HIGH CAPACITY BUS SYSTEMS: A DESIGN METHODOLOGY USING ARTIFICIAL INTELLIGENCE

Nicholas Andrew Tyler

A thesis submitted to the University of London for the degree of Doctor of Philosophy

Transport Studies Group
University College London

November 1991

To Katrina

ABSTRACT

Buses form the basis of the public transport system in many cities of the world: improvements to the operation of bus systems could therefore provide useful benefits in a variety of circumstances. In some cases (notably in Brazil) enhancements to the bus system have been achieved through the design of infrastructure and operations.

This study examines both the nature of these enhancements to bus systems and the methods by which they were achieved, in order to facilitate the use of such techniques to improve bus systems elsewhere. Detailed discussions with engineers responsible for existing Brazilian high capacity bus systems reveal how these were conceived and put into operation. In addition, they indicate important variables and models (existing or new) that could be used during this process.

In order to distinguish between results obtained on the basis of good quality data and those depending on data of poorer quality, existing artificial intelligence methods are investigated for their performance with poor quality data. These methods may provide some interpretative element, but they do not appear to be appropriate for the representation of the engineer's ability to interpret data. In order to include such assessments, a model is presented which represents the use by the engineer of opinion and judgement in the decision process.

The design methodology for high capacity bus systems is represented in a computer model and validated with data representing the situation in a corridor before the implementation of a high capacity bus system. The resulting design is considered to be similar to the design actually implemented. The model is also tested in two other situations where the constraints are more severe, in Perú and in London. The results suggest that it is not possible to obtain a satisfactory design for a high capacity bus system in these corridors.

ACKNOWLEDGEMENTS

In the course of this research, many people have given assistance, and it would be impractical to mention them all by name and to thank them for their various contributions. I would like to acknowledge the help of a few people without whose assistance the work would have been impossible.

I would like to thank Dr L.G. Willumsen, who inspired this work and supervised the research during the time when the majority of the exploratory work was undertaken. His guidance and encouragement have been invaluable during the whole project.

I also wish to thank Dr R.L. Mackett for his encouragement and supervision of the project in its later stages.

Of the various engineers consulted during this research, I would like to name two who have been of considerable help: Pedro Alvaro Szász and Jaime Gibson.

I would also like to express my appreciation to:

> INVERMET and the Royal Borough of Kingston-upon-Thames for the data used as a basis for the experiments,
>
> The Rees-Jeffreys Road Fund for their financial support, and
>
> all my various friends in the various countries visited during this research for their assistance, help and friendship, and my colleagues at the Transport Studies Group who have been so supportive during this research.

Finally, my thanks go to Katrina, who has encouraged, criticized, helped, cajoled and who has been such a support during this research.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

1.1 GENERAL CONTEXT

Human mobility is an essential part of civilized life, and public transport is one way in which mobility can be made available. Where the demand for transport is high, however, the public transport system may often be operating near to (or in excess of) its safe operating capacity. Expansion of this capacity is often expensive and time consuming. There is therefore a need to explore ways in which the capacity of the public transport system might be expanded cheaply and quickly. One form of such a capacity enhancement, the expansion of bus capacity, is considered to be a useful example to study in this context because it represents an expansion of existing (and low-cost) technology.

An important characteristic of the design of infrastructure and operating systems in the transport context is the uniqueness of the conditions and circumstances which apply at a particular site. Transport engineers need to adapt their basic methods so that they are appropriate for each implementation. It is therefore important to consider not only the techniques used to enhance bus capacity, but also the decision processes which underlie the design of these systems.

It is useful to examine how these decision processes may be modelled in order to facilitate the transfer of these methods to new locations. In so doing, it is to be hoped that such techniques may be made available for use by transport planners in many different contexts around the world, and thereby to increase the level of personal mobility in urban areas.

1.2 OBJECTIVES

The relationships between the elements of the bus system (for example, passengers, vehicles, bus stops and other infrastructure) and its capacity should be understood before attempting to determine how this capacity might be increased. To explore these relationships, bus systems must be examined when they are operating near to, and beyond, their capacity. Benefits arise from such analyses in two ways: first, the capacity can be established and critical points identified where this is insufficient for demand. Second, the understanding obtained from such analyses may encourage better design methods to be developed so that more efficient operation results, whether or not the system is operating near to its capacity. The first objective of this thesis is therefore to discover the methods and techniques used to design high capacity bus systems, and to analyze the theoretical basis upon which they are based.

This thesis also considers the decision processes used in the design of high capacity bus systems. These processes include not only the methods, models and techniques used during the design process, but also the way in which decisions are taken. Decisions are often taken in situations where the suitable data are not available, and therefore it is important to consider how decisions are taken in these circumstances.

This is of particular importance where these decisions are taken with respect to technology which exists in one environment but not in another. In such cases it may be useful to transfer the technology between the environments. Technology transfer, however, is only desirable if the technology concerned is appropriate for the new environment. This often means that some form of adaptation is required in order to implement the technology successfully. It is therefore of considerable importance to consider how and why this adaptation is chosen, because this may facilitate better technology transfer in the future.

Before this is possible, however, it is necessary to consider how decisions are taken in the course of the design and implementation of the technology concerned. This may lead to the possibility that the decision process itself may be of considerable importance in the technology transfer process. A particular issue is the extent to which opinion and judgement are used within the decision process. To do this, one approach is to develop a model of the decision process so that the most important aspects can be analyzed.

The objectives of this research are:-

to investigate the methods and techniques used to design high capacity bus systems;

to investigate the methods used to automate decision processes;

to develop a model which includes the use of judgement and opinion in the decision process; and

to use this model within a design model for high capacity bus systems.

1.3 THE STRUCTURE OF THE STUDY

This study has involved discussions with engineers who have designed and implemented such systems in order to ascertain the knowledge used in these processes, and how particular decisions within the design process were reached. As a result of these discussions, the knowledge obtained about the design and implementation of high capacity bus systems has influenced the way in which the decision model could be designed. The results of these discussions also have an important effect on the structure of this thesis.

It is important to consider the context within which high capacity bus technologies have been designed in order to attempt to understand the way in which these decisions are taken, and this is introduced briefly in Chapter 2. This chapter also indicates ways in which a low-cost high capacity bus system might be relevant in the urban areas of the United Kingdom.

Chapter 3 introduces the expertise gained during this research concerning techniques which can be used to enhance the capacity of the bus system. The analysis of such systems is not generally well documented, and much of this expertise is analyzed in order to establish the ways in which it affects the design of high capacity bus systems. During the course of the discussions with the expert engineers, however, it became clear that a large number of the decisions taken during the design process are based on qualitative inputs. These may be the interpretation of the output of a quantitative model, simply the qualitative assessment of an important aspect of the design process which is not amenable to calculation, or the qualitative assessment of data.

In order to investigate the development of a decision model, therefore, it is necessary to consider existing models for the representation of expertise. These are considered in Chapter 4, which considers knowledge-based systems. This investigation reveals, however, that the current methods used in such systems are not adequate for the type of qualitative assessments required for the design of high capacity bus systems.

Chapter 5 therefore discusses these difficulties and presents a model to represent the use of judgement and opinion within the decision process. The use of this model within the decision process requires particular information from the expert engineers, and the acquisition and calibration of this information is discussed in Chapter 6.

Chapter 7 describes the form of the resulting simple computer model of the design decision process for high capacity bus systems.

A problem for knowledge-based systems is the validation of the model. This is because the nature of such systems is such that it is very difficult to test the model against real decisions made by the experts from whom the knowledge was obtained. In this case, however, it has been possible to provide the model with data which refer to a corridor in Brazil which now has a high capacity bus system designed by some of the experts from whom the expertise was obtained. It is therefore possible to ascertain whether the model produces a similar design to that of the experts when presented with the same data. This validation exercise is described in Chapter 8.

Testing the model in a situation in which a high capacity bus system has been implemented is necessary but not sufficient. Chapter 9 therefore describes two experiments which are designed to test the performance in less satisfactory conditions. Data for these experiments represent a corridor in London where the width constraints are potentially too narrow for a typical system, and a corridor in Lima, Perú where environmental considerations place some limitations on the possible solutions.

The designs are evaluated against the expertise in order to attempt to measure whether the evaluation of the decisions taken in each case is feasible given the expertise.

Chapter 10 draws some conclusions from the research and suggests some areas in which the research into decision modelling might be pursued in the future.

CHAPTER 2
BACKGROUND: THE SOURCES OF EXPERTISE IN THE DESIGN OF HIGH CAPACITY BUS SYSTEMS

2.1 INTRODUCTION

The sources of expertise in the area of high capacity bus systems are often to be found in developing countries. This chapter discusses briefly the reasons for this in order to place in context the discussion undertaken in Chapter 3.

A major source of the urban transport problems in developing countries is the problem of urbanization. This is discussed in Section 2.2. Section 2.3 considers some conditions that arise in developing countries that may influence the approach to the supply of urban transport. Section 2.4 discusses the impacts of the availability of resources in the decisions taken about transport infrastructure in these environments. The expertise obtained from these environments may have a potential importance for countries such as the United Kingdom. This is discussed briefly in Section 2.5. Some conclusions are drawn in Section 2.6.

2.2 URBANIZATION

In order to place in context the rationale behind the use of high capacity bus systems, this section considers the demographic changes which bring about the high levels of demand for public transport in these countries.

Beier *et al* (1976) suggest that the most significant aspect of the level of development in developing countries is the level of urbanization. Using the measure used by the United Nations (using as a basis the ratio of urban population to the total population), Beier *et al* suggest that developing countries may be categorized into groups with similar rates of urbanization. It is also suggested that the rate of urbanization varies with time and that therefore it could be expected that developing countries with low current urbanization rates will at some stage in the future have much higher urban growth, thus increasing the global urban population.

Bayliss (1981) predicts that before the new millennium there will be 1 billion new urban dwellers, of which some 870 million will be located in developing countries. A large part of this increase is due to migration from poorer rural areas to the cities which are perceived to be the source of higher incomes.

The position in any one country is much more complex. Chadwick (1987) cites Mexico as an example: between 1940 and 1970, four cities in Mexico each grew at annual rates of over 7 percent, five more were over 6 percent and a further five were in excess of 5 percent. Chadwick points out that the element of urban growth due to migration from the poorer rural areas was not the same for all cities. For example, Mexico City itself accounted for nearly half of the total migration in this period.

In terms of the economy, this variability can have strange effects and this also points to the dangers of using national average figures as indicators. Thomas (1980) shows that poverty in Perú is relative to location because of the higher cost of living in cities. A comparison is shown in Table 2.1. Thus, for example, the price of "the typical diet of the poor" is 25 percent more in Lima than the national average, and in the Rural Sierra the equivalent diet costs 16 percent less than the national average. A similar difference is also noted for minimum wages which are considered to be 40 percent higher than the national average in Lima and 28 percent less in the Rural Sierra. This has profound implications on the availability of resources - of both people and the state - for expenditure on transport.

Table 2.1 Comparison of prices and minimum wage for Perú, Lima and the Rural Sierra region. *Source: Thomas (1980)*

	Perú	**Lima**	**Rural Sierra**
Food	100	125	84
Non-food	100	157	56
Overall	100	141	70
Based on minimum wage	100	140	72

The relative poverty of the migrants makes it likely that they will settle in those areas of the city which are more difficult to occupy. Such areas may be far from the city centre, or on difficult terrain, or in dangerous parts of the city. The migrants are therefore likely to be in a position where they require transport provision in order that they might be able to reach the more lucrative centres of employment in the city centre. Due to the access difficulties mentioned above, transport is difficult to provide cheaply. With the sheer numbers of people involved, this often requires the provision of (expensive) high capacity public transport systems.

This indicates that the perceptions noted earlier in this section that the move to the city is likely to raise incomes is fraught with difficulties. While incomes are higher in Lima than the Rural Sierra, so is the cost of living, and if it is not possible to find employment it is obviously harder to achieve even a subsistence level of income. As a result the informal economy probably accounts for a significant amount of the income in Lima, but of course this is difficult to quantify. Without the informal economy the poverty problems in Lima would be much more difficult to alleviate. This also has implications for the provision of transport.

Hardoy and Satterthwaite (1989), however, suggest that a note of caution should be observed when considering statistics concerning urban growth in developing countries. This is necessary because the statistics upon which these projections are made are extremely poor. These often take no account of differing definitions of urban population and make no statement about the definition of the city boundaries used in the survey. In many cases, the statistics are aggregated for a country as a whole, or even for an area or region. Even when considering one city, it is difficult to estimate whether the city is becoming more dense in terms of population, or is spreading in terms of area. These two possibilities suggest completely different problems for the transport planner, and it should be clearly understood which situation exists before any strategy is considered.

Taking as an example the scenario of an increase in land area which accompanies an increase in population, the possibility exists that as the city grows in area, it engulfs other (smaller) centres of population. These are then subsumed into the metropolitan population figures. As a result, Hardoy and Satterthwaite suggest that the urban population growth may not be as high as that proposed by Bayliss, Chadwick and Beier *et al.* This has implications for the transport sector, especially if it implies that there may be an observable set of transport patterns (and maybe infrastructure) in existence in these population centres before they are subsumed into a grand metropolitan area.

The transport problem in this case could be one of linking the local centres. This also has implications on the level of demand from each centre towards the main city centre. A reduction in the need to travel to the centre of the city obviously suggests that a lower capacity system may be satisfactory on these routes. Greenstein *et al* (1991) show how a bus operational problem in Quito was solved after it was realized that a large number of trips did not need to enter the historic city centre at all.

This presents a different problem to the rural-urban migration described by Beier *et al* where populations arrive in large numbers in the more inaccessible regions of a city, and therefore require transport into the city centre along corridors where none currently exists.

Hardoy and Satterthwaite also suggest that the major problems in the cities of developing countries are not just the result of population growth itself. An institutional structure which is incapable of coping with the needs of the population or the functions involved with running city services, and which does not have the capacity for capital investment exacerbates the problems caused by poverty. Brundtland *et al* (1987) agree with Hardoy and Satterthwaite on this point, pointing out that few city governments in developing countries have the power, resources or staff to ensure that their populations have sufficient facilities (including transport) for civilized life. In such a context, high capacity transport infrastructure, such as a heavy rail system, is often not possible financially. There is therefore a need to examine the possibilities for lower cost options such as high capacity bus systems.

2.3 URBAN TRANSPORT

As indicated in Section 2.2, the economic well-being of potential users is clearly relevant to the provision of urban transport. Low incomes imply low revenues which in turn imply low investment in infrastructure. Even though fares may be low by the standards of industrialized countries, however, they account for a disproportionately large amount of the urban household income. In Brazil, transport expenditure accounts for some 12 percent of wages for low income households (Lindau 1987). As cities grow and distances become greater, modes which do not involve public transport (such as walking) become less feasible and the transport cost becomes higher, thus exacerbating the problem.

Table 2.2 Numbers of buses in private and public ownership in various cities in developing countries. *Source: Armstrong-Wright and Thiriez (1987).*

	Private	Public	Total	% private
Sao Paulo	5850	3280	9130	64.07
Lima	8400	700	9300	90.32
Santiago	8300	0	8300	100
Mexico City	9000	6500	15500	58.06
Jakarta	3915	1940	5855	66.87
Manila	3000	503	3503	85.64
Lagos	5000	371	5371	93.09

vehicle types represented: large buses, small buses, minibuses; not shared taxis or jeepneys

Currently, most urban communal transport in Latin America is provided by the private sector. Armstrong-Wright and Thiriez (1987) show that in Sao Paulo, this sector accounts for about 65 percent of the bus fleet, but in Lima this figure rises to about 90 percent. Santiago has no bus operation in the public sector at all. Some other data for comparison are shown in Table 2.2.

In Lima and Santiago, the private sector consists of mainly owner-operated vehicles. The owners frequently belong to associations, often directed towards particular routes. The proliferation of private owners does tend to be favoured in situations where the bus size can be small since this obviously reduces the capital cost of the vehicle. The associated increase in labour costs is offset by the low wages. Undoubtedly, the corollary also acts on this system: the bus size is determined by the private operator's ability to pay (or raise finance) for the purchase of a vehicle. Particularly where the latter applies, the result tends to be a fairly low quality of vehicle. In Lima, small vehicles are often adapted to carry more passengers (for example, by removing the seats) to enhance the owner-driver's income without the expense of purchasing (and the bother of registering) a larger vehicle.

In order to accommodate high levels of demand with small vehicles, the number of vehicles becomes large, and this results in extremely high flows of buses, particularly near to the city centre. The problem in Quito (mentioned in Section 2.2) arose because of the large number of buses entering the city centre during the peak hours. Greenstein *et al* note that 900 buses per hour were entering this area (which is only 1 km^2). The problems caused by such bus flows are particularly evident at and around bus stops, where the congestion caused by buses is one of the major constraints on operating efficiency.

Although this issue has been raised in the context of small buses, similar problems arise with standard (12 metre) buses. The problems in Brazil described in Chapter 3 involve such vehicles.

Lindau and Willumsen (1988) suggest that in order to meet some of these conditions, transport systems are required in developing countries which should be:-

1. low cost so that they can operate with little or no subsidy,
2. flexible in their ability to meet the constant changes resulting from the high urban growth rate,
3. adaptable in their capacity so that they can be efficient for both low and high demand,
4. technologically secure so that continued operation is not dependent on new technology which is untried in unskilled environments,
5. able to operate in those parts of the city which are poor in terms of infrastructure without the need for expensive capital investment.

Lindau and Willumsen show that bus technology is well suited to all these conditions, the only adaptations required being to extend its normal capacity ceiling to cater for the high demand occurring in urban corridors.

2.4 RESOURCES FOR TRANSPORT

The resources available for urban transport investment in developing countries are dependent on the investment policies of financial bodies such as the commercial banks, the World Bank and the International Monetary Fund. With the problems arising from the debt repayment crisis, such funds are increasingly difficult to obtain.

Armstrong-Wright (1986) suggests that the cost per kilometre for the construction of an underground metro system is approximately US $ 100 million (at 1983 prices and rates of exchange). Armstrong-Wright also estimates that a high capacity bus system, such as those developed in Brazil, could cost less than US $ 1 million per kilometre, and even the most expensive is estimated to cost US $2 million per kilometre. Clearly if it is possible to design such systems a considerable cost saving could be achieved.

Internal resources are another source of finance, and sometimes it is perceived to be in the country's interest to invest in a particular form of transport. Another reason for using internal money is that external sources have refused to fund the project. An example of this is the metro system currently under construction in Lima. When a project is financed internally, it must take its place in the queue for resources. This can have two results. Either the project never happens, and the problem may need to be solved by another method, or it is viewed as a prestigious project and in effect becomes a priority measure. Whether this is appropriate or not is open to question.

This reinforces the advantage of low-cost solutions to the transport problem since these require a smaller allocation of scarce resources. Bus systems, due to the low cost of implementation, can offer such benefits.

2.5 POTENTIAL FOR OTHER ENVIRONMENTS

Scarcity of resources is not, of course, confined to developing countries, and neither is the need to provide transport for large numbers of people in urban areas. The cost of metro construction noted in Section 2.4 is no less high in industrialized countries, where labour costs are higher even if the technology is more readily available. The conditions noted by Lindau and Willumsen and mentioned in Section 2.3 also could apply to industrialized countries (albeit with some modifications).

Cost is a prime concern to Governments, and is therefore a major factor in the decision making process concerned with the choice of technology for transport. The availability of low-cost high capacity systems may be of utmost importance in the future in these environments, and should be seen as a viable alternative during the course of feasibility analyses.

The flexibility of a transport system is essential in developing countries as a result of the high rates of urban growth mentioned in Section 2.2, but it is also important in a different way in industrialized environments. In these environments, it is important to provide attractive and accessible public transport systems in order to encourage their use in place of the car for a wider range of journeys than is currently the case. In this way, car use could become more selective because public transport is easier and more attractive to use, particularly in the context of the introduction of traffic restraint policies in congested areas.

The capacity of the public transport system is often perceived to be a question of mode: Vuchic (1981) for example suggests that bus operation is possible up to about 9000 - 10000 passengers per hour per direction, after which a tram system is needed, with a metro system required for even higher capacities. The high capacity bus systems discussed in this thesis were developed to extend the capacity of the bus system in order to provide a medium term solution to the capacity problem without the necessity of investment in expensive metro technology. There is clearly a potential for such systems within the context of industrialized countries.

Wherever a transport system exists, it is essential that it is capable of reliable operation. This implies that the technology used should be thoroughly tested and shown to be reliable. The use of bus technology for high capacity systems involves existing technology which is well-tried and proven.

The cost of implementation of a high capacity bus system, and the ease with which it could be implemented, suggest that it is possible to provide such injections of capacity cheaply and quickly. This may be of interest to industrialized countries which require a capacity increase within a timescale which is shorter than the time required to implement a full-scale metro system (or even a light rail system).

2.6 CONCLUSIONS

The economic and demographic problems found in urban areas of developing countries have generated the need for low-cost, high capacity transport systems. The resources available for the implementation of such systems are limited, and therefore it has been necessary to adapt cheaper technologies in order to achieve the level of capacity required.

Bus technology has been used for such systems in Brazil, and it is therefore appropriate to examine these methods in order to investigate the possibility of using them elsewhere to alleviate similar capacity problems.

Solutions developed within developing countries are attractive as they may be more suitable for transfer to developing countries than those developed in industrialised countries. This attraction is due to the expectation that developing countries either experience similar problems and have similar difficulties with finding available resources, or can be expected to do so in the foreseeable future.

This similarity is less likely to be found when comparing developing with developed countries. Nevertheless, the most important criterion is that the solutions are appropriate to the new environment. This requires analysis of the basic principles behind any design to establish its particular requirements and effects. Such solutions are discussed in Chapter 3 and the analysis of the principles behind their successes and failures could suggest which aspects of a design require extra care when implemented in a different environment.

These technologies, whether in whole or in part, may be relevant to industrialized countries and several of the analytical problems discussed in this thesis are not without relevance to countries such as the United Kingdom.

CHAPTER 3
BUS SYSTEM CAPACITY

3.1 INTRODUCTION

Cities with different characteristics have different requirements for their bus systems. Since there exists a wide range of city 'types' in the world, there is a correspondingly large number of ways in which bus systems can be operated. Nevertheless, as discussed in Chapter 2, a problem which recurs in the developing world is a lack of capacity within the bus system and there has been a number of attempts to solve this problem.

The means used in developing countries to achieve this vary, but a major difficulty is that the analysis of alternative designs is not always rigorous. There are three possible reasons for this. First, the existing situation is often so difficult that remedial work needs to be done very quickly and the priority is to implement a system rather than to analyze whether or not it will work - if the first idea does not work, another solution can be sought. Second, the resources are not available for extensive surveys to be carried out to provide detailed information which is often already observed in some qualitative sense. Third, in many cases the experience of designers is not geared to analysis in the same way is it is in developed countries, and this way of working is simply not understood.

Some of the current practice is examined in this chapter, and the underlying theory used to resolve capacity problems, with the objective of identifying a suitable range of techniques that may be appropriate for the design of high capacity bus systems.

This chapter is concerned with the expertise which exists with respect to the design of high capacity bus systems. Such systems are not common and the literature is correspondingly scarce. High capacity bus systems are, however, developed using the basic principles which underlie the design of any bus system, and it is possible to suggest factors which may be influential when considering expanding the capacity of a bus system towards the levels described in Chapter 2.

The techniques discussed in this chapter are those which are related to the high capacity bus systems which exist. The discussion therefore reflects the techniques used in these systems to extend the capacity of the bus system. Consequently, some bus priority techniques, for example selective detection and bus priority through traffic signal control, are not discussed here as they are not directly related to the capacity of the bus system.

