runs for passenger level of service.

	Dispatching Strategy		
Item	1	2	3
Demand rate, calls/h/taxi	4.8	4.8	4.8
	2.4	2.4	2.4
	0.6	0.6	0.6
Mean passenger wait time, min	159.6[a]	159.6[a]	19.4
	9.5	7.6	7.0
	7.5	4.5	4.5
Maximum value of passenger wait time, min	312.5[a]	312.5[a]	100.8
	36.3	34.2	32.0
	18.4	19.7	19.7
Standard deviation of passenger wait time, min	84.8[a]	84.8[a]	18.5
	5.8	5.3	4.6
	3.8	2.8	2.8

[a] Demand rate over system capacity.

Table 2. Statistical results of three model runs for overall system performance.

	Dispatching Strategy		
Item	1	2	3
Demand rate, calls/h/taxi	4.8	4.8	4.8
	2.4	2.4	2.4
	0.6	0.6	0.6
Total kilometers traveled	4834	4834	3731
	4912	4477	4403
	4861	3901	3901
Percentage of kilometers of travel in which taxis were engaged[a]	51	51	66
	50	55	56
	51	63	63
Revenue kilometers per taxi hour	24.2	24.2	31.5
	16.0	16.0	16.0
	4.0	4.0	4.0
System load factor[b]	1.20	1.20	0.93
	0.61	0.55	0.54
	0.15	0.12	0.12

Note: 1 km = 0.62 mile.

[a] Percentage of total kilometers traveled with the meter turned on.

[b] Ratio of demand rate to service rate. A value of >1.0 indicates the system is overloaded.

Table 3. Statistical results of three model runs for business-sharing equity.

	Dispatching Strategy		
Item	1	2	3
Demand rate, calls/h/taxi	4.8	4.8	4.8
	2.4	2.4	2.4
	0.6	0.6	0.6
Average number of fares per taxi	80	80	80
	80	80	80
	80	80	80
Number of fares, maximum/minimum[a]	83/77	83/77	83/78
	82/77	85/77	87/72
	81/79	95/67	95/65
Kilometers full			
Average per taxi	493	493	493
	493	493	493
	493	493	493
Maximum/minimum[a]	315/285	315/285	351/301
	318/299	335/278	329/275
	322/292	351/265	315/265
Average kilometers en route per taxi	478	478	253
	490	403	389
	480	288	288

Note: 1 km = 0.62 mile.

[a] Maximum divided by minimum values give range to indicate distribution.

ing technical communication among a problem-solving group of scientists, engineers, and planners.

COST-EFFECTIVENESS OF THE TECHNIQUE

Is the animation technique cost-effective? It is difficult to quantify the value of the detection of a subtle bug in a simulation, of the extra insight into model behavior that may or may not have been obtained without this technique, or of more effective communication within the research community or with the public. However, some rough cost estimates can be made based on the experience gained in this project.

SLOGO currently runs on a medium-scale minicomputer, a PDP 11/45 that has a minimum of 48K words of memory. Sustained use requires that the system have one disk pack unit for movie storage and playback. The total equipment configuration is valued at \$150 000, which includes a \$30 000 graphics terminal on which the movies are viewed. The UNIX time-shared operating system used by SLOGO costs \$20 000 and is commercially available. SLOGO and its underlying graphics support package (3) took roughly 2 person years to design and implement.

If such a system were implemented again in a commercial environment in the late 1970s, the hardware would cost about \$100 000. The UNIX software would still cost \$20 000. A new version of SLOGO with integrated graphics capabilities could be constructed in a year by a very skilled systems programmer. A workable installation would probably sell modeling time at the graphics console for between \$20 and \$50/h. A planner could experiment with several runs of the model in an hour at the terminal.

It is also possible to use the system remotely. SLOGO drives a Tektronix direct-view storage tube over phone lines. Images are painted slowly, depending on the bandwidth of the phone line, and animations must be viewed by superimposing multiple frames and then clearing the screen and starting over. Although this technique is awkward, it is adequate for the initial stages of model construction and debugging. A planner located in Montreal, for example, could carry out most of his or her work remotely and only travel to Toronto for the strategy exploration production runs.