Guided bus systems have been developed in Germany (Baron (1987)) and Australia (Wayte and Wilson (1989)), and can provide high commercial speeds. These systems are not, however, geared to high capacity operation since they are constrained by the inability of buses to overtake each other. As will be seen in Section 3.3.3, this is a major constraint on the operation of bus systems. Consequently these systems will also not be analyzed in detail in this thesis.

This chapter is divided into four further sections. Section 3.2 is concerned with bus lanes. Bus lanes are found in many parts of the world. The reasons for their use vary, however, according to the objectives required and their design is consequently affected. Section 3.3 considers another important element of the bus system: the bus stop. As will be seen, bus stops are extremely important factors in the bus system and the quality of their performance is at the centre of the design of high capacity bus systems. Section 3.4 discusses two types of high capacity bus system which have been developed and implemented in Brazil. Conclusions drawn from this review of existing knowledge are discussed in Section 3.5.

3.2 BUS LANES

Bus lanes are traffic lanes which are reserved for the exclusive use of buses (although in some circumstances other vehicles may be admitted). They are usually designed for use by buses travelling in the same direction as the traffic in adjacent lanes (with-flow lanes), but some are designed to allow buses to travel in the opposite direction to that of traffic in adjacent lanes (contra-flow). For the most part, this thesis is concerned with the design of with-flow bus lanes, but, as will be seen, some attributes of contra-flow lanes are also discussed where this may provide useful information.

In the United Kingdom, a with-flow bus lane is characteristically short, and is designed to alleviate problems in bus operation at a particular bottleneck in the traffic system. Hounsell and McDonald (1988) evaluated the effects of some 22 typical with-flow bus lanes in the United Kingdom of which 2 are longer than 1 kilometre, but 6 are less than 200 metres in length. Bottlenecks in this case are often thought to be located at junctions, particularly where these are controlled by traffic signals. It can be difficult sometimes for buses to reach the stop line at a junction as a result of the queuing traffic, and therefore a bus lane is used to allow buses to pass the vehicles in the traffic queue. In this way, a bus can reach the stop line earlier than it would otherwise, thus reducing the delays to the bus system from the particular bottleneck. These are called 'point measures' and are discussed in Section 3.2.1.

Another type of with-flow bus lane is intended to provide buses with a reserved track over some considerable distance in order to remove them from the influence of other traffic, and to decrease the amount of time it takes to travel along the corridor. Such bus lanes are intended to assist in the provision of a level of capacity for passengers which is thought to be either necessary or desirable, and for which buses are deemed to be the operational solution. These are called 'capacity-enhancement measures' and are discussed in Section 3.2.2.

3.2.1 Point measures

Point measure bus lanes are usually located near to bottlenecks in the traffic system such as junctions. Junctions have a capacity which is determined by a number of factors, one of which being the number of lanes at the stop line. If one of these lanes is reserved for the exclusive use of buses, the capacity of the junction (in terms of vehicles) would necessarily decrease and this could cause additional delay to road users other than buses. In order to alleviate this problem, a major part of the design effort for this type of bus lane is taken up with the compromise between the vehicular capacity of the junction and the benefits to the bus system. This compromise appears in the bus lane design as the setback. A setback is the distance from the stop line to the point at which the bus lane ceases in order to make available more road space for all vehicles at the stop line. The length of the setback is therefore of the utmost concern in the design of point measure bus lanes.

3.2.1.1 The setback

Iunes and Willumsen (1988) suggest that a reasonable setback length can be approximated by the equation:-

$$L_s = \frac{5.5 \;.\; S \;.\; g}{3600 \;.\; n} \tag{3.1}$$

where

L_s	=	setback length (in metres)
S	=	saturation flow at stop line without bus lane (pcu's per hour)
n	=	number of lanes at the stop line
g	=	effective green time at the junction (in seconds)

Equation (3.1) relates the setback length to the green time. Oldfield *et al* (1977) found that this relationship could be as low as 0.75 metres per second of green time or as high as 2.25, depending on the saturation flow at the traffic signals. Hounsell and McDonald (1988) also stress the importance of this relationship and suggest that a value of 2.7 is required to ensure no loss in entry capacity. This suggests a relationship between the number of vehicles it is useful to store in the setback and the number it is possible to discharge during the available green period. This equation does not take account of the number of buses in the traffic stream. Since no provision is made for the second and subsequent buses to cross the stop line during the first green period, this implies that the bus flow is no more than one bus per cycle. It is also implicit that the load factors are constant.

Iunes and Willumsen point out that a setback distance which is shorter than the optimal distance forces a stepped discharge from the stop line which reduces when the vehicles held in the setback have all been released. Iunes and Willumsen suggest that it is reasonable to suppose that the best setback length is likely to be found through simulation, as the trade-off between priority and non-priority vehicles is not amenable to calculation.

Oldfield *et al* (1977) suggest that it is reasonable to reduce the length of the setback according to the number of buses that arrive during the red period. For example, with a large flow of buses and a degree of saturation of 0.9, a reduction in the length of the setback of about 7.5 metres per bus is suggested. This falls to about 4 metres per bus when the bus flow is small and the saturation level is about 0.7.

The economic analysis of Oldfield *et al* assumes the use of the optimum length of setback. A setback which is slightly longer than the optimal would tend to favour the general traffic since the storage capacity for the junction for non-priority vehicles would be increased. A setback which is slightly shorter than the optimal would tend to favour bus users, since this increases the probability that a bus will be able to cross the stop line during the first available green period. A small benefit may also arise if the associated extension of the bus lane would enable a bus to pass more queuing vehicles before the junction. Hounsell and McDonald (1988) note that the amount of green time allocated to the link is also of key importance in this respect. If this is more than 60 percent of the cycle time, the red time is insufficient for the setback to be filled and overall disbenefits will occur.

3.2.1.2 Infrastructure

Since the assessment measure is economic rather than operational (although there are obviously operational benefits inherent in the system), the design of infrastructure hardware is secondary to the estimation of delay. Physical segregation is not generally considered necessary in the United Kingdom for with-flow bus lanes. Lane and Hodgkinson (1973), for example, in an evaluation of a network of bus priorities for central London, do not consider methods of segregation, and consider the major costs of implementation to be attributed to changes to traffic signals, signs and road markings.

In the United Kingdom, these are currently required to conform to guidelines of the Department of Transport (Dtp 1976, 1988). The lane is marked by a solid white line, and the script 'BUS LANE' is inserted to identify the lane at each entry point. Often the lane is marked with a different colour surface. This increases cost (thus affecting the economic assessment) but does help to indicate the special nature of the lane to other road users. Sometimes, only buses are allowed in the lane during operational periods, but generally taxis and/or bicycles are permitted - especially where this increases the net economic benefit sufficiently to justify the installation (Oldfield *et al* 1977).

Contra-flow bus lanes are usually distinguished by the use of physical infrastructure. Two reasons for this may be inferred. First, there is a potential safety problem where the majority of traffic flows in one direction and a (relatively low density) stream of buses is travelling in the other. This is of concern to pedestrian safety, and it is important that pedestrians are aware of the possibility of vehicles approaching from two directions in what appears to be a one-way street. In order to separate the two streams of traffic, a system of kerbs - sometimes with fences to discourage pedestrians from crossing - is often used. A good example of the use of this type of infrastructure is the contra-flow bus lane in Piccadilly, London.

The second reason for such infrastructure is to preserve the exclusivity of the bus lane. Clearly, since all other traffic is travelling in the opposite direction to that of the buses in the bus lane, there is a potential difficulty if a vehicle attempts to use the contra-flow lane in the wrong direction. When drivers are aware of the use of the bus lane (and the direction in which the buses are heading), there is not a particular problem for enforcement with contra-flow lanes. It is interesting to note that a contra-flow lane with none of this infrastructure is currently operating successfully in London along the Charing Cross Road, where the available width is not sufficient to permit the construction of segregation infrastructure of this type.

3.2.1.3 Summary of point measures

Point measures are essentially responses to problems within the bus system which are caused by traffic-induced delay. These are therefore mainly associated with the perceived cause of this delay, and this is commonly associated with junctions. Such measures are important to the operation of a bus system, but they are not designed specifically to increase its capacity. In particular, the usual objective of a contra-flow lane is to permit buses to maintain their route after the introduction of a one way traffic system which has diverted other traffic along a longer route. Such a lane not only shortens the travel time for buses (when compared with the time taken to travel along the one way system), but also ensures that the bus stops are kept in the original area.

In order to enhance capacity of the bus system, it is necessary to use measures which are directed to this purpose. Some of these measures appear to be similar to some point measures (for example, with-flow bus lanes). The different objective, however, demands a slightly different approach to the way in which these are designed. This is now discussed in Section 3.2.2.

3.2.2 Capacity-enhancement measures

All bus systems provide capacity through the use of a number of buses. The problem for the designer of a high capacity bus system is therefore the control of large numbers of buses in close proximity rather than traffic-induced delay. One reason that capacity is not often raised as a problem in industrialized countries is that the bus flows rarely approach the capacity of the system. This situation is very different in developing countries. It is therefore informative to examine the measures taken to assist buses in developing countries in order to obtain information about the design of high capacity bus systems.

Table 3.1 A comparison of urban bus and vehicle ownership levels.
Source: Fouracre *et al* (1981)

	buses per 1000 population	cars per 1000 population
INDUSTRIALIZED COUNTRIES		
United Kingdom	0.84	230
Rest of Europe	0.52	200
N. America	0.36	350
DEVELOPING COUNTRIES		
Asia	0.48	20
Africa	0.3	30
Others	0.63	40

3.2.2.1 General issues

Fouracre *et al* (1981) show that in developing countries, the proportion of the traffic stream made up by buses is much higher than in developed countries, yet the number of buses relative to the population is much lower. Table 3.1 shows some comparisons between bus provision and car ownership in developing countries and those found in industrialized countries. In developing countries, the buses must be accommodated in order that efficient operation can be facilitated. With such a high proportion of buses in the traffic stream, buses tend to impose their own priority in terms of the use of the available road space (Gibson *et al* 1989a).

The major impedance to the progress of buses in this situation is the congestion caused by other buses, especially at bus stops. Bus priority measures in this context therefore need to solve the inter-bus conflicts that occur at bus stops. Thus a central feature of bus priority design in this situation is the design of bus stops to minimize delays caused to buses by activities other than those required to serve passengers. Gibson *et al* (1989a) suggest that the interactions between buses and passengers can also be improved through good bus stop design to reduce the amount of time required to pass through the bus stop area. Bus lanes are therefore the means of delivering buses to the stops in such a way that the operation of the bus stops is enhanced. With improvements to bus stop design, therefore, the delay to buses which arises as a result of the bus stops is reduced. This has an important effect on the overall speed with which buses can operate, and thus on the journey time of their passengers. For this reason one of the major measurement criteria for the efficiency of a bus lane system in this environment is the commercial speed.

3.2.2.2 Commercial speed

Commercial speed is the average speed along a section of a route, including all stops but excluding any time in which a bus is not operating at either end of the section (for example time allowed for meal breaks for the driver, known as layover time). Commercial speed is calculated by dividing the length of the corridor by the time taken to travel from one end to the other including all stops (whether caused by congestion or bus stop delay). The exact nature of the relationship between bus stops and commercial speed is explored in more detail below.

Undoubtedly, with-flow bus lanes, as described in Section 3.2.1 above, can help bus operation in these environments. In Bangkok, for example, the installation of with-flow bus lanes improved both bus and car speeds since the buses were, after implementation, contained within a reserved lane, and cars were able to proceed unimpeded by slow and stopping buses (Marler 1982). Four years later it was found (Jamieson Mackay and Partners (JMP) 1988) that while bus speeds had slightly improved, the variability of travel times had increased. Some data referring to the northbound direction of Pahonyothin Road from these studies is given in Table 3.2. It seems that this improvement in speed was maintained as a result of the use by bus drivers of adjacent vehicle lanes to overtake other buses, or to stop next to other buses at bus stops. This in effect increases the efficiency of the bus stop by allowing more buses to call simultaneously. The tendency of bus drivers to try and enhance capacity in this way has also been noted in Santiago (Gibson *et al* 1989a).

Table 3.2 Pahonyothin Road (northbound), Bangkok. Comparison of survey data from 1980 (before and after implementation of the bus lane) and 1985. *Sources: 1980 data - Marler (1982); 1985 data - JMP (1988).*

	1980 (before)	**1980 (after)**	**1985**
BUSES			
mean travel time	7.47 minutes	5.47	5.35
standard deviation	1.8 minutes	0.7	1.16
mean speed	20.1 km/h	27.4	28.0
flow per hour	241	234	230
OTHER VEHICLES			
mean travel time	6.4 minutes	4.56	5.0
mean speed	23.4 km/h	32.9	30.0
flow per hour	2147	2217	1880

. It has been suggested that the practical capacity of a single-width bus lane is about 120 buses per hour (Webster and Bly 1976). At higher flows, the necessity for some form of additional capacity is therefore required. The flows observed in Bangkok are almost double this level, and this almost exactly accounts for the level of violations by buses of the system (JMP 1988). A similar result can be observed in the Avenida Nossa Senhora de Copacabana in Rio de Janeiro, Brazil, where a double width bus lane has been installed to accommodate bus flows of approximately 200 buses per hour.

The major problem with this type of bus lane is enforcement. JMP notes the high level of violations (both by buses travelling outside the bus lane and by other vehicles travelling inside it) in Bangkok after the policing level had been reduced. The implicit assumption made by JMP is that buses should be constrained to use only the bus lane. As has been indicated above, however, this is a practice which could have resulted from insufficient capacity in the bus lane itself. The capacity of a single width bus lane (permitting only one bus per bus length of road) is limited by the activity surrounding the bus stops where buses are slowing down and stopping.

In these situations, the bus stop as commonly recognized in the United Kingdom seldom exists. In the urban environments of developing countries, a bus stop is more of a bus stop area which is loosely bounded by the furthest point from the local attractor considered practical by the passengers using the stop to board or leave a bus. Bus stops such as these may be the source of delays to buses which are additional to those resulting from traffic-induced delay. While the effect of such stops on associated bus lanes is not currently measured in the United Kingdom, there is a considerable body of knowledge overseas which has been built on the basis of the necessity of solving the problem of bus stop design where capacity is restricted. Wherever such bus stops exist, their effect on the operation of the bus system can be profound, and in order to examine the efficiency and capacity of an adjacent bus lane it is therefore useful to examine analytically the workings of such a bus stop. This is discussed in Section 3.3.

3.2.3 Enforcement

Bus lanes, whether point measures or whether designed to enhance the capacity of the bus system, will suffer a decrease in the economic benefit that could have accrued from their implementation if they suffer from non-compliance by other traffic.

Two types of violation that can occur with bus lanes are identified by Hounsell and McDonald (1988): stationary and moving. Stationary violations are normally the result of the loading or unloading of goods vehicles or car parking. Hounsell and McDonald consider that police enforcement is expensive and will not solve a parking problem: if parking is likely to occur, then a bus lane is not a good solution.

Roark (1982) states that four elements contribute to the enforcement problem:-

1 the cost of enforcement measures
2 lack of understanding of the objectives
3 absence of a clear organizational responsibility for enforcement
4 design of facilities and operations that defy enforcement

Roark also states that it is necessary to place an emphasis on project designs that encourage self-enforcement. Several enforcement strategies are described, but none includes the use of infrastructure such as physical segregation.

Significantly, contra-flow lanes seem to be less liable to violations. Webster and Bly (1976) suggest that to obtain a similar level of compliance for with-flow lanes, it would be necessary to segregate the bus lane physically from other traffic. This can, however, cause difficulties for local commerce who may fear that there would be a loss of access for loading and unloading.

In Avenida Nossa Senhora de Copacabana in Rio de Janeiro (Brazil), no physical segregation is provided since money was not available at the time of implementation. Here the double-width bus lanes are marked by continuous yellow lines. Violation of the bus lanes by non-priority vehicles is common, and tends to detract from the success of the scheme since buses are prevented from passing each other around the bus stop areas.

The level of violations in this bus lane in Rio de Janeiro suggests that physical segregation should be employed where possible, particularly where the lanes are in permanent operation.

3.3 BUS STOPS

3.3.1 Introduction

Bus stops are the most common points of access to the bus system (access in this case includes not only access to the bus system but also egress from it). A bus stop is defined for the purposes of this thesis as an area in which buses may stop in order to allow passengers to board or alight. A formal bus stop is an example of this where there is some form of indication that this is a location at which buses could be expected to stop for this purpose. Bus stops may be 'compulsory', where buses are obliged to stop whether or not there are passengers. 'Request' stops (on-call in American terminology) are formal stops at which a bus stops only if requested to do so by a passenger wishing to board or alight. 'On-demand' stops are informal bus stops, and are established only by the fact that a passenger requests a bus to stop at a particular location. Where demand for boarding and alighting is high, such stops may become quasi-formal in the sense that although no formal stop exists, bus drivers can expect to be asked to stop at a particular point (such a system exists in Santiago, Chile for example). In low demand areas, such stops are only made when a passenger requests a bus to stop, and the location of these stops is determined by the passenger (subject of course to prevailing conditions as interpreted by the driver). This is common in rural areas of Perú for example.

3.3.2 Dwell time

A bus uses a bus stop for a defined period of time, known as the dwell time. Cundill and Watts (1973) show that dwell time is linearly dependent on the number of passengers boarding each bus, and this can be expressed in the following form:-

$$T_{dwell} = t_{dead} + \mathrm{K} \cdot p_{boarding} \tag{3.2}$$

where

T_{dwell} = dwell time

t_{dead} = dead time

K = marginal boarding time

$p_{boarding}$ = number of boarding passengers

Thus dwell time may be considered in two parts: dead time (t_{dead}), discussed in Section 3.3.2.1, and the time spent serving passengers ($K.p_{boarding}$), known as the 'passenger service time' which is discussed in Section 3.3.2.2. Gibson *et al* (1989a) also measure the time during which a bus is unable to leave a bus stop due to obstruction by other buses. The resulting delay is called 'internal delay' and is discussed in Section 3.3.2.1. Cundill and Watts suggest that two elements which influence dead time and marginal boarding time are the design of the doors on the buses using the stop and the fare collection system used. It is useful to consider how these factors affect dwell time.

3.3.2.1 Dead time

In this thesis, dead time is considered to be the time spent by a bus at the bus stop during which passengers can neither board nor alight. This can be calculated by adding the two appropriate time intervals - between the time when the bus stops and the first passenger boards, and between the time when the last passenger has boarded and the bus starting to move away from

Table 3.3 Dead times and marginal boarding times for various types of buses. *Source: Cundill and Watts (1973)* *small sample size

bus type	dead time	standard error	marginal boarding time	standard error
single door vehicles				
two person operated				
open rear platform (London)	0.95	0.2	1.15	0.05
front door	2.3	0.4	1.5	0.05
one person operated (OPO)				
flat fare (Peckham)	4.7	0.35	3.0	0.15
"autofare" (Hull)	0.95	0.15	2.25	0.1
"autoslot" (Slough)	5.5	0.45	2.7	0.25
two door vehicles (OPO)				
traditional OPO (Reading)	4.3	0.6	2.5	0.25
traditional OPO (Bristol)	5.5	0.75	4.65	0.4
flat fare (London)	5.6	0.55	4.6	0.3
"autofare" (Hull)	2.0	0.35	1.85	0.15
split entry (London)	7.65	0.6	6.85	0.4
Red Arrow (London)	5.65	1.05	8.2	1.0*

the stop into the appropriate traffic stream. In practice, dead time is usually taken to be constant. As can be seen in Table 3.3, however, dead times can vary considerably between different types of buses. Cundill and Watts consider that this also tends to be dependent on the safety arrangements (such as door interlocks) concerned with door operation. It is useful, therefore, to consider the issue of the design of doors on buses with a view to their effects on dead time.

When a bus has no doors (for example the 'Routemaster' bus in London), it is possible to board or alight while it is still in motion (making the dead time effectively zero or even negative). It is also possible to board or alight a bus which has no doors at places other than at a bus stop (for example while the bus is waiting at traffic signals). Both of these situations also occur where doors are fitted but are left open for most of the time, such as in Santiago, Chile (where this applies particularly to the front door).

Two effects arise as a result of these possibilities. First, access to the bus system is not confined to specified bus stops and the passengers have a greater choice of access point than is provided by the 'official' system. Second, there is a tendency to reduce the number of passengers boarding and alighting at the more formal bus stops, thus reducing the dwell time at these stops. This is discussed in more detail in Section 3.3.2.2.2 below.

An important element in the consideration of doors is their number. Some buses have one door, while others may have more. As already mentioned with respect to London, some buses have no door, using an open entrance platform to provide access to the vehicle. A common practice is to have two doors, one intended for boarding and the other for alighting. This allows the two activities to take place almost simultaneously, thus reducing the passenger service time.

This concept is taken further by a Brazilian bus design which has three doors (one for boarding and the others for alighting) which is intended to reduce further the passenger service time by improving the flow of passengers inside the bus (EBTU 1988). Although the use of multiple doors clearly reduces the passenger service time, problems may occur at bus stops with a high demand, particularly where this involves a large number of boarding and alighting movements (for example at an intermodal interchange). In this case a problem can occur as a result of congestion between boarding and alighting passengers at the stop itself. A new bus design is being introduced in Sao Paulo to resolve this problem (Transporte Moderno 1990). It has five doors, three on one side and two on the other to allow boarding and alighting on opposite sides of the vehicle. This is intended to improve the circulation of passengers within the vehicle and thus to reduce dwell times. It has not been in service for sufficient time to enable a study to be made to check this hypothesis.

Articulated buses have been in use in continental Europe for several years, and in some developing countries as well. These often have three or four doors. Where more sophisticated fare systems allow, boarding may take place at any door on these buses. South American examples, however, require passengers to board through a particular door since the fares are still cash-oriented, and require collection on entry to the bus.

In Santiago (Chile) there are two doors on most buses. In most cases, the front door is labelled for boarding and the rear for alighting, although this strict allocation of doors to activities is not so evident in practice. Gibson *et al* (1989b) note that there is a tendency for passengers to alight from the door intended for boarding. The reason for this is that the bus system in Santiago tends to operate as the on-demand system described in Section 3.3.1. Passengers wishing to alight from the bus tend to stand near to the driver in order to indicate their choice of stop location. It is therefore likely that they will use the front door to alight.

Another aspect of doors on buses is their size. A typical width in older buses in Brazil is 0.70 metres. This is now felt to be too narrow (EBTU 1988) and newer buses are fitted with doors which are 1.2 metres wide in order to facilitate both boarding and alighting. In London, the doors are normally 1.2 metres wide, although the buses designed for one-person-operation have a handrail inside the bus (designed to ensure that passengers pass the driver in single file) which restricts the effective width.

In the United Kingdom, minibuses usually have only one door, but the design of this door has changed since this size of buses was first introduced. Early (smaller) versions have narrow doors, but more recent vehicles have wider doors (up to the 1.2 metre standard noted above and without the handrail restriction) in order to facilitate simultaneous boarding and alighting.

The 'Routemaster' bus, as mentioned above, has no doors but has an open platform at the rear of the vehicle. These buses tend to have low values for marginal boarding time and dead time, even when compared with crew-operated buses which have doors (Cundill and Watts 1973). Although boarding and alighting movements must take place using the same space on the platform, marginal boarding times are very low when compared with other arrangements (see Table 3.3). These effects combine to produce short average dwell times for these vehicles.

Because of the cost implications, it is often impractical to insist on a new design of bus in order to ensure the success of a proposed bus system. In practice therefore, the designer of a bus system is usually constrained by the design of the doors on the existing buses. It is therefore important that the appropriate characteristics are considered in the estimation of parameters in any model used for the analysis of bus operation at bus stops. As internal delay is not included explicitly in the model of Cundill and Watts, it may be considered to be included implicitly in the dead time. Baeza *et al* (1989) suggest that this could be a significant source of delay at congested bus stops but that analytical evaluation is likely to be difficult. Internal delay is affected by the operation at a bus stop and is therefore unlikely to be constant when this is disorderly. As will be seen in Section 3.3.3, internal delay may be controlled through bus stop design aimed at the reduction of interactions between buses.

3.3.2.2 Passenger Service Time

Passenger service time is the product of the time taken for each passenger to board the bus (referred to as the 'marginal boarding time') and the number of passengers. This is therefore considered to be directly related to the number of passengers. The effects of the number of passengers on the passenger service time are discussed in Section 3.3.2.2.2. The marginal boarding time (and therefore the passenger service time) is also affected by the fare collection system and this impact depends in the first place on who collects the fares. This is discussed in Section 3.3.2.2.1.

3.3.2.2.1 Fare collection systems

Three types of fare collection system are in common use:-

1. fare collection by the bus driver;
2. fare collection by another member of the bus crew on the bus; and
3. ticket sales by other agents away from the bus (for example shops or ticket offices).

If the driver collects the fares, then boarding must take place at the front of the bus in order to ensure that all passengers pass the driver on boarding. If another member of the crew collects the fare two options exist.

First, if the fare collector moves around the vehicle in order to collect the fares (as in the crew-operated buses in London), then boarding and alighting may take place at any door in the vehicle. If, on the other hand, the fare collector remains at a fixed point within the vehicle, then boarding must take place at a door which is upstream of the fare collection position. In these circumstances the door should be located so that passengers must pass the collector in order to alight (this is the typical arrangement in Sao Paulo for example).

In Lima (Perú), a number of fare collection systems exist. The state-owned buses have a front-entry pay-the-driver system similar to that found in London. Some private companies also operate this type of system. Others, however, have a fare collector who collects the fares normally when passengers alight (passengers may pay the fare during the course of the journey, but buses are often too full to allow this to occur).

In several continental European countries, ticket purchase takes place off the bus, and boarding passengers either show their ticket to the driver (as in Paris) or insert the ticket into a machine inside the vehicle for validation (as in Lyon, Milan or Amsterdam). In these cases, boarding may take place at any door in the vehicle and the validation machines can be situated accordingly.

Cundill and Watts compare the marginal boarding times obtained with a number of different fare collection strategies. Some results are shown in Table 3.3. These show the range of marginal boarding times for different types of fare collection system and bus design found in various cities in the United Kingdom. It is evident that both the mean marginal boarding times and the standard errors associated with crew operation (where fares are collected independently of the boarding process) are markedly lower than those for one person operation (OPO) (where the driver collects the fares as passengers board the bus). This suggests that when the action of boarding a bus is independent of the fare collection process, the marginal boarding time is both lower and more consistent than when fare collection takes place during the boarding process. This appears to be so whether the bus has one or two doors.

3.3.2.2.2 Number of passengers

Chapman *et al* (1976) obtain a similar model to that of Cundill and Watts (Equation (3.2)), and in a survey in Newcastle find that the dead time is 3.5 seconds for stops where boarding takes place and 4.7 seconds where alighting is the only activity at the stop. Marginal boarding time is 4.1 seconds, and the marginal alighting time is 1.4 seconds. Analysis of the variance of the boarding time distribution reveals that 72 per cent of variation in stop time can be explained by the number of passengers, 8 per cent by differences between drivers, and 20 per cent by other factors.

Zografos and Levinson (1986) note that as the seating capacity of the vehicle becomes filled, the situation arises when the congestion inside the bus is such that the entry area cannot be cleared. At this point, the marginal boarding time, hitherto fairly constant, begins to increase as passengers wait to board outside the bus. In this way, the value of the marginal boarding time might increase with the number of boarding passengers. It should be noted, however, that this condition only holds where the number of boarding passengers for a particular bus is sufficiently high that some passengers must wait outside the vehicle while previous passengers obtain seats and clear the entrance area. This is more likely to occur if the bus is already nearly full as it approaches the stop.

Guenthner and Sinha (1983) consider the time per passenger spent in boarding and alighting as equivalent. This is stated to be a negative logarithmic function of the number of passengers. It should be noted, however, that the correlation coefficient obtained for this model is rather low. Guenthner and Sinha suggest that this is because the marginal boarding time is also a function of other factors. Those likely to decrease this time include the boarding and alighting characteristics which apply to express bus operation (many passengers but few stops) and an increase in the number of passengers who have paid a pre-paid fare. Factors which may cause an increase include a large number of elderly passengers, a complex fare structure, the number of coins required to pay a basic fare and the use of small buses with only one door. Guenthner and Sinha do not, however, explore these factors further.

The decrease in marginal boarding time is offset by the number of boarding and alighting passengers, and Guenthner and Sinha suggest that a maximum dwell time occurs when the number of passengers concerned is 24. In order to include this, the suggested model is linear if the number of passengers is greater than 24. These models are shown in Equation (3.3):-

$$
\begin{aligned}
T_{dwell} &= p \; . \; [\; 5.0 \; - \; 1.2 \; . \; \ln(\; p \;)] \quad ; \quad p \leq 23 \\
&= 1.2 \; . \; p \quad ; \quad p \geq 24
\end{aligned}
\qquad (3.3)
$$

where

T_{dwell}	=	dwell time per stop
p	=	number of passengers (boarding and alighting)

The empirical evidence of Cundill and Watts, however, tends to suggest that boarding and alighting should be treated separately in terms of the amount of time taken by a passenger to perform these actions.

Lindau (1987) states that in general a constant value for marginal boarding time can be used for any particular number of boarding passengers, for a particular fare collection system and vehicle design. A constant value appears to be obtained in most studies (for example Cundill and Watts (1973), Bly and Jackson (1974), Chapman *et al* (1976), Pretty and Russell (1988)). It does appear, however, that alighting passengers should be included in a model of passenger service time, since these can, under certain conditions, be the determining factor.

3.3.2.2.3 Alighting passengers

A measure corresponding to the marginal boarding time, the marginal alighting time, can be defined as the amount of time taken for each passenger to vacate the departure area of the bus. It is difficult to disaggregate this value any further. Where no mixed door use occurs, alighting becomes important in two circumstances.

First, when the number of alighting passengers exceeds the number of boarding passengers to such an extent that the total alighting time exceeds the total boarding time and becomes the determining factor of the passenger service time. As Cundill and Watts show, marginal alighting times are usually shorter than marginal boarding times so it is important to realise that this effect only occurs when the total time for all alighting passengers exceeds the total time for all boarding passengers.

Second, when passengers waiting to alight obstruct boarding passengers inside the bus. This increases the total time taken for boarding and therefore increases the passenger service time.

Where boarding and alighting takes place through the same door(s) (mixed door use), the situation is slightly different. In this case, boarding may not start until the alighting process has been completed. In this case, the passenger service time is the sum of the total boarding time and the total alighting time.

Whether mixed door use takes place or not, the passenger service time is determined by the door at which the sum of boarding and alighting times is at a maximum.

Taking these situations into account, Equation (3.2) could be altered to allow for all these possibilities:-

$$\begin{aligned} T_{dwell} = \; & t_{dead} + t_{internal} \\ & + \max_{1 \le i \le n} \left\{ \left(K_{board} \cdot p_{board} \right)_i + \left(K_{alight} \cdot p_{alight} \right)_i \right\} \end{aligned} \tag{3.4}$$

where

T_{dwell}	=	dwell time
t_{dead}	=	dead time
$t_{internal}$	=	internal delay
K_{board}	=	marginal boarding time
p_{board}	=	number of boarding passengers
K_{alight}	=	marginal alighting time
p_{alight}	=	number of alighting passengers
i	=	door index
n	=	total number of doors on the bus

The relationships between the marginal boarding and alighting times and the dwell time suggested by this equation do, of course, imply that the bus only leaves a bus stop when all boarding and alighting movements have been completed. In many developing countries, however, this assumption is not necessarily justifiable, since passengers might board, pay the fare and find a seat without the bus stopping at all. This produces a more complex scenario when assessing marginal boarding times and suggests that the time spent by boarding passengers is only relevant insofar as it prevents the bus from leaving a stop and entering the traffic stream.

Having discussed the definition and interpretation of dwell time and its constituent parts, the way in which the dwell time affects the capacity of a bus stop is now discussed.

3.3.3 Bus stop Capacity

3.3.3.1 Introduction

The capacity of a bus system is determined by the bus stop with the highest degree of saturation. The degree of saturation is defined to be the ratio between the number of buses able to pass through the bus stop and the capacity of the stop to accommodate buses. As a result of the relationship (discussed in Section 3.3.2.2.2) between the number of passengers using a stop and the dwell time, the capacity of a bus stop is affected by the number of passengers attempting to board (and alight from) each bus. The capacity of a bus stop, expressed as the number of buses able to be accommodated by the bus stop within a specified time period, is therefore a function of the number of passengers and buses attempting to use the stop.

The capacity of a bus stop is also influenced by the design of the stop (discussed in Section 3.3.4), the design of doors (discussed in Section 3.3.2.1), the fare systems in operation (discussed in Section 3.3.2.2.1) and the associated operating regime of the buses (for example, whether buses may overtake each other or not). Various methods have been proposed for estimating the capacity of a bus stop and these are now discussed.

These methods can be categorized in three ways. First, where there is no control over the way in which a bus operates at a bus stop, this is termed 'non-coordinated operation'. This is described in Section 3.3.3.2. Second, 'coordinated operation' is used to describe the situation where there is some degree of control over the bus operation at a bus stop. This could be the use of overtaking, or the allocation of specific berths to particular bus lines, for example. The effect of such measures on capacity is considered in Section 3.3.3.4. A particular form of coordination is worthy of special mention in the discussion about high capacity systems, so this is treated as a third category. This is convoy operation, which is discussed in Section 3.3.3.3.

3.3.3.2 Existing models for bus stop capacity - non-coordinated operation

The Highway Capacity Manual (Transportation Research Board 1985) suggests that the capacity of a single berth bus stop can be calculated by:-

$$Q_b = \frac{\left(\frac{g}{C}\right) . 3600 . R}{t_c + \left(T_d . \left(\frac{g}{C}\right)\right)} \qquad (3.5)$$

where

Q_b	=	practical capacity of the stop (buses/hour)
g	=	effective green time at the downstream junction (in seconds)
C	=	cycle length at the downstream junction (in seconds)
t_c	=	clearance time per bus (the minimum time between the departure of the rear of one bus and the arrival of the front of the following bus)
R	=	reduction factor to allow for variability in arrival and departure times.
T_d	=	dwell time

If the stop is unaffected by the signals at a downstream junction, (g/C) can be taken to equal 1. R is usually taken to be equal to 0.833 since this suggests that a third of buses are required to queue for access to the stop (Transportation Research Board 1985). R is therefore an allowance for the fact the headways are not precisely equal, which causes the probability to increase that buses may not be able to gain access to the stop as they arrive at the stop area. If arrivals were perfectly regular, R would take a value of 1, and queues would only appear if the bus arrival frequency is greater than the stop capacity. The issue of queues at bus stops is considered in more detail below in this section.

Equation (3.5) therefore compares the amount of time that a bus occupies the bus stop (whether serving passengers or waiting for downstream traffic signals to permit departure) with the total amount of time available, scaled to allow for variations in headway. Equation (3.5), however, includes as a variable neither the number of passengers nor the internal delay, and the dwell time is consequently predetermined. Although differences in demand could be represented by differences in the dwell time, it would be better if the passenger demand were made explicit.

As it stands, Equation (3.5) also allows only for a single berth bus stop. In order to allow for multi-berth stops, the Highway Capacity Manual suggests factors for the scaling of capacity with the use of additional berths. These are discussed further below.

Gibson *et al* (1989a) define the capacity of a bus stop with a linear layout of berths and first-in-first-out (FIFO) discipline (no overtaking is allowed) by:-

$$Q_b = \frac{3600 \,.\, n}{\frac{n}{s} + t} \qquad (3.6)$$

where

Q_b	=	the capacity of the bus stop in buses per hour
n	=	the average number of buses that can enter the bus stop area when this is possible
t	=	the average length of time in which it is not possible for another bus to enter the stop area (blocked time)
s	=	saturation flow of the lane immediately prior to the bus stop area (in buses per second)

Gibson views the operation of a bus stop as a cyclical event, where a cycle consists of a period during which no buses can enter the stop area (because a bus is occupying the berth nearest the entry point) and a saturated period when buses can enter the area. This cycle is represented as the denominator in Equation (3.6). As the number of berths increases, the number of buses that can be accommodated by the stop also increases, and this effect is represented by the numerator (in which the number of berths is converted into bus-seconds and scaled over one hour).

This model therefore treats bus stop capacity as the average number of buses that can enter the stop area in a defined period. The degree of saturation (with its associated effects on the probability of queuing) is, then, the relationship between the amount of time at the stop which is available for the number of buses that can enter the stop area under the prevailing flow conditions and that at the capacity flow, given the average cycle time. The importance of entry to the stop area in this definition of capacity allows for direct consideration of the effect of the stop design on the queue of buses waiting to gain access to the bus stop as well as the delays incurred in attempting to depart.

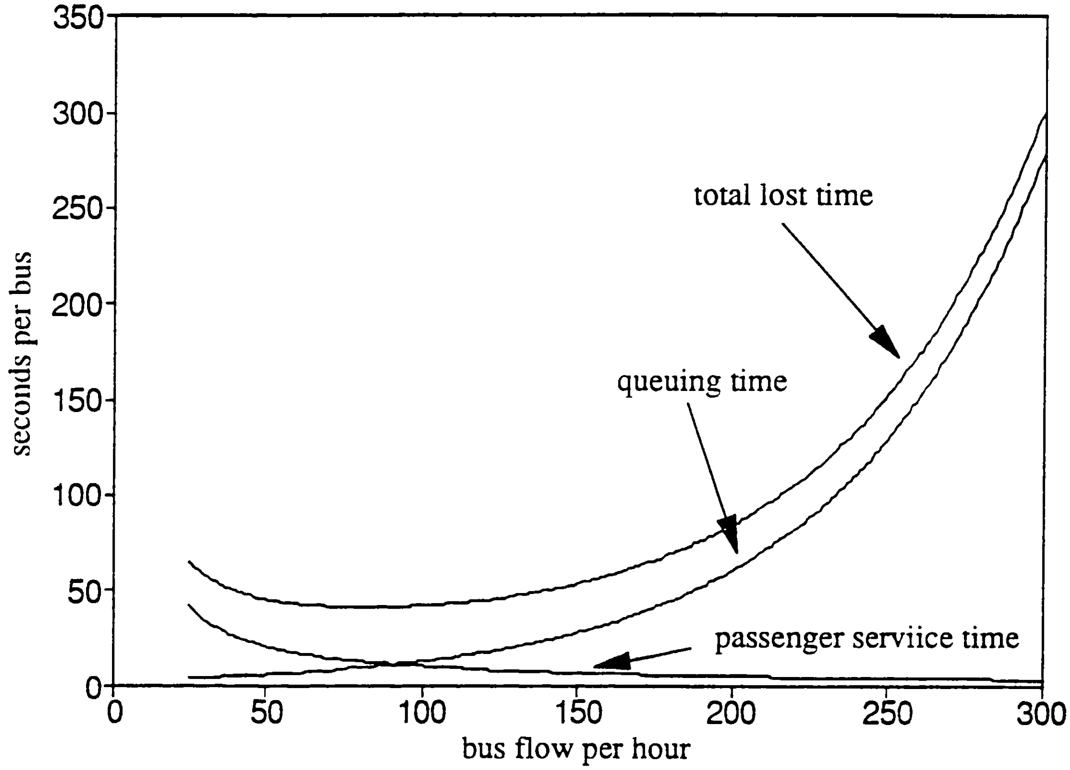


Figure 3.1 Lost time at a bus stop showing the effect of queuing in order to gain access to the stop.

To illustrate the effect of the queuing time at bus stops, IRENE (Gibson *et al* 1989a) is used to simulate the operation of a simple bus stop with two berths. Figure 3.1 shows the results for this example. As the purpose of this exercise is to illustrate the changing effects of various elements of the time a bus spends at a bus stop, passenger demand is held constant while bus frequency is increased. For this purpose, three aspects of the time spent by buses at the stop are considered: the passenger service time, the queuing time (before access to the stop is gained) and the total time incurred by stopping at the stop (total lost time). With a small number of berths and passenger demand held constant, as the bus frequency becomes high, the internal delay becomes almost insignificant. This is because the passenger service time for each bus tends to equality, and buses do not tend to obstruct each other. Internal delay is therefore not included in this graph in order to maintain clarity. The total lost time curve is the sum of the passenger service time and queuing time curves. This curve, however, includes constant values for deceleration, acceleration and dead times, although these are also not drawn on the graph.

Figure 3.1 illustrates the relationship between two elements of the total lost time and the way in which these change as bus frequency changes. This shows that at low frequencies, the dominant element is the passenger service time. If passenger demand is held constant and the bus frequency is increased, the passenger service time (being a reciprocal function of the frequency) decreases accordingly. This has a capacity enhancement effect because of the influence of passenger service time on the dwell time (as discussed in Section 3.3.2 above).

As the bus frequency increases, however, the more dominant element becomes the time spent queuing. Thus as the frequency increases the total lost time (which includes the sum of the two elements) reaches and passes a minimum value. Around this minimum it does not appear to be too sensitive to the bus frequency, but once the degree of saturation reaches a certain point, increased frequency results in a large increase in lost time. The point at which this occurs is difficult to estimate because of the opposing effects on capacity of the passenger service time and the queuing time as bus frequency increases. It is for this reason that bus stop capacity, and therefore the design of bus stops is so important in high capacity bus systems.

Gibson *et al* (1989a) suggests that this cyclical approach to bus stop operation permits analysis of the queuing characteristics using the methods of Kimber and Hollis (1979) and Kimber and Daly (1986). Gibson (1990) states that a reasonable result can be obtained using the time-dependent queuing model of Kimber and Daly, although this is sensitive to the relationship between the distributions of arrival and departure headways at the queue. Gibson (1990) reports that there is a large number of design variables to be considered when evaluating the capacity of a bus stop. These include the number of berths, the proximity of traffic signals (and whether these are upstream or downstream of the stop), the design and operational characteristics of buses and the number of times a bus is likely to stop within the stop area, in addition to the number of passengers using the stop. Gibson states that as a result of this, it is difficult to obtain a value for n (the average number of buses that can enter the stop area when this is possible) analytically since the calculations would quickly become unmanageable. Since t (the blocked time) and n each depend upon the whole process in the stop area it is therefore easier to employ simulation to evaluate their values.

As part of this research, a large number of simulation runs have been made using IRENE. When bus flow is held constant, but the number of boarding passengers is varied, it appears that a negative exponential function, with the number of boarding passengers per bus as the dependent

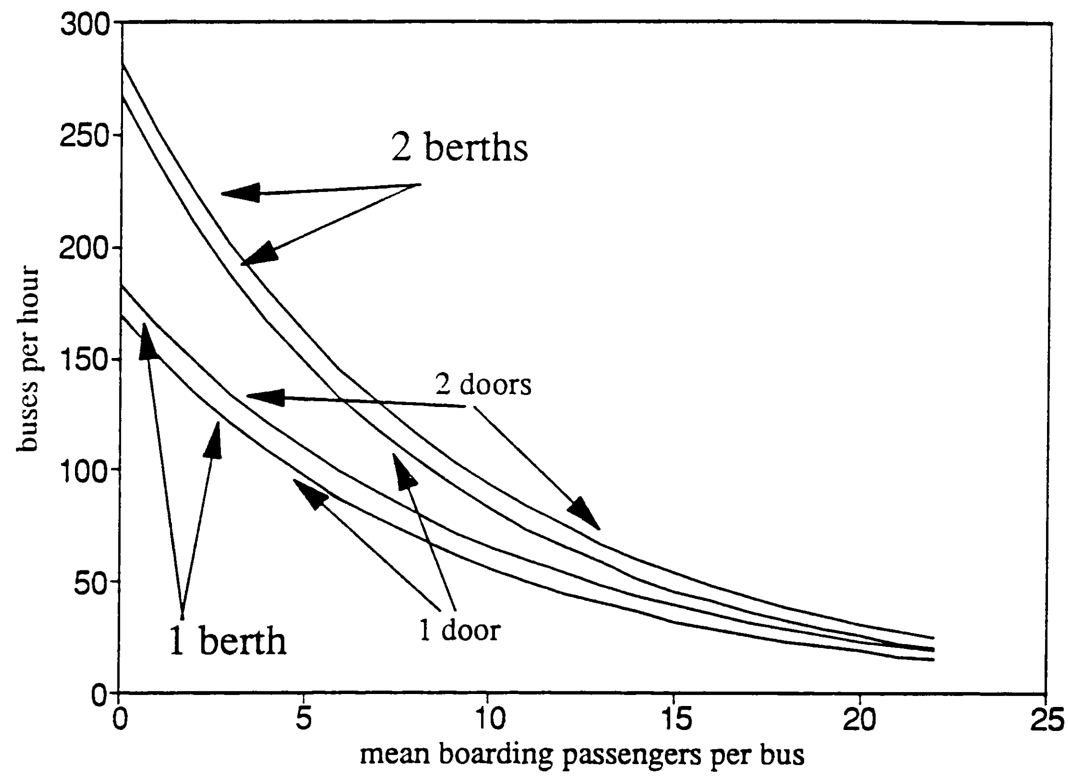


Figure 3.2 Stop capacity simulated using Gibson (1989), as a function of the number of boarding passengers.

variable, could be a reasonable indicator of the bus stop capacity. The resulting curves, for two types of bus and bus stops with one and two berths, are shown in Figure 3.2. Account is taken of alighting passengers through the use of Equation (3.4) in the estimation of dwell time.

3.3.3.3 Existing models for bus stop capacity - convoy operation

Szász *et al* (1978) calculate the capacity of a bus stop under the special operating conditions which exist when the bus stop is used by ordered convoys of buses. Under these conditions (described more fully in Section 3.4 below), overtaking is not possible, but the buses in a convoy can decelerate, stop, open and close the doors and accelerate away from the stop almost simultaneously. Passenger boarding and alighting can also take place more or less simultaneously for each bus in the convoy and in this way the average dwell time per bus is reduced. This reduction is so great that Szász *et al* reports that it was observed in Sao Paulo even though the marginal boarding time increased after implementation of the system.

Since all buses in a convoy do not have an equal number of passengers boarding, an extra element is required in the model to cater for the possibility that there may be extra passengers waiting for one or more of the buses. This effect was tested empirically in Sao Paulo and the resulting expression for the passenger service time is given by:-

$$t_{pst} = \frac{3 \,.\, K \,.\, p_{board}}{2 + C} \tag{3.7}$$

where

t_{pst}	=	passenger service time
K	=	marginal boarding time
p_{board}	=	total number of boarding passengers per hour for all buses
C	=	average number of buses in a convoy

Szász observes that the value of K is slightly greater under convoy conditions than under the previous disorganized regime, but the dwell time per bus in the convoy is less than previously obtained. The time taken by a bus to slow down, stop, open and close the doors and leave the stop is stated to be 8 seconds. A minimum headway of 4 seconds between each bus is added to give a total time for the act of stopping and starting a bus at a bus stop of 12 seconds. With convoy operation, these activities occur almost simultaneously for all the buses in the convoy, so that this portion of the stopping time may be divided between the number of buses in the convoy. The minimum headway remains constant whether buses are operating in ordered convoys or not, so the stopping penalty of 12 seconds is reduced to:-

$$t_{pen} = 4 + \frac{8}{C} \tag{3.8}$$

where

t_{pst}	=	stopping time penalty
C	=	number of buses in a convoy

Using the models expressed in Equations (3.7) and (3.8), Szász gives the following formula (Equation (3.9)) for the maximum theoretical capacity of a bus stop operating under convoy conditions:-

$$Q_C = \frac{3600 - 3 \cdot t_{board} \cdot \left(\frac{p_{board}}{2 + C}\right)}{4 + \frac{8}{C}} \tag{3.9}$$

where

Q_C = maximum capacity of a bus stop under convoy conditions in buses per hour

This formula implies that bus operation takes place at a degree of saturation which is sufficient to provide a continuous flow of convoys, with an average convoy size of C buses. If this is not the case, problems may occur as a result of an increase in the number of boarding passengers per bus. This occurs because p_{board} is treated as a constant and distributed amongst all buses calling at the stop in one hour. A decrease in the number of convoys actually calling at the stop means that the number of passengers per convoy increases. This may magnify any differences between the number of passengers boarding each bus, which could cause the passenger service time to increase - possibly to a greater extent than the decrease in the number of buses would otherwise suggest.