SUMMARY AND CONCLUSIONS

The pilot project described here, in which computer-animated simulation modeling has been applied to the investigation of strategies for taxi dispatching, provided evidence that the technique is valuable for verifying the correctness of simulation models, for augmenting researchers' intuition about the effects of various strategies, and for improving technical communication about models among researchers and with practitioners. This is particularly important when the system is complex for it then becomes increasingly difficult to guarantee the absence of bugs in the model, to predict all the effects of presumed improvements in strategy, and to discuss them without recourse to appropriate still and moving pictures.

Figures 1 through 6 show only one kind of picture by which the simulation can be visualized. Many other kinds of graphical representations of model behavior could be produced during or after execution of the model. Distributions of waiting times and histograms of taxi distance traveled can be displayed and updated as each request is serviced. Means and variances can be computed and displayed in relation to corresponding values for other strategies. Such pictures are easily computed in SLOGO.

The major problem encountered was the need for the user to learn a new language. Planners with an engineering education generally learn only one programming language—typically FORTRAN—whereas researchers developing new simulation and animation systems regard FORTRAN as archaic and awkward and base their new systems on modern computer languages. There are only two solutions: The planner must either (a) invest several weeks of intensive effort in becoming fluent in the new langauge or (b) work through an "interpreter" who translates the planner's intentions into programs.

In conclusion, the results reported here, in an accompanying film (4), and in the earlier pilot study (1) seem sufficiently promising to warrant further work. In spite of the numerous shortcomings of SLOGO, many of which are too technical for this report and will be dealt with in the design of a subsequent simulation and animation system, the results justify the undertaking of a longer range and larger scale use of this technique on some significant transportation planning problem.

ACKNOWLEDGMENTS

Tom Duff developed SLOGO and did the bulk of the SLOGO programming. The Transportation Development Agency and the National Research Council of Canada provided financial support. The views expressed are ours.

REFERENCES

1. R. M. Baecker and T. R. Horsley. Computer-Animated Simulation Models: A Tool for Transportation Planning. TRB, Transportation Research Record 557, 1975, pp. 33-44.
2. T. D. S. Duff. Simulation and Animation. Department of Computer Science, Univ. of Toronto, MSc thesis, 1976.
3. W. T. Reeves. A Device-Independent, General-Purpose Graphics System in a Minicomputer Time-Sharing Environment. Department of Computer Science, Univ. of Toronto, MSc thesis, 1976.
4. R. M. Baecker, G. Hill, and M. Tuori. Computer Animated Simulation of Taxi Dispatching Strategies. Dynamic Research Group, Univ. of Toronto, 10-min, 16-mm, black-and-white, sound film, 1976.

Computer Graphics Human-Figure System Applicable to Transportation

William A. Fetter, Southern Illinois University

A study of improved graphic representation that has usefulness for transportation research is presented. Applications that parallel the transportation design process in the variety of levels of detail required are emphasized. For example, a human figure represented by a single point can be useful in overview plots of population density and consumer areas. A crude 10-point figure can be applied to studies of queuing theory and the simulated movement of groups. A 100-point figure can be animated to scale in a design showing different overall body shapes, including male and female figures. A 1000-point figure, similarly animated, can be used in anthropometrics for work-station designs, gross body movements, and occupant motion in vehicle crashes. Extrapolations of this order-of-magnitude approach should ultimately result in very complex data bases and a program that automatically selects the correct level of detail.

The need exists in transportation research to clearly display information about the human figure—information ranging from large numbers for population and other demographic studies to single figures for studies of occupant safety, all accessible in a single system (6). Many computer simulations and modeling systems are under development. Program output is often already in machine-readable form that can drive improved graphic displays. The ability to relate different research results may be simplified and made more coherent by adopting some graphic conventions of the human figure (3).

Figure 1 shows the limited graphic quality of some research software (3, 5). Quality improvements in graphics are sometimes deferred in current research because the specialists must concentrate on the functional aspects of their particular fields of study. When graphic output is used, the means of achieving an improved display are not always readily available. In addition to animation capability, management of graphic parameters is needed so that end products have quality and consistency (5).