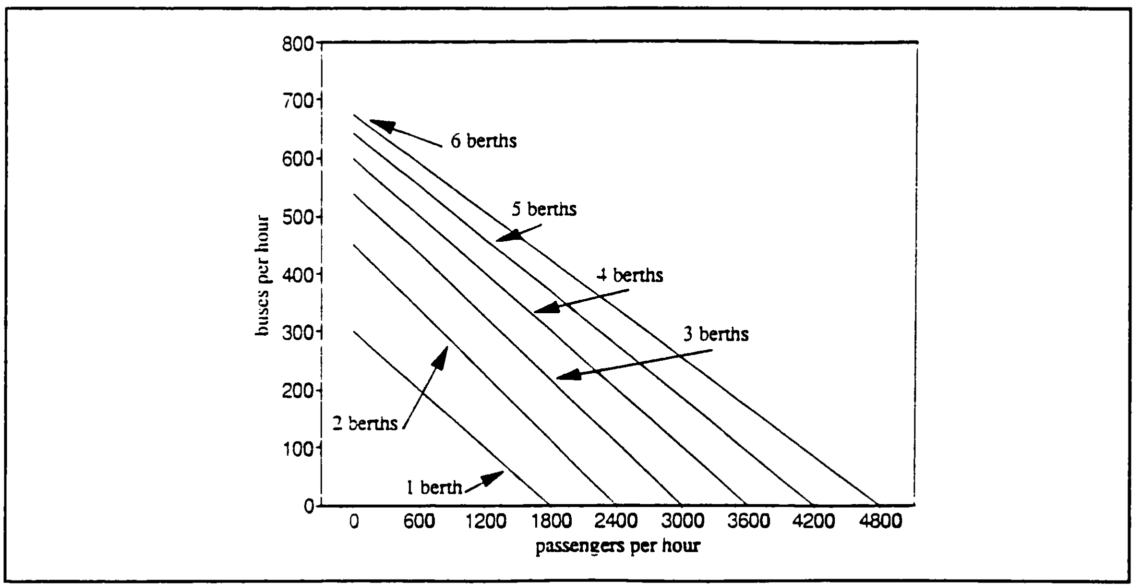


Figure 3.3 Capacity of a bus stop under convoy conditions, with different average convoy sizes. *Source: Szász et al (1978)*

Similarly a reduction in C (the average convoy size) affects both the numerator and the denominator in this expression. This serves to increase the size of the denominator which then tends to decrease the magnitude of Q_c. Since C also appears in the second (subtracted) term of the numerator, this also tends to reduce the value of Q_c. As a result, Equation (3.9) suggests that the capacity of the stop decreases if the convoy size decreases. Figure 3.3 illustrates how the bus stop capacity changes as the number of passengers per hour increases and as the average convoy size changes.

The expression in Equation (3.9) assumes that the convoy frequency is at saturation level, and it does not allow for the eventuality that convoys do not arrive continually at the stop. Thus it is useful mostly as a guide to the maximum possible number of buses that can pass through the bus stop within a specified time period under convoy conditions.

3.3.3.4 Existing models for stop capacity - coordinated operation

In this thesis, coordination of buses is considered to be any measure which constrains the bus driver to operate in a certain way. Such measures include, for example, allocating bus lines to particular berths and combining bus lines into groups which are then treated as single units. The control of the actual route over which a bus line operates, and the frequency required are also types of coordination, but for the consideration of bus stop capacity the determination of these forms of coordination is beyond the scope of this discussion.

The coordination under discussion includes the allocation of bus lines to berths. It is therefore useful to continue the discussion started in Section 3.3.3.1 by considering the capacity implications of the number of berths at a bus stop.

The Highway Capacity Manual (Transportation Research Board 1985) offers some advice on the number of berths required in order to provide sufficient capacity. This consists of reduction factors which estimate the number of 'effective' berths at the bus stop. The number of effective berths is less than the number of actual berths as a result of the fact that it may not always be possible for an empty berth to be filled by a waiting bus (the reasons for this are stated in Section 3.3.4.2.1 below in relation to Gibson's analysis of bus stops in Santiago). The reduction factors given in the Highway Capacity Manual are obtained from empirical observations in a terminal in New York, and some doubt has been expressed over their validity for on-street stops (Gibson

1990). Using the simulation model IRENE (Gibson *et al* 1989a), it is possible to obtain an estimate of the capacity that may be expected under specified operating conditions with different numbers of berths.

Szász (1990) suggests that the effective number of berths which results from an increase in the number of linear berths is not a linear relationship. Additional berths provide a reducing increase in the capacity of the stop. In order to calculate this reduction effect, the average effective use of each berth can be given by the relationship:-

$$R_{berths} = \frac{3}{2 + N} \tag{3.10}$$

where

R_{berths} = the effective capacity of one linear berth

N = the number of actual linearly adjacent berths

Table 3.4 The effective number of berths resulting from a specified number of linear berths. Source: Transportation Research Board (1985)

Number of actual berths	Highway Capacity Manual	Szász	Gibson *et al*
1	1	1	1
2	1.75	1.5	1.12
3	2.25	1.8	1.54
4	2.45	2	1.89
5	2.5	2.14	2.34

The sets of values obtained by each of these methods is shown in Table 3.4. In this table are shown, for various numbers of actual berths, the number of effective berths estimated by the

Highway Capacity Manual and the Szász models, and from the use of the simulation model IRENE of Gibson *et al.* It can be seen that there is a reasonable agreement between the Gibson and Szász models, and that the Highway Capacity Manual appears to estimate a higher number of effective berths than the other models for a given number of actual berths. This difference may be due to the fact that the Highway Capacity Manual uses data gathered from observations of a bus terminal in New York for these estimates, whereas the other models are derived from observations of on-street stopping practices. The reduction in efficiency of a bus stop is an important factor in the estimation of capacity and therefore of the design of a bus system. This is therefore an important area for further research.

If overtaking is allowed, this has a substantial effect on the capacity. In this case convoy operation is not relevant because by definition this does not allow for overtaking to take place. Where overtaking is permitted, a distinction must be made between a bus stop which consists of a number of berths which are linearly adjacent, and one which comprises a number of groups of berths.

Linearly adjacent berths are juxtaposed in such a way that it is only possible to reach a forward berth if all the previous linearly adjacent berths are vacant. Thus the stops considered in Section 3.3.4.2 (and the example illustrated in Figure 3.4) are made up of a number of linearly adjacent berths and the capacity of the stop may be described in terms of Equation 3.6.

A group of berths, on the other hand, is a set of linearly adjacent berths which constitutes only a part of the total number of berths at a particular bus stop. Although each group consists of a number of berths which are linearly adjacent, the groups themselves may be located with a substantial distance between them. This permits the buses from one group to overtake the buses from another group in order to gain access to (or from) their allocated berths. Thus the operation of one group is functionally independent of the operation of all others using the same stop. In this way, buses from one group do not interfere with buses from another group. A design for a bus stop with functionally independent groups of linearly adjacent berths is shown in Figure 3.7.

The capacity of a bus stop with functionally independent groups may be estimated using an adaptation of Equation (3.10):-

$$Q_{stop} = Q_{berth} \cdot \left(\sum_{i=1}^{G} \left(\frac{N_i \cdot 3}{2 + N_i} \right) \right) \tag{3.11}$$

where

G = the number of functionally independent groups at the stop

N_i = the number of linearly adjacent berths in each group i

The important result of Equation (3.11) is that the capacity of a bus stop designed to have a number of independent groups is the sum of the capacities of all the groups. Because the groups are independent, reductions in the effective use of berths only apply to the internal arrangement of each group. Thus once the groups are functionally independent the operation in one group does not affect that in another, and the capacity of each group of berths is therefore also independent.

An example serves to illustrate the effect on capacity of designing a bus stop with overtaking facilities. If a particular operating regime were to provide a capacity of 100 buses per hour at a single berth, four linearly adjacent berths would provide (using Equation (3.10)) a capacity of 200 buses per hour. If the same number of berths were arranged in two groups of two (linearly adjacent) berths, the capacity would be (from Equation (3.11)) 300 buses per hour.

A bus stop with a defined number of functionally independent groups can be considered as independent sets of linearly adjacent berths with the appropriate flows of buses attempting to gain access to their allocated berths. The capacity of each group can then be estimated using Equation (3.6). As mentioned in Section 3.3.3.1, the values of the variables in Equation (3.6) are difficult to evaluate analytically, and therefore this is facilitated by a simulation process. IRENE (which allows for more accurate representation of the design parameters than the Highway Capacity Manual) may be used for this procedure. The capacity for the whole stop is then the sum of the group capacities, using Equation (3.11) which allows for the reduction in capacity which arises as a result of the use of linearly adjacent berths within each group. Thus in order to estimate the capacity of bus stops where bus operation is coordinated, it is possible to use a model devised for the analysis of non-coordinated bus stops, with the appropriate parameter values.

Having discussed the calculation of bus stop capacity, it is now necessary to consider the ways in which the capacity, as calculated, may be enhanced in practice. The discussion therefore now turns to the question of bus stop design.

3.3.4 Bus stop design

When considering high bus flows, the key to bus system design lies in the design of the bus stops. In Brazil, for the reasons stated in Chapter 2, these flows are not uncommon, and it is essential that a bus system can deliver the appropriate capacity at the stops. The capacity effects of various operating strategies have been discussed in Section 3.3.3. This section considers ways in which bus stops may be designed in order to facilitate the operation of these strategies. Bus stops at which buses are not able to overtake each other are considered first, and the differences in design which arise as a result of overtaking are considered later.

3.3.4.1 General issues

Gibson *et al* (1989b) give four factors which influence the operation of a bus stop:-

1 Characteristics of the buses;
2 Demand at the stop (passengers and buses);
3 Method of operation of the stop; and
4 Physical design of the stop.

The first includes aspects of the design of buses such as the number and width of doors as well as operational elements such as the use of the doors and the fare collection system. These have been discussed in Section 3.3.2 above. The effects of passenger and bus demand on the capacity of a bus stop have been discussed in Section 3.3.3 above.

The method of operation at a bus stop has an important effect on the design of the stop Gibson *et al* (1989b) give five operational characteristics which can influence stop design:-

1 **whether stops are compulsory** - if not, some provision must be made for non-stopping buses;
2 **berth assignment** - whether berths are allocated to particular bus lines (or a group of lines) or whether any bus may stop at any berth;
3 **stopping regime** - if more than one berth is assigned to a particular group of lines, it could be more effective to instruct the drivers to stop in the berth nearest to the exit whenever this is possible. If the berths are not assigned, it may occur

that a bus will stop more than once in the stop area to allow boarding and alighting;

4 **entry discipline to the stop area** - whether a bus is constrained to wait for a preceding bus to leave a berth before it is possible to enter the stop (FIFO), or whether it is possible to overtake a stopped bus before obtaining access to a berth;

5 **exit discipline from the stop area** - it may be possible for a bus to leave a berth when it has completed its business at the stop, or it may be necessary to wait for other buses to depart.

It is through the interaction of these operational methods that the physical design of a bus stop can be altered to enhance its capacity. Of all these factors, only demand is variable with time; the others are all constant for a particular situation. Thus it is possible to measure not only how a bus stop performs as the demand varies, but also how this performance can be changed by an alteration to the design of the stop. Gibson *et al* (1989b) do not advocate a particular design for a bus stop, rather that the design depends on the particular combination of these factors which applies in a particular situation.

3.3.4.2 Bus stops with no overtaking

Bus stops which do not allow overtaking are those in which a bus may be dependent on the activities of another bus for its access to the stop or for its ability to leave when free to do so. Thus a bus may be prevented from leaving a stop when boarding and alighting have been completed because another vehicle which is still loading is physically obstructing the exit path. Similarly, a bus may be unable to enter the stop area because another bus is occupying the berth nearest to the entry even when other berths are unoccupied. The combination of these two effects tends to result in a reduction in the capacity of the bus stop which has been discussed in Sections 3.3.3.1 and 3.3.3.2 above.

Bus stops may be designed in order to ameliorate this reduction in capacity, by increasing the number of berths or by defining a particular operating system to reduce the probability of inter-bus interference. An interesting example of the latter is the convoy system described in Section 3.3.3.3 above. Before considering bus stop design for convoys in Section 3.3.4.2.2, however, the design of stops for non-coordinated bus operation without overtaking is discussed.

3.3.4.2.1 Non-coordinated operation

In this case, the bus stop consists of a number of berths arranged in a linear sequence (whether at the kerbside or in the median). Berths for non-coordinated operation are available for any bus to use. In many cases, berths, as such, are not delineated and are only defined informally by the bus drivers. In these informal conditions, the buses often interfere with each other and as a result cannot reach or leave the stop without previously waiting in a queue. The effects of this on bus operation have been discussed in Section 3.3.3.2.

Passengers may attempt to alleviate the problem by moving from the bus stop area towards their chosen bus: Szász *et al* observes instances of passengers being required to walk 100 metres in order to find their bus. The capacity of the stop is also reduced. Figure 3.4 shows a bus stop with no facilities for overtaking. The berths are marked in this illustration for reasons of clarity - such markings may not exist in reality.

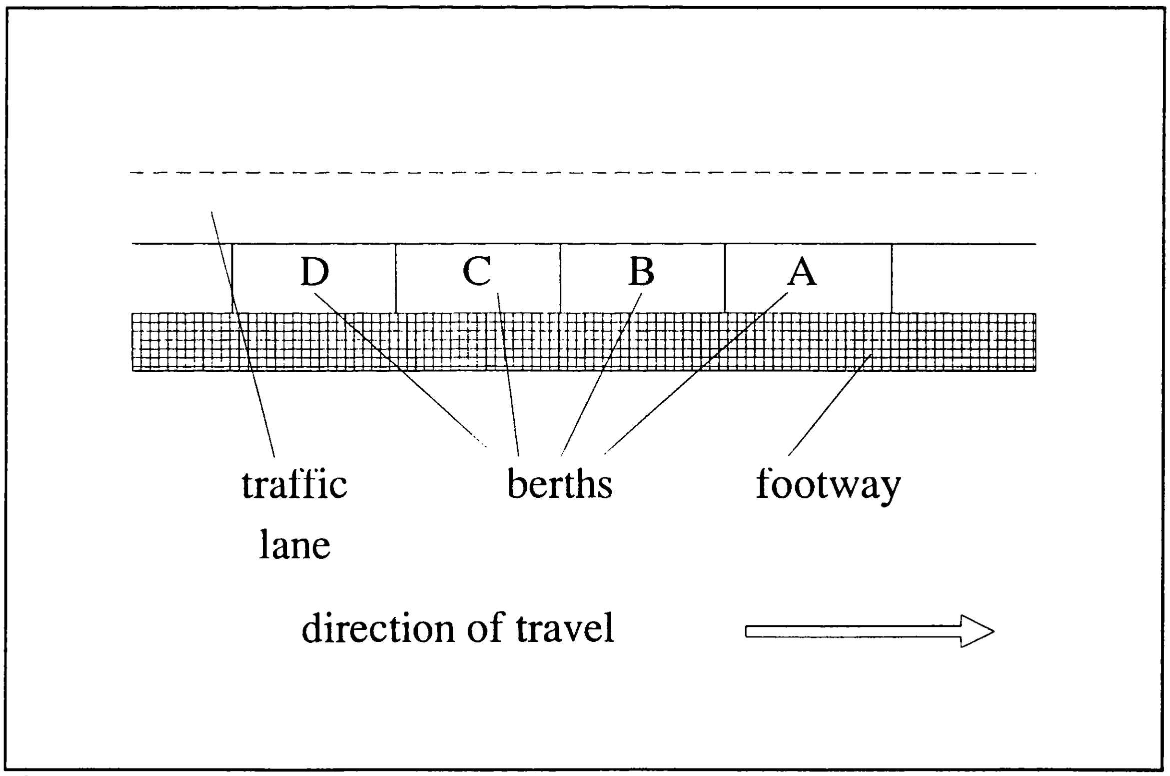


Figure 3.4 A kerbside bus stop with no overtaking facilities. Four berths are shown.

EBTU recommends the dimensions for a single berth kerbside bus stop such that a length of 10 metres is allowed for entry to the stop, 8 metres for exit, a bus length and a two-metre length to be positioned at the rear of the bus for rear-boarding buses and at the front of the stop area if boarding takes place at the front of the bus. A bus stop with one berth is illustrated in Figure 3.5. This stop is 33 metres long and is intended to allow a 13 metre bus free access to and from the kerb.

Access to a berth under non-coordinated conditions is limited by two factors: a berth may not be entered if it is already occupied, and it may not be entered if there is an occupied berth through which a bus would have to pass in order to reach the desired berth. This restriction exists because it is not possible for one bus to overtake another in order to attain an empty berth in front of the one that is occupied. In high demand areas, the high flow of buses may spill over into an

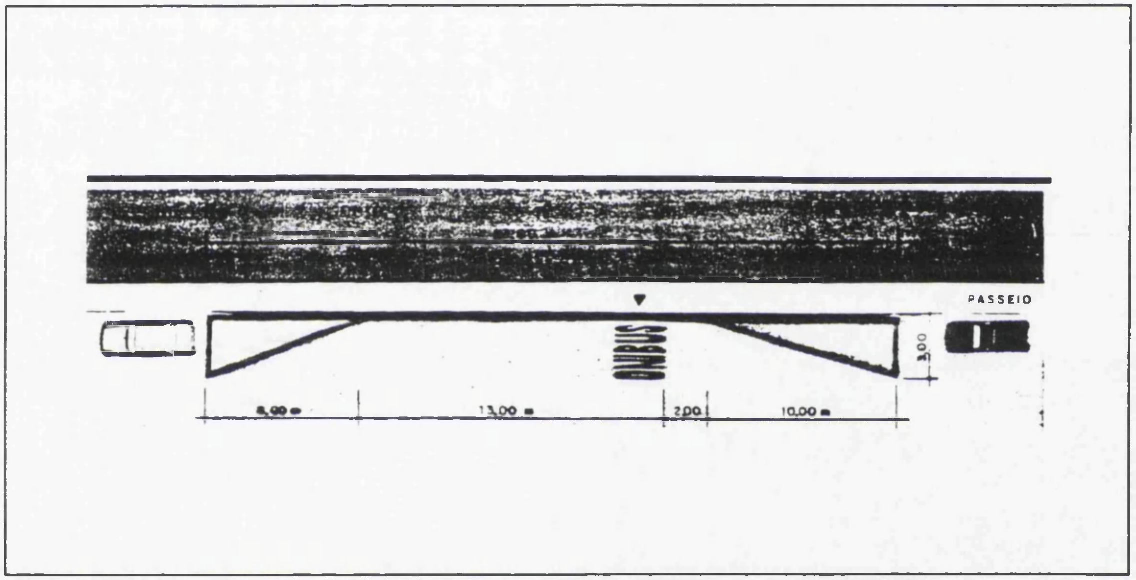


Figure 3.5 Bus stop design for a single berth bus stop. *Source: EBTU 1982*

adjacent traffic lane, and it is possible that a bus driver uses this lane for overtaking in order to reach an available berth. It is also possible that this lane is used to allow passengers to board and alight: a frequent sight in Santiago is the use of two lanes for stopping buses with a third used for overtaking (sometimes also for boarding and alighting). Another example of this, in Bangkok, has already been discussed in Section 3.2.2.1. This type of operating practice is clearly unsafe, and has a devastating effect on the traffic capacity of the corridor (especially where the stop area is near to a junction).

The reason for this practice is that the capacity of the stop area is insufficient for the demand, and drivers are attempting to enhance the capacity by using adjacent lanes for overtaking and passenger service. It is clearly preferable, for reasons of both safety and general traffic capacity, to try to design a bus stop with sufficient capacity. Not surprisingly, this may involve permitting overtaking and, in some cases, stopping at parallel locations. One way of achieving an increase in capacity is to coordinate the buses so that those elements of their operation that cause interference occur for different buses at the same time. This reduces the interactions and therefore the associated delay. This was attempted in the design of the ordered convoy system.

3.3.4.2.2 Ordered convoys

The operation of an ordered convoy bus system is discussed in detail in Section 3.4. At this point, however, it is useful to consider the design implications of the bus stops when convoy operation is used.

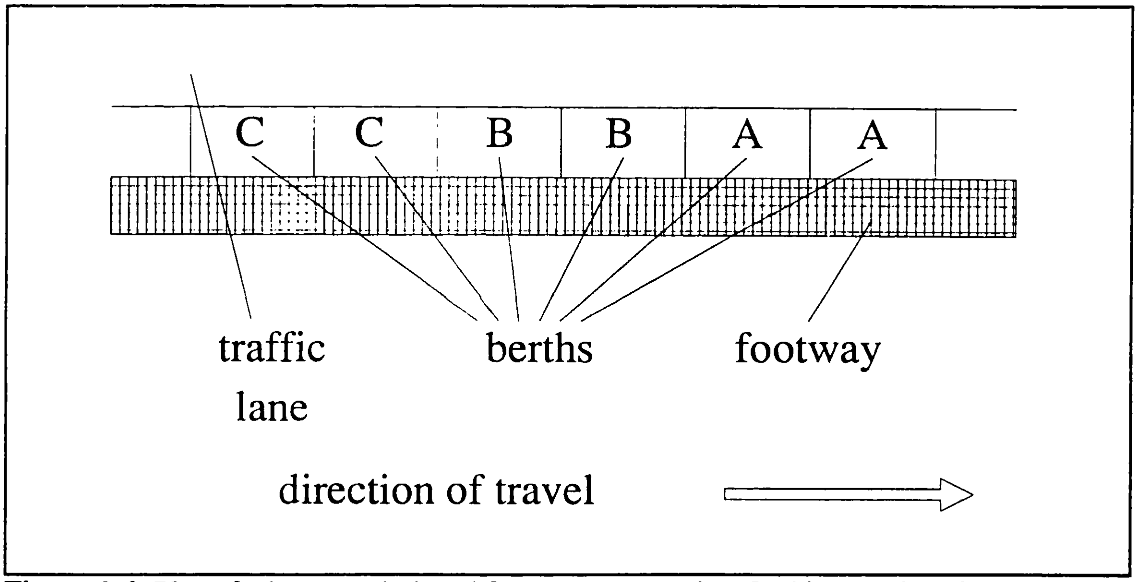


Figure 3.6 Plan of a bus stop designed for convoy operation. In this case three bus groups are allowed, with two berths allocated to each group.

Figure 3.6 shows a bus stop designed for use by an ordered bus convoy, as used in Sao Paulo. The convoy is divided into a number of groups (in this case three). The number of groups is dependent on the relationship between the flows of passengers and buses. It is interesting to note that the convoy stop shown in Figure 3.6 is 77 metres long. This is a little over twice as long as the single berth stop illustrated in Figure 3.5. This is because the convoy stop is situated in a bus lane and therefore does not require additional space for access to the kerb from a traffic lane. When the stop is located in the median, and a platform has to be accommodated (for example as shown in Figure 3.7), sufficient length must be allocated for this manoeuvre.

The bus stop shown in Figure 3.6 is intended for a convoy of up to 6 buses, divided into three groups and therefore consists of six linear berths. The total length of the bus stop platform required for each direction is therefore:-

$$L_{stop} = (B \cdot L_{bus}) + ((B - 1) \cdot L_{gap}) \tag{3.12}$$

where

L_{stop}	=	length of the bus stop platform in metres
B	=	number of berths required
L_{bus}	=	length of buses using this stop
L_{gap}	=	length of gap between buses when standing at the stop

EBTU (1982) gives some iso-saturation curves to indicate the levels of saturation at which the number of groups should be changed. These are based on the premise that the bus stop should be designed so that the degree of saturation does not exceed 0.8, and that a degree of saturation of less than 0.4 indicates that the stop is not being used sufficiently. In this context, the degree of saturation is calculated (Szász 1990):-

$$X = \frac{t_{lost} \cdot F + K \cdot p_{board}}{3600} \tag{3.13}$$

where

t_{lost}	=	lost time per bus including minimum headway, but excluding acceleration and deceleration time
F	=	bus flow in buses per hour

The number of bus lines within each group depends on the frequencies of all the lines concerned. In the case of Porto Alegre (Brazil) Szász (1980) indicates that the combined frequency of all buses within each group should be approximately equal for all groups. It is possible to have a range of up to 15 per cent of this value between the group with the lowest and that with the highest frequency. The destinations of the bus lines within a group should be similar.

In Sao Paulo, the convoy system was located at the kerbside and the amount of road space taken by the stop was simply the width of the bus lane. In Porto Alegre, however, the bus stops were located in the centre (median) of the roadway and extra space is required for the provision of platforms for the passengers. In order to reduce the total width required the platforms are staggered as shown in Figure 3.7.

3.3.4.2.3 Number of berths

As discussed in Section 3.3.3.2 above, the capacity of a bus stop is limited where it is not possible to accommodate overtaking. Ways to increase the capacity have been described in Section 3.3.3 and include the use of more berths and operational systems such as convoys. As Gibson *et al* (1989b) note, the capacity of the stop is reduced if buses are allocated strictly to particular berths at the stop: capacity can be enhanced simply by encouraging the bus drivers to use the berth nearest the exit because this leaves more accessible vacant berths behind the bus for other vehicles. Where a large number of bus lines is using the stop, this could be confusing for passengers, however, and in this case thought should be given to the allocation of berths to individual or groups of bus lines.

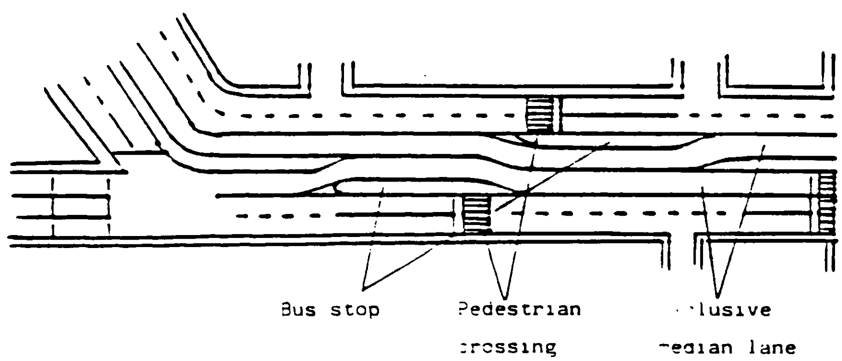


Figure 3.7 Convoy bus stops located in the median in Porto Alegre. *Source: Lindau 1987*

The grouping of bus lines enables stops to be designed with a higher capacity with a smaller number of berths. This is because the reducing effect noted in Section 3.3.3.4 which reduces the effectiveness of successive berths does not apply to functionally independent groups. A theoretical example has been used in Section 3.3.3.4 to illustrate the increase in the stop capacity which may be achieved through the use of groups. It is interesting to note that the double group, double berth stop in this example, which has four actual berths, has (from Equation (3.10)) 3 effective berths. Four linearly adjacent berths would produce only 2 effective berths. Although the constant values in Equation (3.10) preclude the possibility of calculating the value of actual berths necessary to obtain 3 effective berths, simulation runs with IRENE suggest that at least six linearly adjacent berths would be necessary.

The convoy system resolves this problem as it allocates the berths to groups of bus lines, but it always ensures that the order of buses arriving at the stop is the same as the order of berths.

3.3.4.3 Bus stops with overtaking

Bus stops without overtaking, therefore, are likely to have a severe restriction on their overall capacity. Section 3.3.4.2 has discussed one way of alleviating this restriction through the use of convoy operation, which is discussed in more detail in Section 3.4.2. Section 3.3.4.2 has suggested that it is possible to enhance the capacity of a bus stop if it is possible for buses to overtake each other as a normal part of their operation at the stop. This enhances capacity by permitting the bus groups to operate independently of each other. For this to be achieved, it is necessary to design bus stops with overtaking facilities. This is discussed in this section.

In order to ensure that the groups are functionally independent, it is necessary to allow a bus to enter or leave its allocated berth independently of buses from other groups. This requires an overtaking facility (which is uninterrupted by other traffic flow) and sufficient space for the relevant turning manoeuvre. Two manoeuvres are considered in order to establish the minimum amount of space required. The first is the exit manoeuvre of a bus attempting to leave its berth and overtaking a bus stopped in the rear berth of the next group. The second is the entry manoeuvre of a bus which must pass a bus which is stopped in the front berth of a group before turning to gain access to its allocated berth.

A bus which passes a stopped bus before stopping at its own berth requires more space for this manoeuvre than the bus which is required to pull out from its own berth to pass a bus stopped further along the bus stop. This is because the initial speed from which each manoeuvre takes place is different. In the former case the bus will be making the manoeuvre from a running speed even allowing for deceleration into the bus stop. In the latter case, the bus is stationary prior to making the manoeuvre.

Engineers in Sao Paulo observed drivers attempting each manoeuvre, and obtained an approximate value for the distance to separate the two groups of buses. This was found to be about 18 metres for the former manoeuvre and about 11 for the latter. These figures are notable because they seem to apply for both standard 12 metre buses and the 18 metre articulated vehicles. The articulation seems to be well designed in this case. Minibuses were not tested.

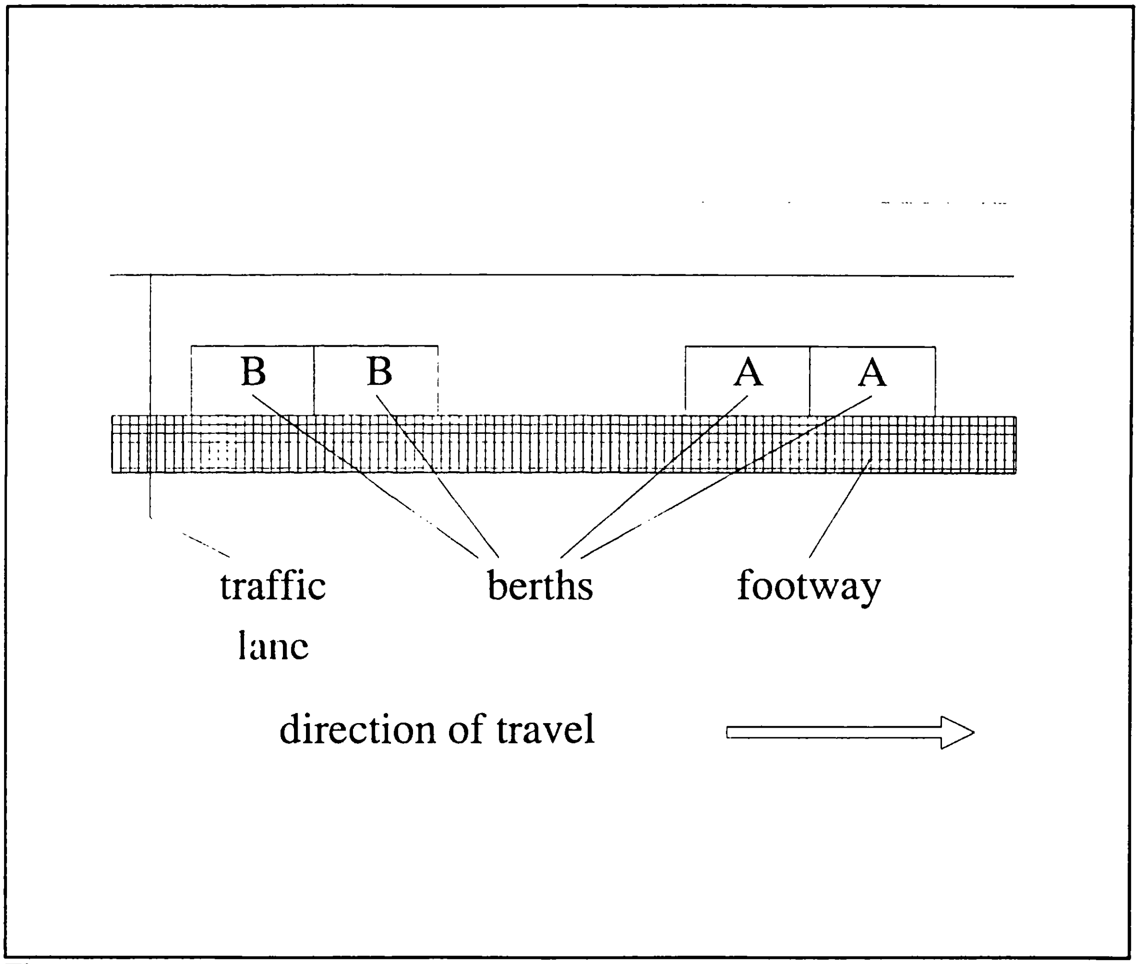


Figure 3.8 A double group, double berth sequential stop.

In order to reduce the number of potential conflicts, the Brazilian designers tend to position the berths for the group of bus lines with the highest flow at the beginning of the stop.

Groups may be arranged sequentially as shown in Figure 3.8 or they may be arranged in parallel as shown in Figure 3.9. Care should be taken with parallel stops that the groups are functionally independent, especially where the inner lane does not permit overtaking.

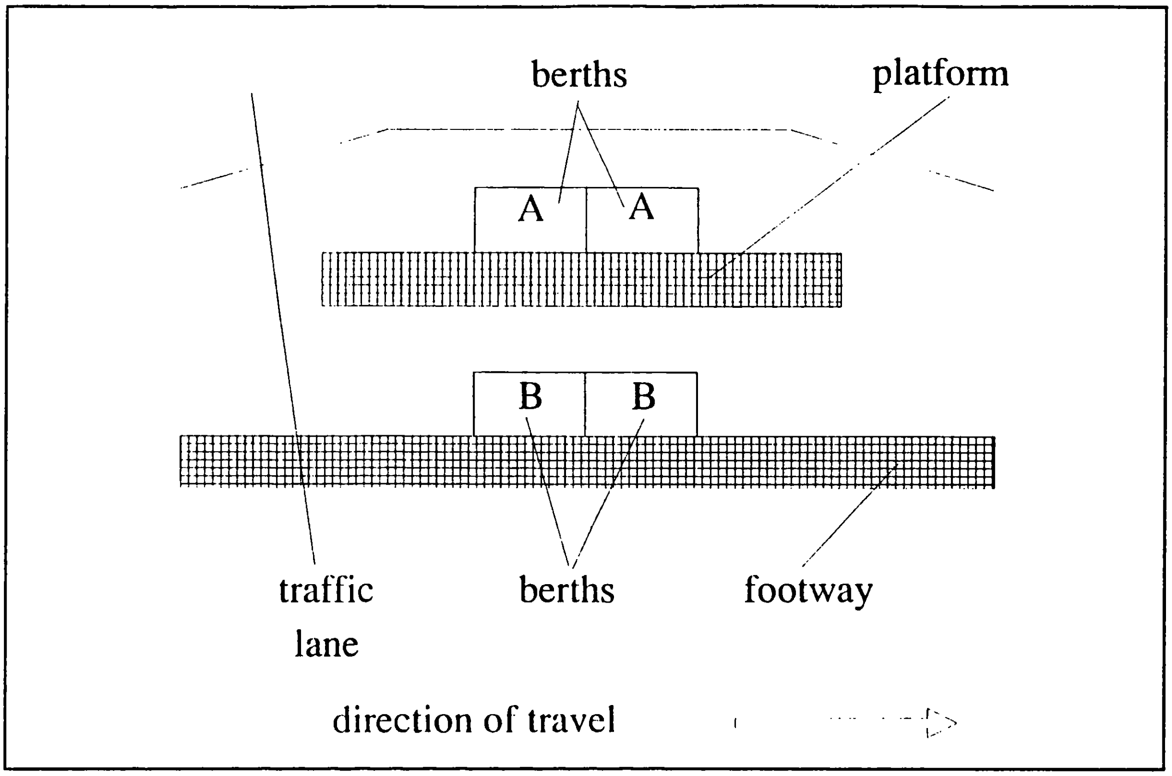


Figure 3.9 A double group, double berth parallel stop

3.3.5 Location of bus stops

As can be seen from the discussion so far, a major factor in the determination of the capacity of a bus stop is the number of passengers boarding buses at a bus stop. This can be affected by three factors: the passenger demand, the bus flow and the distance between the bus stops. As mentioned in Section 3.2.2, commercial speed is often used as an indicator of the performance of a high capacity bus system. This is because it is affected by the time spent by a bus when stationary (whether at bus stops or whether involved in other congestion). As discussed in Section 3.3.4, this is determined by the stop design, but it is also affected by the number of stops made by the bus as it travels along the corridor. This is, in turn, a function of the location of bus stops along the corridor.

3.3.5.1 Stop location effects on commercial speed

Table 3.5 shows some results from a study in Milton Keynes (Hutton and Clark 1989). This shows a reduction in commercial speed of 4.73 km/h for hail and ride operations when these are compared with the mean speed for the whole route (including hail and ride sections where no stop was requested). When compared with the average data for all minibuses which appear in the third column, the reason for this difference becomes evident. The number of passengers per stop is much lower in the hail and ride sections than in the system average. The difficulty which arises from this is that the amount of time spent stopped at bus stops as a proportion of the journey time is not reduced proportionately. It seems that although the stop time per passenger (which presumably is the dwell time divided by the number of embarking passengers) is roughly equivalent, the number of passengers embarking at each stop is much less. This increases the number of stops, thus increasing the amount of time spent accelerating and decelerating. This causes the reduction in commercial speed.

Table 3.5 Comparison of operational characteristics of hail and ride operation with conventional bus services in Milton Keynes. *Source: Hutton & Clark (1989)*

	sections with hail & ride	hail & ride sections with stops	all minibuses	large bus
% time stopped	13.4	15.9	17.7	23.4
stoptime per stop		17.4	26.4	32.4
stoptime per passenger		9.6	10.2	10.8
passengers per stop		1.83	2.58	2.92
commercial speed	28.63	23.9	28.39	27.37

This relationship between the number of stops and commercial speed was not quantified by Hutton and Clark, but attempts have been made to do so elsewhere. Both Cohen (1984) and CADE-IDEPE (1988) show a strong negative relationship between commercial speed and overall stop frequency in such different environments as Besançon in France and Santiago in Chile. This can be expressed (Gibson *et al* 1989a) by:-

$$V_c = V_o . e^{(-\alpha . F_s)} \qquad (3.14)$$

where

V_c = commercial speed

F_s = number of stops (for any reason) per kilometre

V_o, α = parameters for calibration

F_s is in effect the sum of the number of commercial stops and stops due to congestion per unit distance. Cohen points out that with the calibration made in Besançon (V_0=32, α=0.122), a reduction by one third of non-commercial stops would increase the commercial speed by 1 kilometre per hour, while if commercial stops are fixed, an increase of 10 per cent in the commercial speed would necessitate a reduction in the frequency of non-commercial stops of 75 per cent. Cohen makes the point that the problem in Besançon is mainly concerned with intersections, which accounted for 20 per cent of bus journey time. The bus priority system introduced there set out to reduce this delay and after implementation intersection delay accounted for less than 9 per cent.

In developing countries, where the commercial stops are less fixed, and a bus may stop several times at each bus stop to collect passengers (each of which is technically a commercial stop), this relationship is more dependent on the control of commercial stops. Gibson *et al* (1989a) suggest that differences between parameter values obtained in the calibration in Santiago with those obtained for Besançon (the Santiago values corresponding to those given above for Besançon are V_0=35, α=0.144) can be explained by the different environment to be found there. The postulation is that the higher value for V_0 could be explained by different vehicle driver characteristics and lower car flows, while the higher value for α may be due to greater activity and congestion at bus stops, thus increasing their effect on commercial speeds.

The following example from Santiago quoted in Gibson *et al* (1989a) indicates the difference in the problem from that in Besançon. In a link where the stop frequency is 10.4 stops/km, 4 stops/km are attributable to the setting down and picking up of passengers and a further 4.7 are caused by congestion at bus stops. The commercial speed in this section is less than 8 km/h.

The negative exponential model is appropriate for these purposes because of the relationship between the two parameters. The parameter α is a measure of the effect on the commercial speed of the act of stopping: as the stop frequency increases, this effect becomes smaller since the maximum running speed becomes lower as a result of the smaller distance between the stops. This parameter therefore affects the degree of curvature of the negative exponential curve. The parameter V_0, however, is a measure of the free (or running) speed. Where this is high, the effect of stopping is great, causing a large drop in commercial speed for small increases in stop frequency. This effect is represented by the form of the negative exponential curve, with the degree of the stopping effect represented by the value of α. The curves obtained for the Besançon and Santiago are shown in Figure 3.10.

3.3.5.2 Stop location effects for passengers

3.3.5.2.1 Introduction

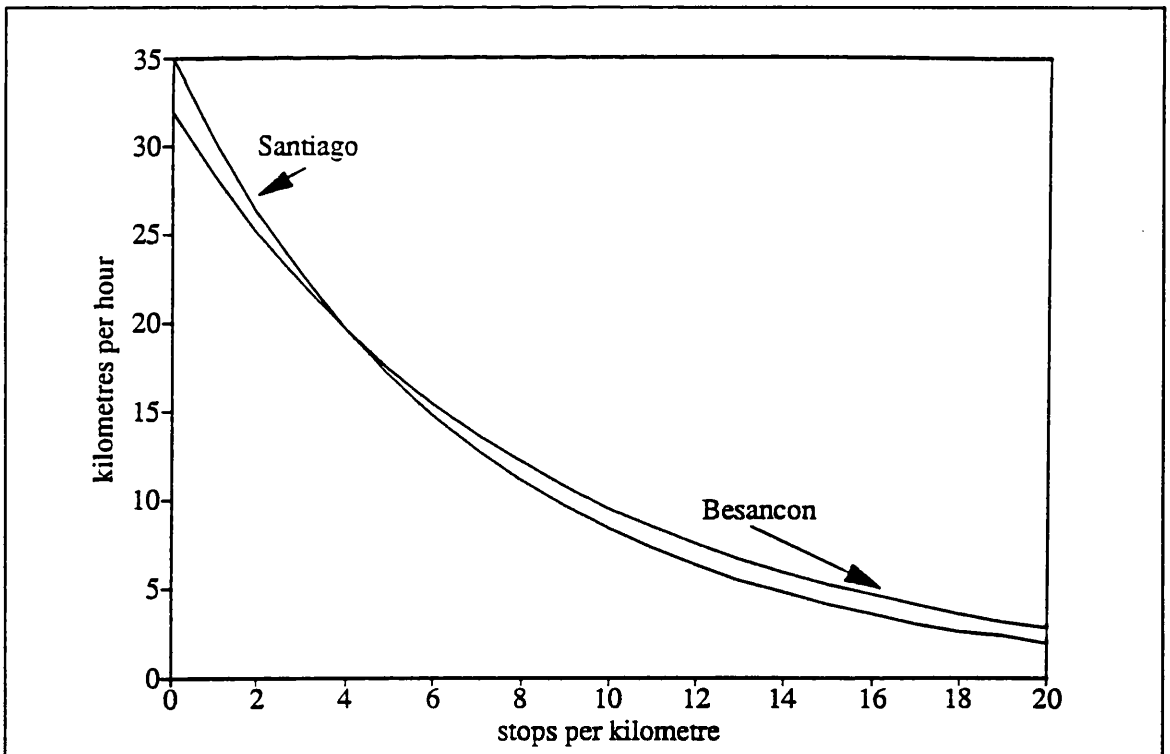


Figure 3.10 Negative exponential model for commercial speeds in Besançon and Santiago. *Source: Cohen (1984), Gibson et al (1989)*

One of the first aims of a bus system designer is therefore to regularize and reduce the overall stop frequency. The spacing between bus stops is a trade-off between the desires of the operators and the requirements of the passengers. As stop spacing increases, the operating efficiency increases. As the spacing increases, however, passengers are required to walk further to reach the bus stop, so that the benefit for these passengers decreases. The exact relationship between these factors is now explored.

3.3.5.2.2 Access distance

The distance between bus stops affects the access distance for passengers. Most stop location models assume that the street network is a grid pattern and that passengers therefore walk in parallel with, or orthogonally to, the corridor. For a corridor considered in isolation, (so that the proximity of access points in other corridors is excluded) access distance is calculated as the distance parallel to the corridor that it is necessary to walk in order to reach a bus stop. For this reason, access distance has a maximum value of half the distance between the stops, with an average of half the access distance.

Investigations have been made into the shape of the access area (Danas 1983), and into the exact points between stops at which a passenger chooses to walk in the direction of travel to the next stop, or to walk in the reverse direction towards the previous stop (Vuchic and Newell 1968). The former work requires extensive knowledge about the nature of the area surrounding the corridor, and is therefore not considered further here. The latter work is now considered.

3.3.5.2.3 Allocation of passengers to stops

Vuchic and Newell consider the optimal spacing between stations on a commuter rail line. It is assumed that passengers choose the station at which they board their train on the basis of the shortest overall travel time. There is therefore a point between stations at which the sum of the time taken to walk to the station (k) and take the train to the next station ($k+1$) is equal to the time taken to walk to station $k+1$. This point is not half way between the two stations because of the difference between walking speed and the running speed of the train.

The distance along the corridor between two stations k and $k + 1$ is denoted S_k and is divided into two parts, G_k and H_k, where G_k is located after station k and H_k before station $k + 1$ (in terms of the direction of travel). The time taken to reach station $k + 1$ is the time of interest, since the objective is to locate the point at which the subsections H_k and G_k meet. Assuming contiguity between the sections, the lengths of the subsections G_k and H_k can be established relative to the position of stop k as a function of the travel time:-

$$\frac{G_k}{V_a} + t_f + T_k = \frac{H_k}{V_a} + t_f \tag{3.15}$$

where

V_a	=	access speed
t_f	=	transfer time at the station
T_k	=	travel time from station k to station $k + 1$

Since S_k is the sum of G_k and H_k, these can be determined by:-

$$\begin{aligned} G_k &= \frac{1}{2} \cdot (S_k - V_a \, . \, T_k) \\ H_k &= \frac{1}{2} \cdot (S_k + V_a \, . \, T_k) \\ &= G_k + V_a \, . \, T_k \end{aligned} \tag{3.16}$$

Thus assuming a uniform distribution of passengers along the section, the maximum access distance is H_k (for passengers using stop $k + 1$) and the average would be a quarter of the section distance S_k. It should be noted that the difference between G_k and H_k becomes more pronounced as the travel speed of the trains between the two stations becomes lower.