The wider communication capability required to reach specialists in rapidly subdividing technologies often calls for reliance on graphics to help translate or bridge specialized jargon. There is a relatively new, growing need to communicate simultaneously with other disciplines and with the public, which also suggests more reliance on graphics. A consistent conceptual framework applicable to transportation research is needed in which data may be accessed and displayed in controlled degrees of detail.

HUMAN-FIGURE COMPUTER MODELING

Problem

The problem is how to develop a computer graphics system, for independent use or for use in work by other researchers, which is applicable to an array of man-machine research and application purposes including transportation. Development of a comprehensive program and anthropometric data bases must ensure a widely used language in a straightforward data format. The program must be reliable and provide relatively easy manipulation of three dimensionally defined data in a number of basic projections including map projections. The data must be developed to accommodate many levels of detail and, ultimately, easily calculated costs.

The system capabilities must be complete enough for relatively independent direct applications. A transportation-related example is population mapping of a mass transit system, enlarged to figures to display queuing through a station, then enlarged again to indicate space requirements in vehicles by means of many figures occupying a single car. Detailed studies of car occupants during decelerations should also be possible.

The system should be relatively convenient for display of other ongoing research. For example, existing data on spatial needs related to anthropometry should be assimilated and graphically portrayed. Another use is the translation of existing biomechanical dynamic models into more easily communicated, animated predictions.

Research

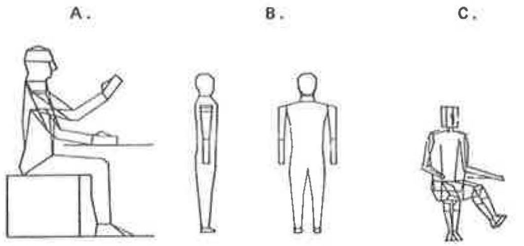
Research in computer models of the human figure is often principally involved in functional aspects of biomechanical modeling in which only modest attention is paid to display; such work is addressed to a very specific research area or relies on the capabilities of specific computer hardware-software systems.

Some of the work that has emphasized body functions includes that of Springer and Katz of the Boeing Company's personnel subsystems development for the Joint Army-Navy-Air Instrumentation Research (JANAIR) program (Figure 2). Other extensions of the Boeing work are proceeding at the Naval Air Development Center. Computer models have been developed by, among others, Kilpatrick (3) and Reed and Garrett (5).

Work that addresses very specific research areas is exemplified by studies at the University of Utah and the Naval Aeromedical Research Laboratory. Parke (4) has developed advanced specialized representations of the human figure; equations have also been developed that express possible variations in leg movements. Naval researchers have done voluminous studies of the biomechanical effects of head-neck motion.

Much of the work described in the examples above relies on widely available computer systems—the University of Utah being at least one notable exception. In many cases, plotters, a growing number of which access computer output microfilm (COM) or interactive cathode ray tube (CRT) displays, are the extent of graphic communication hardware. It appears that the system under development could be a useful aid to a number of these systems in communicating research results.

Figure 1. Limited graphic quality of some research software.



THE STUDY

Background

Earlier developments by Fetter in human-figure computer graphics occurred in an industrial setting at the Boeing Company where it was important to produce meaningful anthropometric contributions to the design process, to reduce such research to profitable practice, and to bring about reliable results and multipurpose graphics output (2). An eight-step system was developed to integrate graphic and technical qualities:

1. Aims definition to set goals,
2. Communication design to provide early visualization of the end product,
3. Data transcription to encode data,
4. Computing-programming to manipulate the data,
5. Automatic drawing to produce the essential image,
6. Final rendering to complete a master copy,
7. Reproduction to produce further copies, and
8. Distribution to ensure the communication reached the audience.

Figure 2. Sample of graphics for a joint military study.