In Brazil, the bus system designers are particularly interested in the operational efficiency and tend to increase the stop spacing to about 500 metres. As the spacing between bus stops tends to be smaller than that typically found between stations on commuter rail systems, the Brazilians do not use the Vuchic and Newell formulation to determine the source of passengers for each stop. Instead, they simply use the half way point between the two stops as the dividing line, giving a maximum access distance of 250 metres. With an access speed of 1 metre per second and a travelling speed between the stops of 20 kilometres per hour, the Vuchic and Newell model (Equations (3.15) and (3.16)) gives values of approximately 205.4 metres for G_k and 294.6 metres for H_k. If passenger density is constant along the corridor this is not significant, since the decrease on one side of a stop is balanced by the increase on the other. If demand is not constant, or if the length of the sections S_k varies, this could affect the modelling of demand at the bus stops. In these circumstances therefore the extra calculations required for the Vuchic and Newell model would seem to be worthwhile.

This design feature is reinforced if the bus lane is located in the median of the road, thus reducing the possibility of passengers hailing a bus between stops. This system seems to be less successful in Perú, for example, where the buses often do not use the (Brazilian-designed) bus lane until the buses are full. This is apparently in order to be able to service passengers more frequently at the kerbside. The full significance of this to the design process will be discussed below.

3.3.5.3 Stop location models

By reducing the number of commercial stops the commercial speed is increased, and the question of increasing the efficiency of the remaining stops can be addressed. The principal problem concerning Szász *et al* (1985) is the conflict occurring between buses as a result of uncoordinated arrivals at the stop. This conflict reveals several problems. First, the buses are unable to reach the stop because previous vehicles obstruct the entry to the stop area. Second, in order to board these buses, the passengers must walk towards the vehicle (Szász observed walking distances in excess of 100 metres). Third, buses in the queue are unable to leave when boarding is complete because leading buses are still loading. In each case, the dwell time tends to increase as a result of the particular problem. This naturally tends to reduce the commercial speed. It is therefore of interest to find the stop distance which minimizes the time which is lost at bus stops, the time spent by passengers walking to or from stops and the time spent waiting for buses.

A number of models have been presented with the objective of obtaining a good stop spacing, and these are now discussed. This objective may be seen as the reduction of journey time (for example, Vuchic and Newell 1968) or as the maximization of ridership (for example Vuchic 1969) or as some form of cost-minimization (for example CADE IDEPE 1988).

Vuchic (1969) and Vuchic and Newell (1968) present models to optimize the stop spacing for a single commuter rail line with one terminal station to which all passengers travel. This is not usually the case with urban bus operations which are more likely to call at a number of trip generators and attractors, and is therefore not pursued here. Nevertheless, a major conclusion of this work is that stop location is a function of the passenger density (passengers per unit distance) along the corridor.

Lesley (1976) presents a model which optimizes bus stop spacing by minimizing the generalized cost for passengers. This is based on a circular catchment area with a radius of half the optimum stop distance. The optimum stop distance, d, is given by:-

$$d = 2 \cdot \left(\sqrt{\frac{\frac{l_v \cdot v}{4} \cdot \left(\frac{1}{\alpha} + \frac{1}{\beta} \right) + \frac{l_v \cdot t_l}{2}}{\frac{l_v \cdot f \cdot g \cdot \pi \cdot t_b}{2} \cdot \frac{8}{3 \cdot \omega}}} \right) \tag{3.17}$$

where

l_v	=	average in-vehicle trip length of passengers
v	=	average operating speed of buses
f	=	scheduled service frequency
g	=	rate of trip generation per area per unit time
ω	=	average walking speed of passengers
t_l	=	constant lost time at each bus stop
t_b	=	marginal boarding time
α	=	acceleration rate of buses
β	=	deceleration rate of buses

This assumes that all stop spacings along the route are equal, and that the demand generation is constant for all stops. Thus the passenger demand per stop is a function of the area of the circular catchment areas and the trip generation rate, g, and is equal for all stops. If this increases (as a result of an increase in g, for example), the resulting stop distance becomes smaller. A point of interest in this model is that the stop frequency increases as the service frequency (f) becomes greater.

One effect of this relationship is that optimal stop spacing appears to be smaller for peak operation than for the off-peak. Another point of interest with this model is that although a value of time is used in the theoretical derivation, this cancels out and thus there is no reliance on such a value. This model does not, however, consider operational costs. The lack of these elements of cost in the model has been cited as an advantage when data for their evaluation is scarce (Fernandez 1990). The model also introduces the concept that stop spacing may have a profound effect on the passenger service time. This is because the number of passengers wishing to use a particular stop for access is a function of the stop spacing.

Kikuchi and Vuchic (1982) draw a distinction between a bus stop and the action of a bus in stopping. thus a stop is defined as the physical location of a bus stop, and a stopping is defined as the act of a bus stopping at a particular bus stop. Kikuchi and Vuchic then derive an expression for the average number of stoppings made by a bus along its one-way trips:-

$$E(s) = \mu . (- e^{(-\lambda . h)}) \qquad (3.18)$$

where

$E(s)$	=	expected number of stoppings
μ	=	number of possible stops
λ	=	$2p/\mu$ and is assumed to be constant at all stops along the route
h	=	headway between buses
p	=	number of passenger boardings

As p increases $E(s)$ tends to μ, and as p decreases $E(s)$ approaches $2ph$ (one passenger boards or alights at each stopping). When passenger demand is low, the average total user time on the line is minimized when μ is at its maximum possible value ($+\infty$). This suggests an on-demand stopping regime where a bus is only required to stop if and where a passenger requires access to the system. When passenger volume is large, the optimal number of stops is obtained by differentiating the total user time with respect to μ to obtain:-

$$\mu^{*} = \frac{L}{\sqrt{2.l.v_a.t_l}} \qquad (3.19)$$

where

μ^{*}	=	optimum number of stops
L	=	line length
l	=	average trip length
V_a	=	access velocity
t_d	=	time lost at each stop

which is independent of both p and h. In this case the optimum number of stops is a function of line length, average user travel distance, access speed and the time lost in the action of stopping.

If μ^* is used in place of μ to calculate $E(s)$, the average number of stoppings is obtained for different passenger volumes. Discontinuities occur where on-demand stopping changes to on-call (where stops are defined, but buses stop only when requested by a passenger), and where on-call changes to all-stop operation (where all buses stop at all stops, irrespective of the passenger demand). Heuristics are suggested for the determination of these regimes by evaluating the number of average stoppings per stop:-

$$\gamma = \frac{E(s)}{\mu *} \tag{3.20}$$

where

γ = stopping regime indicator

Kikuchi and Vuchic suggest the following regimes:-

on-demand	$\gamma <= 0.2$
on-call	$0.2 < \gamma <= 0.9$
all-stop	$0.9 < \gamma$

Kikuchi (1985) uses Equation (3.18) to establish a relationship between the number of stops, the headway and the fleet size. The number of stoppings, and the dwell times both increase as passenger volume increases. To maintain a satisfactory user travel time, the number of stops should be reduced. Kikuchi does not consider whether the possibility that a larger number of shorter stops could be advantageous in terms of total user travel time as a result of the reduction in access distance.

CADE-IDEPE (1988) considers three aspects of bus stop location: access time for the passengers using a bus stop, the time spent by passengers already on the bus when it stops at a bus stop, and the operational cost of stopping the bus. The optimum stop spacing is therefore the distance at which the sum of these costs is minimized. The total cost of stopping at bus stops where the stop spacing is given by d is:-

$$C_t = \frac{L_p \, . \, d \, . \, \alpha \, . \, C_p}{4 \, . \, vp} + Z + W$$

$$Z = \frac{1000 \, . \, L \, . \, \alpha \, . \, q}{d} \, . \, (2 \, . \, P_c \, . \, C_d + t_L \, . \, (2 \, . \, P_c \, . \, C_r + T_o \, . \, C_p)) \tag{3.21}$$

$$W = 1000 \, . \, L \, . \beta \, . \, P_L . \, (2 \, . \, P_c \, . \, C_r + T_o \, . \, C_p)$$

where

L	=	length of the corridor
P_L	=	passenger demand per unit distance (per hour)
d	=	distance between bus stops
C_p	=	value of time (passengers)
α	=	weighting for access time
V_p	=	walking speed
q	=	bus flow
λ	=	proportion of bus flow stopping in one bus stop
P_c	=	social cost of fuel
C_d	=	fuel consumed in stopping
T_L	=	lost time per stopping
C_r	=	fuel consumed while engine is idling
T_o	=	bus occupancy in the corridor
β	=	stop time per passenger

The first term represents the cost of access to the bus stops for passengers and the second term, Z, is the cost of stopping at the bus stops when these are d metres apart. The third term, W, represents the cost of the time spent by boarding and alighting passengers in the corridor.

The constant 2 used in the terms Z and W is the result of an assumption that operating costs are twice the social cost of fuel. Minimizing the total cost with respect to d gives:-

$$d = \frac{\sqrt{4000 \, . \, \lambda_q \, . \, (2 \, . \, P_c \, . \, C_d + t_L \, . \, (2 \, . \, P_c \, . \, C_r + T_o \, . \, C_p)) \, . \, v_p}}{P_L \, . \, \alpha \, . \, C_p} \tag{3.22}$$

This model, however, requires inputs such as the social cost of fuel and the fuel consumed during the action of stopping which may be difficult to obtain.

Szász (1990) also approaches this problem as a cost minimization problem. In this formulation, the operating costs and values of time are based on values commonly used in Brazil. Thus the model is both coarser and simpler than that of CADE-IDEPE. Szász considers whether in the system of interest overtaking is allowed at bus stops and therefore includes another element of flexibility. This is the assumption that the proportion of buses stopping at each stop is a factor which affects the stop spacing (Szász 1991). This is based on the relationship between the flow of buses and the passenger density. This model provides a contour specified by the number of passengers on a bus as it arrives at a stop, and which gives the optimum stop spacings according to the relationship between the bus flow and the passenger density.

The Szász model given in Equation (3.23), like that of CADE-IDEPE, evaluates the stop spacing at which the costs incurred by stopping are minimized. Total cost is the sum of the access cost and the cost of stopping a bus. Therefore costs are minimized when these two costs are equal. This gives:-

$$d = \sqrt{\frac{K \cdot 4 \cdot V_p \cdot T_p \cdot (C_b + (P_v \cdot C_v)) \cdot F}{C_p \cdot P_a}} - \frac{F}{P_a} \tag{3.23}$$

where

d = distance between bus stops

T_p = lost time per bus per stop (excluding passenger service time

C_b = operating cost of buses (per hour)

C_v = value of passengers' time in the bus

C_p = value of passengers' access and waiting time

P_a = access passengers per kilometre (passenger density)

P_v = average number of passengers in a bus as it approaches a bus stop

F = bus flow (buses per hour)

V_p = walking velocity (metres per second)

K is the probability of a bus stopping and is given by:-

$$K = \frac{P_s}{P_s + 1} \tag{3.24}$$

where

P_s = the number of boarding and alighting passengers per bus per stop

Where there is no overtaking, K is set equal to 1 since all buses must pass through the stop area, and the final term of Equation (3.23) disappears. In this case, the form of Equation (3.23) is similar to that of the other square root functions used to describe stop spacing, with a slowly decreasing increase in stop spacing as the bus flow increases and the number of boarding passengers decreases in relation to the number of passengers on the vehicle.

Where overtaking is allowed and buses are not obliged to stop at bus stops if there is no access passenger, K takes its calculated value and the final term is used. In this form Equation (3.23) initially has the same square-root form of the CADE-IDEPE model described above, but the final term causes the curve to decrease to zero as the number of buses per access passenger tends to one. The curves representing a contour of optimum stop spacing for operation where overtaking is allowed and that where it is not are shown in Figure 3.11. These curves have been calculated assuming an average loading (P_v) of 40 passengers per bus. It is interesting to note the degree to which the decrease in stop spacing is affected by the possibility of overtaking.

This decrease occurs because the distributions in Figure 3.11 are of the spacing between stoppings. If it is possible for buses to overtake at the stops, a stopped bus does not impede a following bus. When the passenger density is relatively low and the bus flow relatively high, the probability of a bus encountering a passenger is relatively low. Under these conditions a bus may therefore stop wherever a passenger requests without imparting a disadvantage to other buses. In this situation, a stop location is defined as the point at which a bus stops. Subsequent buses may stop infinitely close to this place (where the first bus stops) and therefore the spacing between the stoppings is infinitely small. This provides an extremely good level of accessibility to the bus system. Szász (1990) suggests a value of 0.35 for the bus flow/passenger density ratio above which it may not be necessary to provide formal (on-call) bus stops.

Where overtaking is not allowed, however, the stop spacing becomes progressively greater as the number of buses per access passenger tends to one. This is because the probability of a bus

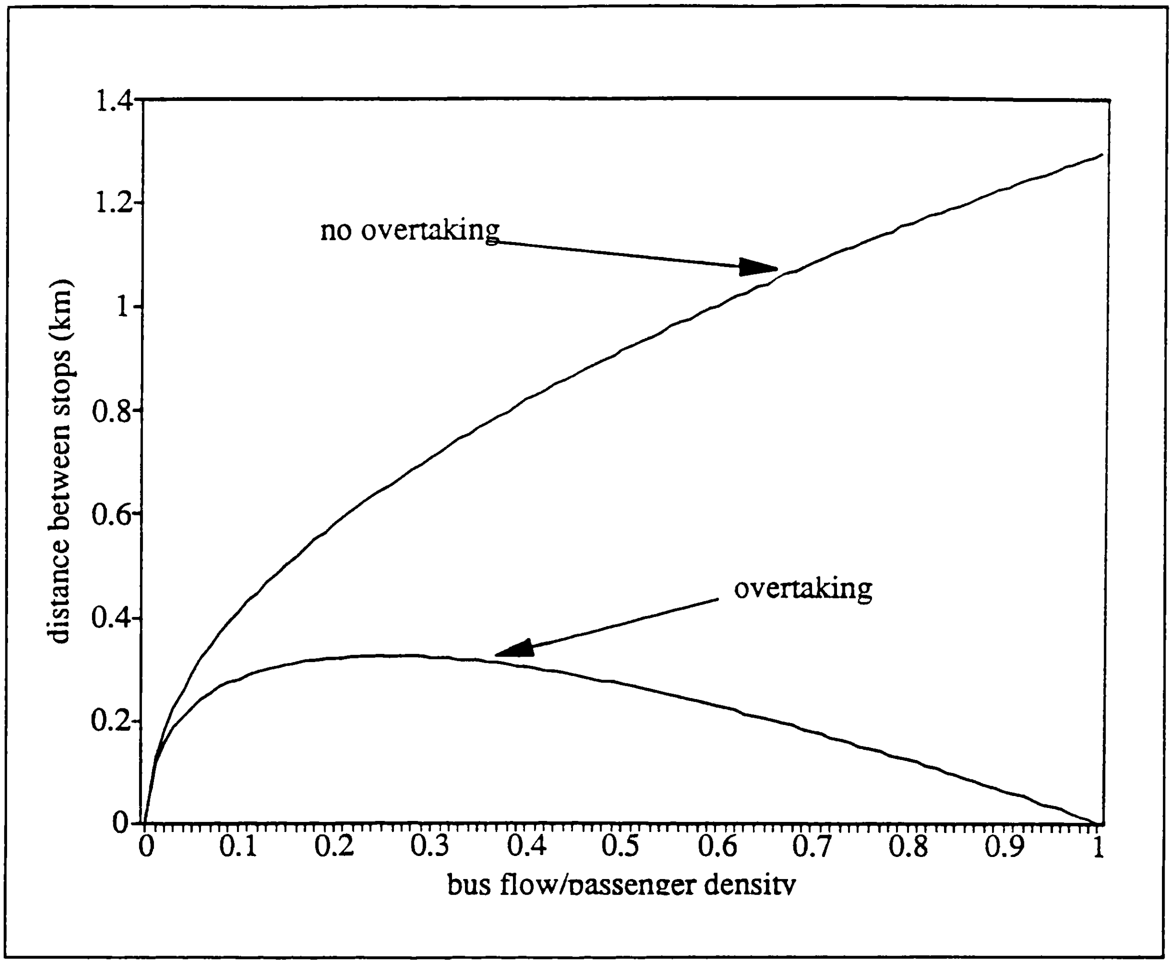


Figure 3.11 Stop spacing as a function of the ratio between passenger density and bus flow for overtaking and no-overtaking stops (average number of passengers in a bus is shown in parentheses).

encountering a passenger is low and therefore requiring a bus to stop at a particular point should be limited as far as possible so that it does not impede other buses. By spacing the stops at a greater distance from each other, the number of stoppings is reduced, and the disadvantage to buses caused by stopping at bus stops for a small number of passengers is reduced.

Where overtaking is allowed, a bus need only stop when a passenger requires it to do so. Thus it is possible that in practice, a particular bus operating within an overtaking regime may stop at intervals similar to those calculated for the non-overtaking regime. The exact locations of these stoppings are, however, specified by the passengers instead of the system.

3.4 BUS OPERATING SYSTEMS FOR HIGH CAPACITY

Normally, bus operation is not expected to achieve a passenger throughput in excess of 9-10000 passengers per hour (Vuchic 1981). Brazilian experience shows, however, that it is possible to design operating systems which enable buses to exceed this capacity by a considerable margin. This research is concerned with the expertise developed during the design of high capacity bus systems. Two types of high capacity system have been designed in Brazil, and these are now discussed.

3.4.1 Convoy systems

The convoy system was originally implemented in Sao Paulo in 1979. COMONOR (Szász *et al* 1985) is an attempt to organise the buses in such a way as to iron out the problems encountered at bus stops. Convoy operation is not, however, a Brazilian invention. Studies made in the United States of America show how convoy operation can enhance the capacity of a bus system (Scheel and Foote 1968, 1969). In these studies, convoys of as many as twelve buses were operated on a test track at speeds up to 70 miles per hour with headways of 4 seconds. These experiments concentrated on the ideal gaps between the buses, both when stationary and when in motion, which would enable the bus flow to be maximized. They did not, however, examine the operation of bus stops.

3.4.1.1 Convoy operation

Bus convoys, as seen in Brazil, operate in a similar fashion to a train. A number of buses travel and stop at bus stops together. As they are not physically coupled together, each vehicle can have a different ultimate destination. For this reason, the bus stops are designed so that buses with the same destination always stop in the same part of the bus stop so that passengers know where to wait. This requires that each convoy should be ordered so that the buses always arrive at their correct berth in the bus stop. These bus stops have been discussed in Section 3.3.4.2.2.

In order to reduce the number of berths necessary to serve all the possible destinations, bus lines are grouped according to their destination area and the level of demand at the critical stop. Generally the group with the highest demand is positioned at the front of the convoy, since this helps to keep the convoy together along the corridor. In Sao Paulo, it was found that the best combination was to have three groups, with two buses from each group represented in a convoy. The groups are colour coded and denoted by an identity letter such as 'A', 'B' or 'C'.

It is not necessary to have three groups, and some consideration should be given to this matter. A higher number of groups has the advantage that fewer passengers will be waiting for each group, and embarkation times will be shorter. However, three disadvantages of increasing the number of groups tend to become evident. First, the number of bus lines in each group reduces, resulting in a lower arrival frequency of buses within each group. This tends to increase the delay caused by convoy assembly. Second, as each group requires at least one allocated berth at each bus stop the length of the bus stops can become extended as the number of groups increases. Third, an increase in the number of groups means a finer delineation between the destination areas served by each group. This means that the probability of a passenger being able to choose between two groups for any particular destination increases, and this may cause confusion at the bus stop.

If for some reason the convoy becomes disturbed (for example if a bus in group B or C is unable to leave the stop when the group A buses have left the stop) the convoy continues and reforms at a subsequent stop. It is essential for the efficient working of this system that the convoy order is maintained even if the ideal spacing suggested by Scheel and Foote (1969) is not possible. If the convoy order is broken, successive buses may be unable to reach the allocated berth at a subsequent bus stop, and the congested conditions observed before implementation are likely to return. Indeed this situation occurred in Sao Paulo, and contributed to the abandonment of the COMONOR system and its replacement by the express bus system described below.

3.4.1.2 Convoy assembly

Clearly, the critical point of convoy operation is the assembly of the convoys themselves. As mentioned above, this is a function of frequency of arrivals of all groups, and of all bus lines in each group. If convoys can only be released when they are complete, the rule is simple, but the delays caused by waiting (particularly for the second bus of the last group) can be considerable even when the aggregate frequency is high. To prevent these delays, and also to reduce the space required to store queuing buses, it must be possible to release incomplete convoys.

There are several possible operational strategies that could be used to decide when to release an incomplete convoy. The most simple is to release convoys on a regular basis, whatever the composition at the time. This can be controlled by a traffic signal dedicated to the purpose (Szász *et al* 1985). In this instance, 8 seconds is allowed for each group, on the assumption that this is sufficient for two buses to pass the stop line. Between the green time for each group is 2 seconds when all signals are red, and the cycle is made up with 32 seconds when all groups are retained by red signals. In this way the cycle repeats every minute. Thus the convoy frequency is 60 per hour and the maximum bus frequency is 360 buses per hour.

The convoy system requires an area for convoy assembly. The space is divided into two parts which may or may not be contiguous. The first is the assembly area. This consists of sufficient space for the maximum number of buses of each group in the convoy. These buses will form the next convoy to depart. The second space, the storage area, stores the third and successive buses of each group that arrive while the assembly area allocated to that group is occupied. The size of this area depends to some extent at least on the relationship between the release algorithm for the convoys and the arrival frequency of the buses at the assembly area. This is because the storage space must be sufficient to store all the extra buses without obstructing the remaining road space. This is much easier to achieve if the frequencies of each group are nearly identical, hence the conditions placed on this in the allocation of bus lines to groups in the convoy (described in Section 3.3.4.2.2).

In Porto Alegre, this space occupies three lanes (one for each group), each of which is long enough for two buses, which is quite a considerable space to be taken out of the existing road system. Such a system is illustrated in Figure 3.12.

It is important that the delays caused by convoy assembly are more than offset by the time benefits gained through the convoy operation itself. It is difficult to obtain data which reliably describe this delay, but the benefits gained in the 9 de Julho corridor in Sao Paulo were considerable. Over the 4 kilometres of the system which comprised 6 bus stops and 5 signalised intersections, the commercial speed increased from 10 to 19 km/h (EBTU 1982).

3.4.1.3 Operational results

The requirement noted above that the convoy order must be maintained was not enforced in Sao Paulo, since the bus lane was situated at the kerbside and there was no physical segregation between the bus lane and the lanes for non-priority vehicles. Drivers could therefore overtake at

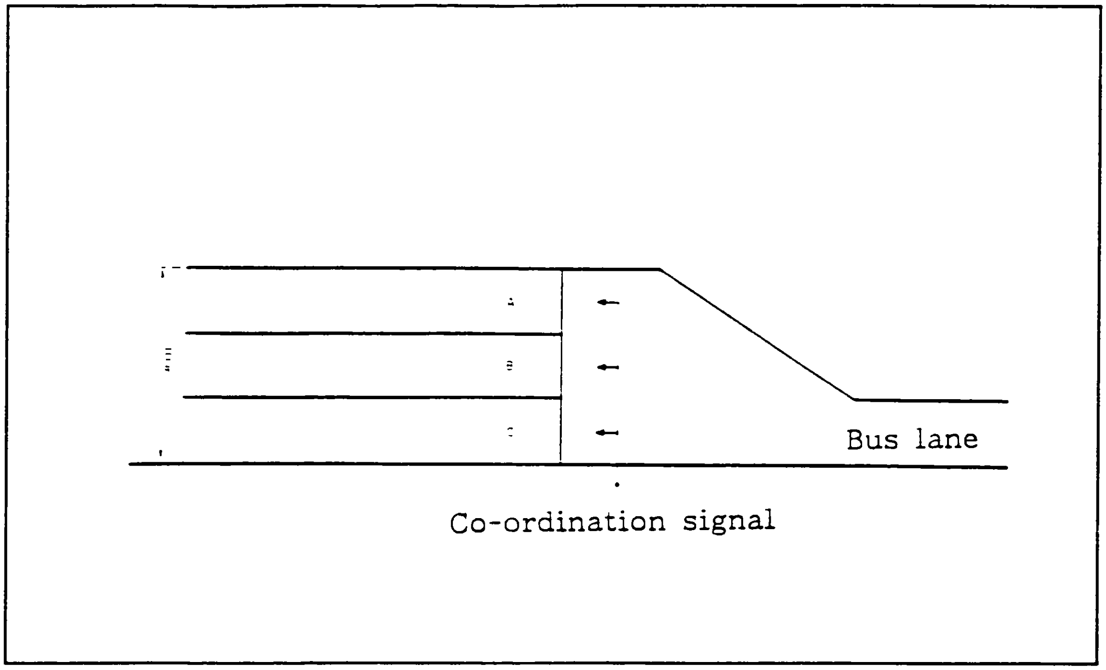


Figure 3.12 Convoy assembly area. Source: Szász *et al* (1985).

bus stops if they felt that the bus in front was too slow, and in this way the convoy order was broken. Another crucial cause of the breakdown in the ordering of the convoys was that the police (who were required to oversee the convoy assembly) failed to understand that the capacity benefits from the convoy operation were achieved as a result of the (almost) simultaneous boarding of all buses at each stop. As a result, there was a tendency to allow buses to leave the convoy assembly area independently of each other and constrained only by the occupancy of the bus lane immediately ahead of the assembly area. The system deteriorated until it was removed.

The COMONOR system was also implemented on two corridors (Farrapos and Assis Brasil) in Porto Alegre in the south of Brazil. Here the convoys operate along segregated lanes in the median of the corridors. In this case, the segregation ensures that the drivers are unable to overtake each other, with the result that the convoy order is maintained. A comparison (SPM 1981) of the situation in 1978 before implementation with that in 1981 after implementation shows that the results are mixed. In Avenida Farrapos, the average speed increased, the level of the increase depending on the time of day, and ranging between about 6 per cent and over 30 per cent. A similar comparison for Assis Brasil is not so encouraging, with speeds observed in the two studies showing at worst a decrement of about 15 per cent and at best no change. Although the study attributes some of the increase in journey time to an increase in the length of the route, it does not explain why the average speed should be affected in this way. A further indication of the effect of this implementation is that there is a slight reduction in the ranges of observed average speeds.

3.4.2 Express bus systems

In Sao Paulo in 1984, Projeto Trolebus was begun to design a new trolleybus system for the Santo Amaro/9 de Julho corridor, which extends for some 12 kilometres. The design team chose to try a new method of extending the bus capacity, this time locating the bus lane in the median. This project (CET 1984) involved extensive research before implementation, and to date no published review of the success of the system (which was implemented in 1987), has been forthcoming.

3.4.2.1 Express bus operation

The principle behind the design of this system is that the journey profiles of the passengers using the corridor should be examined to discover the actual passenger loadings of buses along the corridor. In this case, it was discovered that the major demand along the corridor was in the first section (approximately one third of the total length) near to Santo Amaro, the suburban end of the corridor. If the bus capacity were increased to satisfy this level of demand there would be an excess of capacity elsewhere in the corridor with an overprovision of buses. The solution devised was to match the supply to the demand and to create an operational system that permitted the variations in demand to be matched with the lowest number of buses.

A variation in demand of this order is not amenable to convoy operation, which would tend to provide too many buses in the parts of the corridor with lower demand. Also, the design team seem to have felt that convoy operation was no longer the best solution to this type of problem, and a different solution was sought. With the highest demand in the suburban part of the corridor, it is reasonable to suppose that the problem for those passengers wishing to travel further along the corridor is that they would be delayed by the high level of activity at and around the bus stops in the first part of the corridor. To avoid this, and to free the buses that would also be caught in this congestion, a facility is required to permit overtaking at the bus stops. Also, the lessons of the earlier design being learnt, it was felt that the design should incorporate physical segregation of the bus lane. A special kerbstone was designed to provide the physical segregation. This is significant because the cross-sectional shape is eccentric, making it easier for buses to mount than for cars. Buses may pass other buses which have broken down, thus reducing the impact on the system of such events. Similarly, a car could cross the kerb if this were necessary (for example to avoid an accident), although the design of the kerbstone is such that this manoeuvre would not be particularly comfortable. The physical segregation in effect provides a reserved track for the buses, and to avoid problems with loading and parking for the commercial activities along the corridor, it was decided that the system should be located in the median.

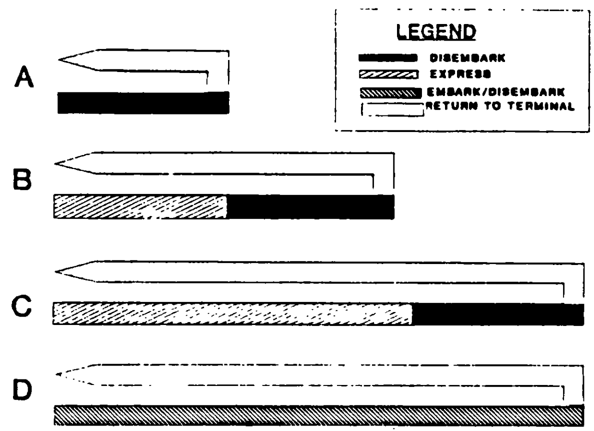
Having decided on the location of the lane and the nature of the bus stops, attention was turned towards the actual operation of the bus system. The corridor was divided into three parts, according to the demand profile. As mentioned above, the first third of the corridor has the highest demand, and to accommodate this, the first division of the corridor was placed at the end of this stretch. A schematic diagram of the operating system is shown in Figure 3.13. To serve this demand, some buses (Group A in Figure 3.13) call at every stop in this section, but only to allow passengers to alight. When they arrive at the end of the section, they return to the terminal at Santo Amaro to make another journey. Some buses call at bus stops on the return, while others return empty. These buses can therefore provide a high level of service over this busy section.

In order to serve the next section, some buses (Group B) run non-stop to the end of the first section, and then call at all stops in the second section to allow alighting. As with the first group, they then return to the terminal at Santo Amaro for another journey.

A third group (Group C) runs express to the end of the second section, stops at all remaining stops to allow passengers to alight, and returns from the end of the corridor.

The last group of buses (Group D) stops at all stops along the entire corridor, for both boarding and alighting. Clearly the usefulness of the overtaking facility at the bus stops can be seen as it allows the buses in groups B and C to pass the buses which are stopped at the bus stops.

These 'trunk' buses are fed in two ways. First, the terminal at Santo Amaro allows passengers arriving from the outer suburbs to transfer from their local buses to the trunk buses for the remainder of the journey. Second, although normally only trunk buses operating from the terminal at Santo Amaro are allowed inside the bus lane, there are special stops along the corridor where local buses can set down or pick up passengers for transfer to trunk buses. There is no payment at the point of transfer: the passengers pay a special 'integraçao' ('integration') fare



NOTE:

Groups A, B and C only allow alighting;
Group D is available for boarding and alighting.

Figure 3.13 Schematic diagram of bus operation along the Santo Amaro/9 de Julho corridor in Sao Paulo, Brazil.

which allows the two journeys to be made for a lower price than two normal single journeys.

The number of buses required to operate each type of service is determined by calculations to determine the length of time taken to perform the round trip. An obvious part of this is the time spent at bus stops, which since in the cases of the first three groups only permits alighting does not include an analysis of the marginal boarding time. Where this has been used, in consideration of the fourth group, a constant value has been assumed and this is determined to be 1.67 seconds. This is of a similar value for the convoy system above, and there is no quantitatively supported justification for this value.

3.4.2.2 Transfer terminals

Using the Santo Amaro terminal as a transfer point provides the benefit that the buses operating in the outer suburbs are no longer operating along the whole corridor. This has three main advantages. First, they are not involved in the congestion associated with the central area of the city, and thus can operate a more reliable service in the outer suburbs. Second, their journey lengths and times are shorter, allowing a higher frequency in the outer suburbs with the same (or perhaps a smaller) number of buses. Third, the trunk buses operating along the corridor have in effect been consolidated, which tends to reduce the number of buses required to satisfy the demand. The disbenefit is that passengers are required to change buses at the terminal. In the short term, the overall time savings on the journey from the outer suburbs to the city centre are such that this does not appear to be too damaging. In the longer term, however, this may become an issue that needs to be resolved: in Porto Alegre, such transfer terminals have recently been removed because of passenger dissatisfaction with the transfer process. In Sao Paulo, plans have been made for a similar (although simpler) system from Rio Branco in the north west of the city to Largo do Paissandu near the centre (CMTC 1985).

The special transfer stops along the corridor permit better access to those buses operating from other outer suburbs which can join the corridor more efficiently at these points than travelling to the Santo Amaro terminal. Interestingly, this facet of the operation would be very difficult to manage with a convoy system. While it would be feasible for buses to leave the convoy along the corridor, it is very difficult to arrange for buses to join a convoy along a corridor without imposing extra delay on the whole convoy.

The design of the infrastructure required for the express system is that of bus stops in the median with a lane for overtaking. The stops are linked by a segregated median bus lane. These are described in more detail in Section 3.3 above.

The express bus system uses a transfer terminal to consolidate the passenger demand from various outer suburbs to reduce the number of trunk buses actually required to use the corridor into the city centre. This terminal (Terminal Padre José Maria) has to be able to process the transfer of some 7300 passengers per hour in the morning peak from the outer suburban buses to the trunk buses and another 4000 passengers whose journeys terminate at the terminal (CET 1984). The total area for this terminal is approximately 15000 m^2.

There are platforms for the 21 outer suburban bus lines that terminate at this terminal at a frequency of 275 buses per hour. 4 trolleybus trunk lines commence at the terminal to operate along the corridor in the manner described at a frequency of 99 buses per hour. There is also an area for excess buses to park and for cleaning. There is another suburban terminal (Terminal Borba Gato) which accommodates 81 suburban (12 lines) and 39 (2 lines) trunk buses per hour. The third terminal in the system (Terminal Bandeira) is located at the city centre end of the corridor and is much simpler in concept than the others, notably because fewer bus lines are involved in its operation. In terms of operation, these terminals operate quite conventionally. The only facilities required which could be considered additional to the basic terminal designs discussed, for example by White (1988), are the fare collection facilities used by passengers joining the system at the terminal. As mentioned above, passengers arriving on other buses do not need to pay and are therefore able to change buses without additional payment.

3.4.2.3 Operational results

The express system in Sao Paulo was implemented in 1987, and to date no detailed data are available which can shed light on the capacity implications of the bus stop operations, or of the efficiency of the groupings and trunk-haul operations. A study undertaken on behalf of the Transport and Road Research Laboratory (TRRL 1990) indicates that the commercial speeds obtained in the corridor are of the order of 22 kilometres per hour, and that the observed bus flow is approximately 320 buses per hour, giving a nominal line-haul capacity of 22400 passengers per hour. It should, however, be pointed out that this measure fails to measure the true capacity of the bus system which is also a factor of the accessibility of the system. For this reason, such capacity measures can be slightly misleading and understate the real benefits of using a bus system (rather than for example a metro system) for such demand.

Results of this nature, although better than nothing, are not sufficiently detailed to give insights about the bus stop operation, or what the effective capacity of the bus stops actually is. This is certainly an area which should be investigated further and more deeply.

3.5 CONCLUSIONS

This chapter has been concerned with the design characteristics of high capacity bus systems. Some of these characteristics have constraints or difficulties which determine the way in which they may be represented in a computer model. The constraints and difficulties have been discussed in this chapter, and it is helpful to bring them together in order to draw some conclusions about the way in which a computer model must be designed in order to model the design of a high capacity bus system.

It appears that high capacity bus systems benefit from the use of a bus lane that is physically segregated from the adjacent traffic lanes. This protects the general traffic from the high volume of buses, and restricts the amount of parking and unloading activities which may interfere with its operation. Such physical segregation should be breachable: the eccentric kerbs mentioned in Section 3.4.2.1 allow buses to escape from the bus lane if there is a broken-down bus obstructing the lane. Physical segregation, however, restricts buses to the bus lane under normal circumstances and therefore it is extremely important that the bus lane has sufficient capacity to cope with the flows of buses which may be expected.

In terms of physical infrastructure the UK designs are not comprehensive. The requirements of the Department of Transport are geared to the conversion of a normal traffic lane to special use - often for only part of a day - and no particular infrastructure is used to enhance further the efficiency of bus operation. Generally, bus stops are not moved or redesigned with a view to making their operation more efficient, the only mention of bus stops is that they should not be positioned in the setback as this would reduce junction capacity when occupied by a bus.

Both in Brazil and Chile, the principal restraint with respect to the capacity of a bus lane is seen to be the capacity of bus stops. For the design of a high capacity bus system, therefore, it is essential to provide bus stops which have sufficient capacity for the number of passengers and buses expected to use the system. Ways in which the capacity of a bus stop may be estimated have been discussed and this analysis has indicated that the design of a bus stop may have a considerable impact on the stop capacity - and therefore on the capacity of the bus system in the corridor as a whole. Some of these bus stop designs require particular constraints on the operation of the buses. Examples of these constraints include operating in ordered convoys, overtaking at bus stops, division of bus lines into groups, and the allocation of specific berths to these groups. The capacity of the bus stops in a corridor is a major determining factor in the commercial speed obtained in the corridor. A major conclusion from this review is that bus stop capacity is the main driver of the decisions concerning the type of bus system to adopt.

The questions also arise about the nature of the relationship between inputs to the design process (such as bus flows and passenger flows) and the outputs (for example, the number of berths at a bus stop). Such inputs are distinguished by the fact that they may take any positive value - if averages are used, these may even be real numbers. In many cases, however, the outputs are discrete: it is not possible to have one and a half berths at a bus stop. A major requirement for this research is therefore to explore how the design engineers decide about the magnitude of a discrete output when the inputs are continuous.

In attempting to make their analysis more rigorous, however, one of the main problems for bus system designers in developing countries is that there is little or no history of data collection. Methods that require extensive data analysis are in danger of producing spurious results if the reliability of the data being used is not taken into account. In order to facilitate the use of good quality design with respect to high capacity bus systems, it is therefore of great practical importance to establish a model which can identify, and proceed with, data which are of poor quality.

In developing countries, the number of buses per car is much higher, and the number of buses per inhabitant is much lower. The combination of these two factors points to the necessity of designing systems to facilitate the efficient operation of large numbers of buses.

In the United Kingdom such high bus flows are not common, although with the introduction of smaller buses, problems may occur if the bus flows increase. The work discussed in this Chapter, however, with its emphasis on the importance of bus design and the sensitivity of the operation of the bus system to the capacity of its bus stops, may have implications for the design of critical bus stops in the United Kingdom. It is therefore of interest to explore whether these ideas may be transferred to this operating situation.

The first decision in the design process is, however, whether or not to proceed with the design. This should not be overlooked as there is a definite cost attached even to pre-feasibility studies. Brazilian engineers use a simple model to check the potential viability of a feasibility study, using commercial speeds, which may be characterized by the fact that the relevant data are relatively cheap to obtain. One conclusion to be drawn from this process concerns these data. The data are not exact, and this feature characterizes much of the data concerned with bus system design. It is therefore necessary to consider how the decision process may be modelled to take account of the use of imprecise data.

CHAPTER 4
ARTIFICIAL INTELLIGENCE TECHNIQUES

4.1 INTRODUCTION

Artificial Intelligence (AI) is an area within the field of computer science which studies the representation of human thought by computers (Boden 1987). Computers are machines which are able to perform numerical calculations using a binary representation of numbers. This binary representation means that at the most fundamental level, a computer calculates only with the knowledge of two states: ON and OFF. With the use of sophisticated computer languages, complex calculations can be represented in terms of this binary form. AI, however, attempts to include in a computer program those elements of a process that may be characterized as 'human'. Indeed, Boden (1987) states that AI is the study of methods to enable computers to perform the type of functions normally expected of humans. The breadth of this description is important because there are many things that can be done by humans that are not immediately possible on a computer. Examples of AI work may be found in research into computer vision, robotics, theorem-proving, problem solving, machine learning, and systems in which the reasoning is dependent on expert knowledge (Boden 1987).

AI work in each of these areas is concerned with the qualitative assessment of data, where this is extrinsic to the quantitative evaluation of particular values within the data set concerned. For example, a data set which contains the recorded times at which buses pass a certain point may be analyzed quantitatively to obtain the headway distribution. Before this can be used in the design of a high capacity bus system, however, it is necessary to interpret the information in terms of its effect on the various design possibilities. The design engineer performs this task routinely, using knowledge developed over years of experience. This type of activity is an intelligent approach to the data, where it is the meaning contained within the data set rather than the specific values which is of importance. The interpretation of data is therefore one area in which a design engineer uses knowledge to provide qualitative assessment and this part of the decision process could be an area in which AI work could make a contribution.

This research is concerned with the representation of a design process undertaken by experienced engineers. As noted in Chapter 3 above and discussed further in Chapter 5 below, a large part of the design engineer's time and effort is taken up with the interpretation of data. In order to reproduce the design process in the form of computer software it is therefore necessary to represent this interpretation process in a way that is appropriate. The area of AI work which appears to be most relevant to this work is that concerned with the use of expert knowledge, known generally as Intelligent Knowledge-based Systems (IKBS).

In order to maintain an element of intelligence in a knowledge-based system, it is necessary to model, not only the factual knowledge held by the engineer, but also the way in which this knowledge is used. In particular it is important to represent the way in which the engineer works with partial information, imprecise data and uncertainty. Gaines and Boose (1990) emphasize that experts are not simply stores of knowledge and skills. Rather they are 'managers of the knowledge acquisition process itself, arbitrating the inductive process whereby knowledge is created and updated through experience'. This suggests that expert knowledge consists not only of facts, but also elements of process: how to use these facts in order to infer a solution, even where such a solution has previously never been adopted. Gaines and Boose also note that experts are often occupied in problem-solving where they have at least some degree of uncertainty about their data and knowledge. This source also notes that a major objective of research into the technology of knowledge-based systems has been involved in the capability of these systems to 'emulate human skill in coping with uncertainty'.

This research explores the possibility of using artificial intelligence techniques to assist in the design of high capacity bus systems. This chapter, together with the two chapters which follow, considers the use of these techniques to represent the knowledge used by the design engineers concerned. This chapter is a brief review of the main techniques that are currently available. Chapter 5 presents a model which is intended to surmount some of the difficulties of representation noted in this chapter. Chapter 6 describes how this model is used to obtain the expertise necessary for inclusion in the decision model.

This chapter has four further sections. The first section is a discussion of various methods for the representation of this expertise, the second considers inference and the third explores techniques used for the representation of uncertainty and imprecision. The relevance of these techniques to the design of high capacity bus systems is discussed in the fourth section.

4.2 INTELLIGENT KNOWLEDGE-BASED SYSTEMS

4.2.1 Introduction

While all computer software has at least some element of knowledge implicit in its construction, knowledge-based systems are different in that this is made explicit. Often it is entirely separate from the main procedural part of the software and is located in a self-contained 'knowledge base'. To obtain a solution, an inference process attempts to match inputs with the declarations in the knowledge base. Various methods are used to perform this pattern-matching task, and three of these are described below.

This section is divided into three parts. First, the methods used for the representation of knowledge are described. Second, the issues underlying the inference process are discussed and third, the various treatments of uncertainty and precision are considered.

4.2.2 Knowledge representation

A knowledge-based system is a system by which expert knowledge is encapsulated within a computer program. The way in which this knowledge may be acquired is considered in Chapter 6. This section considers the methods used in order to represent the knowledge of experts in an appropriate form.

Knowledge representation has been defined as 'a set of syntactic and semantic conventions that make it possible to describe things' (Winston 1984). Thus a knowledge representation system uses a formal syntax to represent knowledge. The syntax must be sufficiently robust that the knowledge-based system can have reliable access to the relevant knowledge, but should not be so rigid that the richness of the original knowledge is lost.

It has been suggested (Jackson 1986) that a knowledge representation system should be judged on three main grounds: logical adequacy, heuristic power and notational convenience. Logical adequacy ensures that the knowledge representation system truly represents the appropriate logic, and is logically valid and consistent. Heuristic power refers to the process of inference and is a measure of the ability of the representation system to solve problems. Notational convenience is deemed important because of the complexity of most knowledge-based systems.

It should be noted, however, that these measures confine the accuracy of the representation to a logical dimension. The question of whether the knowledge is fully represented (however subtle the distinctions between various elements of the knowledge) is not approached in this formalism except with respect to its logical content. The semantics underlying the expression of the knowledge by the expert - the meaning of the expert's statements - is therefore dependent upon its logical consistency, and the semantic considerations are reduced to logical formalisms. While this may be consistent with the definition of Winston and the assessments of Jackson (both noted above), the quality of a knowledge representation system rests on its ability to represent the knowledge as expressed by an engineer in all its richness. Thus the logical formalism should serve the semantic consideration rather than the reverse.

A major part of an expert's activity is concerned with the definition of the problem to be solved, and the formulation of the problem is a critical part of the decision process. Brachman *et al* (1983), however, recommend that a problem can be reformulated so that it becomes converted to a form which is appropriate for processing by one particular form of representation system (expert rules in this case). Clearly such reformulation can only be appropriate where it does not alter the nature of the problem. Care should therefore be taken when engaged in this activity so that the problem which is solved is the same as the original problem. The knowledge representation system should be determined by the way in which the expertise is used by the human experts, and this implicitly incorporates their model of the problem. Within the well-defined problem domains normally encountered by knowledge-based systems, this should be appropriate for the particular problem at hand without the need for substantial reformulation.

Three types of knowledge representation system are frequently used in knowledge-based systems: rules, frames and blackboards.

4.2.2.1 Rule-based systems

4.2.2.1.1 Rules

Rule-based representation stems from a common syntax for the expression of knowledge, usually called production rules:-

'IF condition C holds, THEN consequent Q follows'

This can be represented in the form of a logical sentence in propositional logic:-

$$C \supset Q \tag{4.1}$$

thus introducing an element of logical syntax to the representation of knowledge. In this example, if the truth value of C is 1 (in other words the condition is true), a truth value of 1 is implied for the consequent Q. For representation in a form suitable for computer processing, propositional sentences are converted to predicates. This process is described in detail by Thayse (1986). Thus Rule (4.1) then becomes:-

'consequent Q is TRUE IF condition C is TRUE'

which in predicate form is:-

Q :- C. **(4.2)**

Rule-based systems are characterised by the use of a set of production rules in which a consequent is established if the relevant conditionals are found to be true. For the purposes of this thesis, the more familiar 'IF ... THEN' propositional form is used in preference to the (semantically more correct) predicate representation '... is TRUE if ... is TRUE'. The use of knowledge represented in this form depends on the evaluation of the truth value of each rule. If a rule can be proved to be true, it is said to succeed and has a truth value of 1. Otherwise it is said to fail and takes a truth value of 0. This binary nature of rule-based representation is an important factor because it underlies the power of the representation system. The expertise, if it can be represented in this way, is then a matter of the evaluation of the truth values of a series of production rules. This evaluation can be undertaken as a series of logical tests of these rules. The testing of rules is called 'firing'.