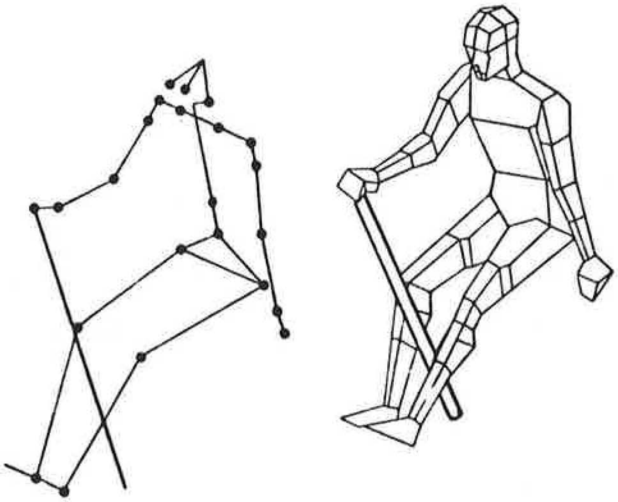


Figure 3. Seven-system man-in-cockpit studies emphasizing graphic modeling.

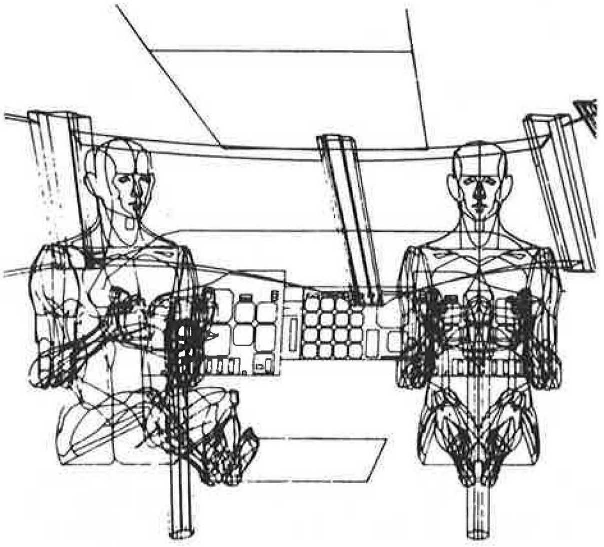
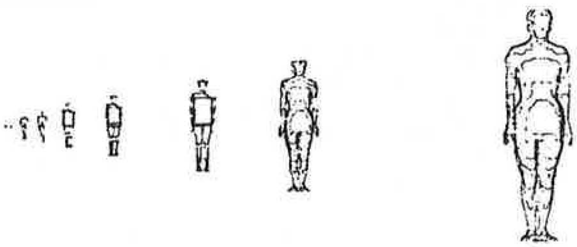


Figure 4. Human-figure data for a range of number of points per figure.



The visual module system permits one image to be immediately useful in a number of communication media without costly reworking. The most common media in use have been document illustrations, presentation charts, 35-mm slides, 16-mm motion pictures, closed-circuit television, and interactive CRT displays. This system has demonstrated its usefulness in industry.

Human-figure work originally centered around male, variable-percentile representations of airline pilots. A rendition of a landing signal officer was used in early carrier landing simulations. Figure 3 shows the first articulated model, known as the "first man," which involved the use of seven systems. The "second man" (using 19 systems) was then developed for more complete articulation and applied to a number of aircraft design projects and cockpit evaluation studies.

Objectives

The objectives of this study are to resolve display problems in a cost-effective, reliable system. The quality of the graphic representation of each human figure is limited in number of data points but also controlled as to size to ensure appropriate levels of completeness and finish in each representation. The visual module system is the prime mover of information, controlling visual angle, line weight, size of images, and other graphic standards for use in any media.

To achieve a number of levels of detail, a system of several data subbases, each of which increases by one order of magnitude in number of points, is under development (Figure 4). A 1-point figure can be used for statistical mapping; a 10-point figure for large crowds for, say, queuing-theory studies; a 100-point figure for gross bodily motion; a 1000-point figure for detailed motion; and so on. Ultimately, the choice of data bases will be made automatically within the system based on user requirements.

Demonstration applications under way in the field of design include population as one aspect of transportation information display. Anthropometric applications can include studies of a single-percentile comfort analysis of designed artifacts such as interior dimensions of a vehicle. [Anthropometric data used in this research are drawn from those of Diffrient, Tilley, and Bardagjy (1) (Figure 5).] Applications include display with more complicated systemic details.

Integration

By integrating the eight-step approach, including the program VIEWIT, and the visual module and the order-of-magnitude human-figure data, the following results have been achieved. [All figures are based on the 50th-percentile data of Diffrient, Tilley, and Bardagjy (1) unless otherwise noted.]

The 1-point man and woman are single points defined

Figure 5. Sample of anthropometric data now in use.

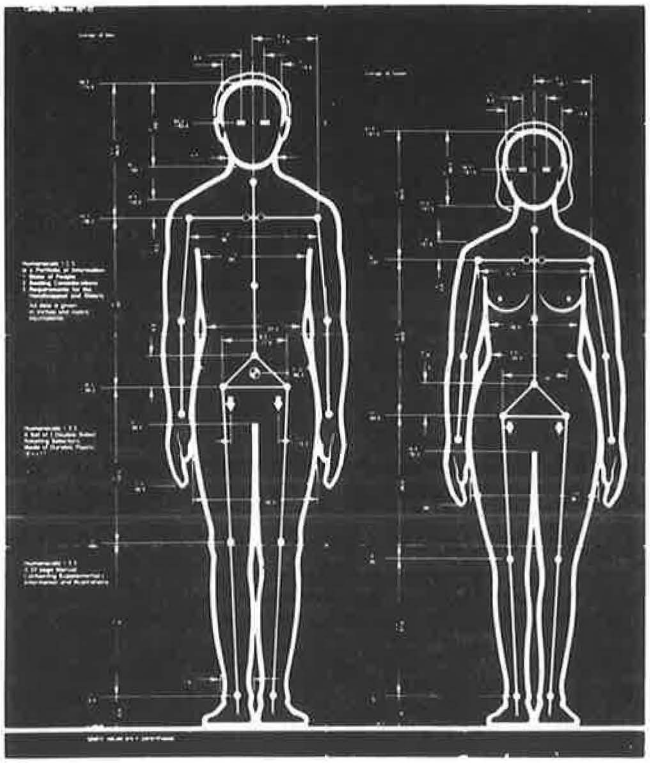


Figure 6. Male and female 1000-point figures showing motion in the male.

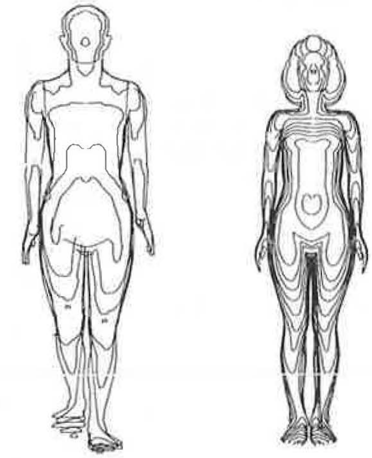


Figure 7. Frame from animated film showing human figure in a monorail design simulation.

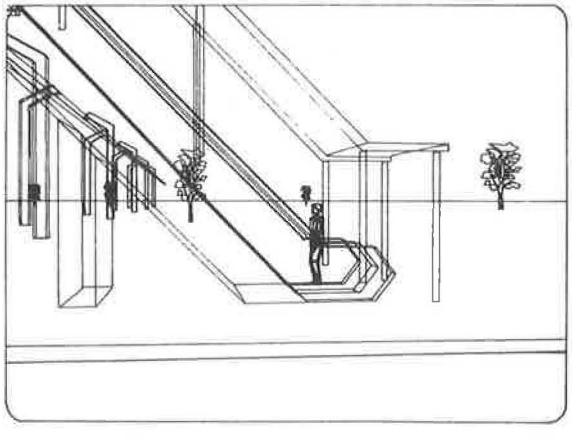
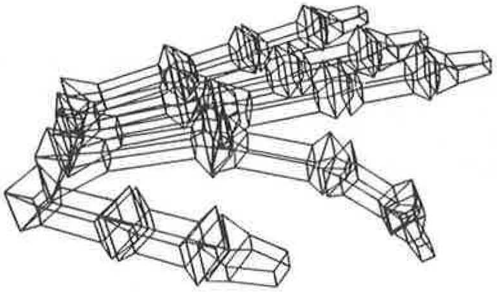


Figure 8. Skeletal hand with a high number of data points.



in three-dimensional coordinates and of a height consistent with 50th-percentile center of gravity depending on the application. Because of the distances viewed, the single point represents all figure elements.

The 10-point figures become stick-figure lines simultaneously representing the bones and the surface of the body—again, because of the thickness of lines needed in the small sizes to be used. This will give way to a polygonal form in the new data bases.

The 100-point figure has a system of 10 moveable linkages in which the data represent the exterior surface only:

1. Torso,
2. Head,
3. And,
4. Upper arms,
5. And,
6. Lower arms and hands,
7. And,
8. Upper legs,
9. And,
10. Lower legs and feet.

New polygonal data will contain surfaces and five systems.

The 1000-point figure (Figure 6) has a 10-linkage system that may be articulated. The data describe only the exterior surface, which is composed of points from a biostereometric data representation of the surface. The figures shown here are based completely on hand-drawn biostereometric contours, but tapes made available by Herron of Baylor University should allow the point definition of surfaces to be converted more readily to this system for animation with greater accuracy and ease of data transcription.

Demonstration

Demonstrations of potential applications of this system to date have included a seated figure, representation of figure scales in a monorail design simulation (Figure 7), and a motion simulation of a skeletal hand (Figure 8). In two cases, the 100-point figure has been used in exercises showing gross body motions. The process of seating the figure in a chair demonstrated the potential testing of a chair design against multiple-percentile male and female figures.

A slide and film simulation of a contemplated monorail system at Southern Illinois University permitted the inclusion of human scale in the presentation (Figure 7). The animation of a skeletal hand (Figure 8) demonstrates a biomedical training example of a figure that will contain 10 000 data points. The skeleton of the hand was simplified to the basic representative forms and data transcribed. Then the limits of motion and the hierarchical array of motion were programmed for an animation sequence demonstrating the hand motions.

Application

An important part of this effort is the constant growth of the interface between the generalized system capability and potential users. Promising areas of application are being studied to determine the best means of collecting and coding data. The completed media will be shown to others interested in this field of research so that feedback on the usefulness of the system can be obtained.

ACKNOWLEDGMENTS

This effort in extended human-figure computer graphics could not have been carried out without the support of the National Science Foundation and the following persons at Southern Illinois University: Stanley Smith, dean of the College of Human Resources; Michael Dingerson, director of the Research and Projects Office; and John Lonergan, chairman of the Department of Design. The computer programs used have been developed over a period of years under my direction by assistants in design, graduate assistants, student workers, and students who have incorporated their work into an overall unified software system. These include Dennis Cagle, John Buyer, William Shaw, Frank Juzwik, Albert Allen, Lewis Wright, Dennis Andrews, Joe Tumminaro, Craig Schilhahn, Lawrence Evans, and Frank Crow. Students who contributed directly to the system application discussed in this paper are Robert Charneski, Richard Rovens, Philip Hoekstra, Todd Nickle, and John Virruso.

REFERENCES

1. N. Diffrient, A. R. Tilley, and J. C. Bardagjy. Humanscale 1/2/3/. MIT Press, Cambridge, MA, Nov. 1974.
2. W. A. Fetter. Communication. McGraw-Hill, New York, 1964.
3. K. W. Kilpatrick. A Biolinematic Model for Workplace Design. Human Factors, 1972, pp. 237-247.
4. F. I. Parke. Computer Generated Animation of Faces. Univ. of Utah, Salt Lake City, Computer Science Technical Rept. UTEC-CSc-72-120, June 1972.
5. W. S. Reed and G. E. Garrett. A Three Dimensional Human Form and Motion Simulation. Kinesiology Review, 1971, pp. 60-65.
6. A. Roozbazar. Biomechanical Modelling of the Human Body. Presented at National Human Factors Society Meeting, Arlington, VA, Oct. 16, 1973.