Most early knowledge-based systems were constructed using this type of representation. MYCIN, a knowledge-based system for the analysis of blood disorders and appropriate drug therapies (Shortliffe 1976) is one example. MYCIN contains rules of the form:-

IF the stain of the organism is 'gramneg' **(4.3)**
AND the morphology of the organism is 'rod'
AND the aerobicity of the organism is 'aerobic'

THEN there is strongly suggestive evidence (0.8) that the class of the organism is 'enterobacteriaceae'

Rule 4.3 describes an expert's conclusion about the class of an organism given the confirmation of the stain morphology and aerobicity parameters. The derivation and use of the number which appears in the consequent is explained in section 4.2.3 below. The principle behind the operation of this knowledge-based system is that the user should identify the truth values of these parameters. Therefore the system asks the user whether the organism is gram negative. In this case the rule is crisp: if, for example, the stain of the organism is not gramneg, then the rule fails. If the answer is affirmative, then the next conditional is tested. Confirmation of all the conditionals leads to the consequent being established, and the success of the rule. In fact, the consequent necessarily follows from the confirmation of the truth of all the conditionals. Conversely, if any of the conditionals fails, then the rule also necessarily fails.

On the other hand, the truth of a rule in particular circumstances may not be so easily determined by such a simple test. In such circumstances the rule may need to be qualified in order to establish more precisely the occasions when it is true. In any case, the rule-based system still has the binary true-false characteristic: the use of qualifications in the rule simply defines the rule more precisely so that its meaning can be expressed in these terms.

The qualification of a production rule is obtained by the use of one or more clauses. A clause is a part of a production rule with its own conditionals. In the same way as a production rule, each clause can have more than one conditional which can be combined using the logical connectives 'AND' and 'OR'. In the first case, each conditional must be established to be true before the consequent can be inferred. In the latter case, only one conditional needs to be true in order for the rule to succeed (although, in fact, more than one could be true). Rule 4.3 is an example of a production rule with only one clause. Further clauses might be added to allow alternative conditionals to lead to the same consequent, or to allow different consequents to be attained from different sets of conditionals.

Clauses are effectively linked by an 'OR' operator because within a particular rule a clause is only tested if a previous clause has failed (or if it is the first clause in the rule). Thus 'OR' is used to represent 'ELSE' in this case.

Rule (4.3) is a very simple example. A more complex rule contains 'OR' operators. These can be used in two ways. The first is where a consequent can be inferred from different sets of conditionals:-

IF there is not a lot of parking along the kerb **(4.4)**

OR

IF there is a lot of parking

AND enforcement is very strict

THEN the bus lane can be located at the kerbside.

In this case, the production rule contains two clauses. The first clause contains the first and last lines of the rule, and the second clause comprises the last three lines. In this simple case, the consequent (the last line) is common to both clauses: the rule can succeed under two different conditions. Here, if the first conditional succeeds, the consequent follows. If it fails, however, the second clause is tested. In this case the second clause contains two conditionals linked through the 'AND' operator. The consequent also follows if the second clause succeeds (which is in turn dependent on the success of both of its conditionals).

The second way in which the 'OR' operator is used is where two or more different consequents can be inferred, depending on the truth of different sets of conditionals. To illustrate this, Rule (4.4) is expanded to allow for the failure of both conditionals by adding a third clause:-

IF there is not a lot of parking along the kerb **(4.5)**

OR IF there is a lot of parking

AND enforcement is very strict

THEN the bus lane can be located at the kerbside.

OR

IF there is a lot of parking along the kerb

AND enforcement is weak

THEN the bus lane should be located in the median.

In this case, the third clause contains a different consequent: the rule can therefore succeed with different results. An important point to note is that because the truth value of each conditional, clause and rule is established in order, it is extremely important that the sequence of tests is correct.

If the whole rule fails and there is no alternative clause, it is necessary to return to the previous successful goal and try another branch of the decision tree from that point. This process is called backtracking (which is also the method by which explanations are derived in rule-based systems).

4.2.2.1.2 Combinatorial explosion

The effect of these operators is to increase the number of clauses in the rules in the knowledge base. This has some unfortunate effects. The first effect is that the rule base becomes very large and therefore more difficult to manage. The second problem is a result of the binary nature of the representation system. For every rule, two possibilities exist: success or failure. Thus the space in which the solution exists (the search space) is a function of the number of conditionals. In fact, this is an exponential function to base 2. Thus a rule base with three conditionals has eight possible solutions and one with four has sixteen. The addition of one more conditional increases the search space to thirty two possibilities, and so on. This effect on the size of the search space is known as 'combinatorial explosion', and has important implications for the availability and use of computer memory.

Rule (4.5) has two conditionals which need to be established: the level of parking and the effectiveness of the enforcement. The search space therefore contains a total of four possible solutions:-

(PARK = *not a lot*) AND (ENFORCEMENT = *not strict*) IMPLIES <KERBSIDE>
(PARK = *not a lot*) AND (ENFORCEMENT = *strict*) IMPLIES <KERBSIDE>
(PARK = *a lot*) AND (ENFORCEMENT = *strict*) IMPLIES <KERBSIDE>
(PARK = *a lot*) AND (ENFORCEMENT = *not strict*) IMPLIES <MEDIAN>

The advantage of the logical dimension can be seen in this simple example. If there is not a lot of parking, the bus lane will always be at the kerbside. Therefore it is possible to establish the truth of the rule without testing the level of enforcement if the level of parking has been established to be '*not a lot*'. In this case, the search space can be reduced to three possible solutions, as is the case with Rule (4.5).

The problem of combinatorial explosion can be tackled in two ways. The first is through the careful definition and use of conditionals as in Rule (4.5) above. Notice in this case that the levels of parking and enforcement are only defined in broad descriptive terms. For these conditionals to be included in a rule base, it would first be necessary to have some definition of '*a lot*' and '*strict*' as well as any other descriptions of these parameters. Such definitions would be the subject of other rules (which would therefore be added to the knowledge base).

The second way to tackle combinatorial explosion is to reduce the size of the search space by reducing the scope of the problem domain. In this way the problem is reduced and therefore the number of rules and conditionals required in order to define it is also reduced. There is clearly a trade-off here: if the problem is reduced too much, it becomes trivial.

The 'OR' operator is the rule-based system's approach to flexibility. This is because the conditionals can be used to differentiate between fine details (for example, between an engineer's interpretation of a value of 3.4 metres for a bus lane width and 3.5). Such sensitivity could be achieved with the use of rules with suitably defined conditionals. This, however, has the combinatorial consequences which arise from the use of large rule bases, and therefore such sensitivity is unusual in a rule-based system.

This approach is adopted by Tomii *et al* (1989). The solution presented by Tomii *et al* is to provide a basic system which does not contain the site-specific expertise. This is enhanced by local domain experts who add their own knowledge concerning local sites. The result is a site-specific knowledge-based system which does not deal with the fundamental principles of the problem at hand. This is therefore a way of reducing the problem domain to one of proportions which are manageable within a rule-based context. It does, however, rely on the availability of domain experts with local knowledge of the sites in question in order to provide the missing expertise. This system does not work if such expertise is not available. It should be noted, however, that as noted by Cagan and Agogino (1987) such decisions are better modelled from first principles.

4.2.2.1.3 Problems with rule-based representation

Dreyfus and Dreyfus (1986) argue, however, that rule-based systems are not appropriate for expressing expert knowledge. It is suggested that the difficulties found in the elicitation of the expert's rules (also noted by Witten and Macdonald 1990) are based on the fact that the experts are not using any rules at all. Rather they are using a set of thousands of special cases. Thus an expert, when considering a problem, rather than perceiving it in terms of conditionals and consequents views the potential solutions in terms of a list of previous examples, one (or a combination) of which will provide the solution.

Witten and Macdonald also observes that experts find it easier to demonstrate their expertise by using specific situations than to try and formulate rules to define their knowledge.

Rule-based systems are useful when the problem can be defined by the statement of a succession of rules. For best results, these rules should be interrelated in a hierarchical form such as a tree structure. The problem domain should be sufficiently small to prevent a combinatorial explosion during the search. The rules should not require fine distinctions to be made in the conditional test, as this tends to increase the size of the rule base.

The problem which arises from fine distinctions is that for each distinction, it is necessary to have a separate set of conditionals. One way of reducing this problem is to devise a representation system which can define objects in a wider sense. For example, when considering a simple piece of furniture, such as a table, it is not always necessary to consider all of its fine details. In certain circumstances it may be sufficient to know only that an object has a horizontal surface and that its size is irrelevant. A representation system which treats objects in this way is known as a 'frame' system.

4.2.2.2 Frame-based systems

To avoid the inflexibility of rule-based systems, some researchers approach this problem with frame-based systems (for example Hendrickson *et al* (1987)). These are more sophisticated than rule-based systems because they define aspects of expertise as classes of objects. Each class has various attributes, and any object, once it has been identified as a member of a particular class, inherits the attributes of the class. Generally, the higher levels of the frame system represent things that are typically true about the object, and so are fixed. Lower levels (slots) are filled with actual data which can, if necessary, be constrained to comply with defaults. This principle can be used to describe objects which are basically the same, but which differ in detail.

Thus a frame to describe a particular person might be constructed as follows (it is assumed that a THING is an object with attributes):-

[**PERSON** is a kind of **THING** with **(4.6)**
AGE
HEIGHT
WEIGHT]

[**PROPERTY** is a kind of **THING** with
UNITS
RANGE]

[**AGE** is a **PROPERTY** with
UNITS: YEARS
RANGE: 0-120]

[**HEIGHT** is a **PROPERTY** with
UNITS: METRES
RANGE: 0-2.5]

[**WEIGHT** is a **PROPERTY** with
UNITS: KILOGRAMS
RANGE: 0-150]

[PETER is a **PERSON** with
AGE: 36
HEIGHT: 1.85
WEIGHT: 75]

In this example, the frame representing person indicates the various attributes of a person: age, height and weight. Each attribute is also represented as a frame with its own properties which are defined in 'slots'. The acceptable range for the age of a person is defined in the person frame to be between 0 and 120 years, and the actual use of the age slot in the definition of Peter is that it is given the data value 36.

Sometimes this can be extended to include default reasoning, where actual values are suggested in the frame to be used if no other information is forthcoming.

The problems associated with frames are mainly concerned with the movement from one frame to the next. Dreyfus and Dreyfus (1986) suggest that rules are required to determine when this event should occur. The relationship between one frame and another is therefore defined by rules and has the same advantages and disadvantages of rule-based systems described above.

Frame systems are appropriate where classes of objects (rather than individual objects) are relevant to the problem domain and where members of a class may be slightly different. Since a frame allows such dissimilarities to be represented within a single type of data structure, inferences can be made independently of the detail of any particular object.

An example of this is the consideration of the width of a bus lane. In a rule-based system, if an engineer interprets a width of a bus lane of 3.5 metres differently from one of 3.4 metres it is necessary to increase the number of conditionals. It is unlikely, however, that every aspect of a bus lane is sensitive to this difference. A frame-based system would be an advantage in this case because it need only differentiate between the two widths when it is absolutely necessary to do so: in other cases, the 'bus lane' frame could be used independently of the value in the 'width' slot.

Hendrickson *et al* (1987) suggest two frame types: variable frames and equation frames. Four types of slot are postulated in the former case:

is-a	to indicate the equation frame from which the instance of the variable has been derived;
value	the current value of the variable;
units	in which this variable is measured; and

source-value

which indicates whether this value is calculated, input or unknown.

Slots for equation frames include:-

is-a input arguments used during the equation process;

calculation-rules

rules for each of the possible instances of the equation; and

status indicator of previous instances of the use of this equation.

Hendrickson *et al* also point out that the use of control rules adds to the efficiency of the method, which would imply at least some agreement with the postulation of Dreyfus and Dreyfus noted earlier in this section.

4.2.2.3 Blackboard systems

Blackboard systems are more complex still, but are based on a simple paradigm. The paradigm is one of a group of experts sitting near a blackboard, each writing their own ideas about the problem on the blackboard. Successive entries may suggest that previous ones are now irrelevant, or that they are confirmed. Similarly, each new piece of information is assessed to see which of the currently known knowledge could possibly use the new information. The computer model of such a blackboard system places every attempt at the firing of a rule into a special part of the computer memory known as the blackboard. The problem with this type of system is that it is extremely memory-hungry and slow. The blackboard architecture advocated by Hayes-Roth *et al* (1983) consists of three parts: the Plan, the Agenda and the Solution.

The first activity of the group of experts is to define the problem and to decide on a general approach to adopt in order to solve the problem. In blackboard architecture, this is known as the PLAN. As the plan is executed, data are needed. For each data input, there exists a number of possible actions. These actions can be compared with the plan to eliminate those which are not relevant to the problem. Those actions that remain are referred to as the AGENDA which therefore contains all the potential actions awaiting execution. The results from each of these actions are analyzed and the resulting decisions and any hypotheses derived from them are referred to as the

SOLUTION. Clearly, as the decision process progresses, each part of the architecture changes (just as in the paradigm, the actual writing on the blackboard is altered, erased or supplemented during the group discussion).

An example of this approach may help to clarify this procedure. The approach to the design of a bus lane may be defined in terms of the identity and sequence of data inputs. For example, in order to design a bus lane it may be necessary to know the commercial speeds in the peak and off-peak periods, the number of buses per hour and the passenger demand. This list is placed in the plan. The agenda then has a list of all program elements in which these elements feature. As each data value is obtained, the agenda is updated so that the list of future actions can be modified to suit the particular value in question. For example, knowledge that the number of buses per hour is 500 would allow all the program elements which depend on lower bus flows to be removed from the agenda. The solution is also updated to add any new decisions and hypotheses which may be inferred from the knowledge of the new data.

Thus the Plan elements describe the general approach that the system will use to solve the problem. The Agenda sets out all the potential actions awaiting execution while the Solution stores all the possible hypotheses and decisions that the system has already made.

Blackboard architecture is useful where the actual course of the inference and decision process is important, and needs to be transparent to the user. At any stage in the process it can identify what it has done, and what it is intending to do. The difficulty arises when the blackboard becomes very large, as the various databases, inference systems and explanation facilities can occupy a large amount of memory. The result is that systems using this architecture tend to be slow and memory-hungry.

4.2.2.4 Summary

It is thus clear that each representation system has its advantages and is capable of representing certain types of knowledge. The conclusion to be drawn from this brief analysis of knowledge representation systems is that a comprehensive decision model requires some aspect of all three types of representation system. The method to be used in a particular case is dependent on the type of knowledge and the type of inference required.

4.3 INFERENCE SYSTEMS

4.3.1 Introduction

Inference is the central activity of a knowledge-based system, and concerns the techniques used to infer decisions from data. The methods of inference used in a particular system are dictated by the knowledge representation system used, and thus the techniques are allied to the various representation systems described in the preceding section.

Kulikowski (1980) lists five emphases in Artificial Intelligence work concerned with medical diagnosis systems, and these sum up the traditional view of inference taken at that time. First the knowledge base should be clearly separated from the reasoning and control strategy which is in turn also separate from the current data. Second, the knowledge base consists of production rules. Third, careful knowledge representation helps to justify both the course and the outcome of the consultation. Fourth, there should be flexible strategies for controlling the reasoning process and last, there should be a good user interface.

This indicates the importance of knowledge representation and the structure of the system: inference is only seen as a function of the relationships between rules. Thus the consequent of the consideration of two facts is determined by the representation of their combination. The emphasis on production rules dictates the type of inference used. With knowledge represented in the form of production rules, inference is the way in which a set of conditionals is combined in order to produce a consequent. Thus inference is performed by the set of operators ('AND', 'OR', 'IF','THEN') described in Section 4.2.2.1. Since rules contain these operators as a matter of course, inference is predetermined by the particular structure of the rules: thus it is a result of the knowledge representation scheme itself. The fourth emphasis (a flexible reasoning process) is difficult to obtain outside the rules themselves. As mentioned above, the knowledge base is a part of the knowledge-based system that is separate from the procedural part of the program. The inference process is controlled by another part of the program which determines the order in which the rules should be fired and the basic strategy to be employed by the program . This could be, for example, whether the method is to prove an existing hypothesis, or to determine what alternative solutions may be available as a result of existing knowledge of the data.

This simple view of knowledge-based systems has become a little outdated. Spiegelhalter (1986), for example, argues that since this time it has become evident that while predicate calculus is suitable for the representation of deterministic systems such as legislation, medical systems require methods which can qualify the knowledge and current beliefs during the consultation.

This view is also stated more recently in Pawlak *et al* (1990), where the distinction is drawn between systems where the knowledge is sufficient and precise (that is without uncertainty), and those such as the physician or business manager where it is difficult to model unambiguously their decisions where the knowledge is often imprecise and incomplete. The question of the treatment of uncertainty and imprecision in knowledge-based systems is considered in more detail in Section 4.4 below.

4.3.2 Inference strategies

4.3.2.1 Introduction

Most knowledge-based systems structure the knowledge base in the form of a decision tree. The reason for this is that it is perceived that an expert, when making decisions, either searches backward from some hypothesis to find the causal situation (backward chaining) or searches forward from the current situation to the eventual goal (forward chaining).

4.3.2.2 Backward chaining

Backward chaining search processes are restricted to those paths that are relevant to a particular goal. This has the advantage that the search space is constrained, but also the disadvantage that the choice of possible solutions is restricted to those defined by existing rules.

4.3.2.3 Forward chaining

On the other hand, forward chaining tends to open up the possibility of a wider choice of final decisions, but can produce a memory problem through the generation of a large number of possibilities which may result in combinatorial explosion.

In either case, each decision point is represented by a node on the tree. The options available at the node are tested and the appropriate path is chosen. The difference between the two methods is one of direction: the backward chaining process starts with a particular goal, and moves down the decision tree testing the data against the knowledge base in order to prove that the goal can be supported. The forward chaining process starts with the known data, and moves up the decision tree testing the knowledge base to ascertain which goals are supportable with the known data. Figure 4.1 serves to illustrate the point.

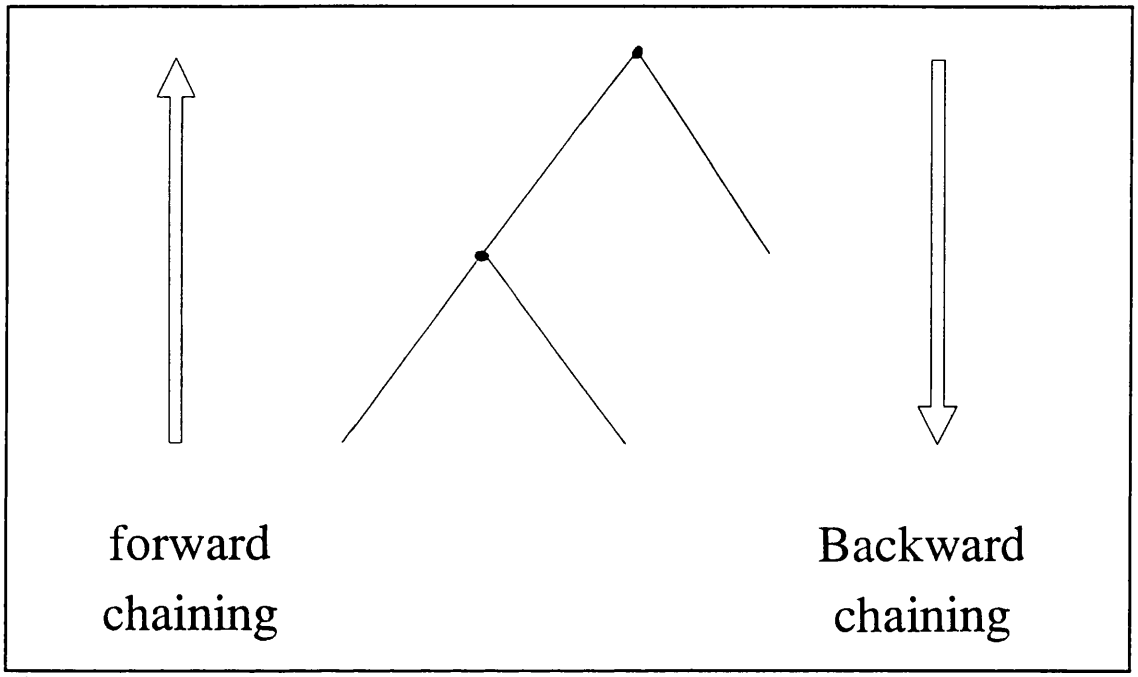


Figure 4.1 A simple decision tree, showing the backward and forward chaining processes

4.3.2.4 Search algorithms

To make the search more efficient, search algorithms are derived which reduce the search space. Hayes-Roth *et al* (1983) suggest that there are three major categories of search algorithm: depth-first, breadth-first and best-first.

Depth-first search tests all the nodes along one branch of the tree for success. When a node fails the test, another branch is tried. This procedure continues until the conclusion is reached or the whole system is unable to find a conclusion and fails. Breadth-first search tests all possibilities available from the current position in the tree before moving to the next level of nodes. All successful tests at one level are carried forward to the next and each path is tested again. This procedure also continues until a successful conclusion is obtained, or the system fails.

These search strategies are grouped together by Stefik *et al* (1983) as 'blind search' techniques because they have no limiting algorithm beyond the success or failure of the rules. The search continues until all possibilities are exhausted.

A refinement of blind search presented by Stefik *et al* is heuristic search. This aims to stop when a satisfactory solution is reached. One example of such a search algorithm is the best-first search suggested by Hayes-Roth *et al* (1983). Best-first search attempts to find the best solution at each node. This is usually much faster than the other methods, but the risk is that some solutions which are possible may be missed. In particular, the situation where the best overall solution is one consisting of a particular set of suboptimal solutions is likely to be missed unless there is a suitably clever algorithm which can identify the best choice overall.

Obviously, combinations of these methods are used in knowledge-based systems. EDESYN (Maher 1989), for example, uses depth-first search for the first four node levels, and asks the user whether this should continue or whether another branch should be attempted. Thus the user is given the opportunity to initiate a breadth-first search strategy.

4.3.2.5 Synthesis : Analysis

Winston (1984) divides rule-based systems into two types, synthesis and analysis, and suggests that synthesis systems tend to use forward chaining inference while analysis systems tend to infer solutions using a backward chaining process. This seems a reasonable conclusion, since synthesis tends to start from data describing some current situation and aims towards a (possibly) changed situation in the future. Analysis, on the other hand, examines the current situation and attempts to discover how this situation could have come about.

4.3.2.6 Global : Local

It has been suggested that inference can be divided into global and local (for example, Jackson (1986)). Global inference is generally active throughout the program: it could for example infer solutions on the basis of the maximization of some value which is obtained as a result of the tests performed on the rules (for example, truth values, or the support values described in Section 4.4.3.2). Thus global inference is largely domain independent. Local inference, on the other hand, is dependent on the particular domain being investigated. An example of local inference might be that since the width of a bus lane is a function of the width of the buses using it, the appropriate width can be inferred from knowledge of the width of the buses. Since, however, the domains of most knowledge-based systems are narrowly defined, this distinction can be difficult to maintain within a particular system.

Another way of approaching inference is to divide it into two areas: strategic and tactical inference.

4.3.2.7 Strategic : Tactical

Strategic inference refers to that part of the decision process that determines the actions to be taken. This is therefore derived from the concept of a PLAN and AGENDA used in blackboard architecture and described in Section 4.2.2.3. Blackboard architecture is the only one of the three types of representation system described in Section 4.2.2 to approach strategic inference explicitly. The other systems either approach strategic inference implicitly, or induce a strategy through the design of the rules and frames themselves. It is interesting to note that the third emphasis given by Kulikowski and noted in Section 4.3.1 (that careful knowledge representation helps to justify both the course and the outcome of the consultation) suggests that the strategic inference is implicit in the rules being used.

Rule-based systems use a pattern matching system of inference. The conditional part of the rule is tested in two ways. First, the database of information already obtained is searched for values that indicate that the conditional has already been satisfied. If this fails, the user is asked to supply the appropriate information. As mentioned in Section 4.2.2, the conditional can contain calls to another rule, and this is how strategic inference is performed from inside the rule base itself. Thus the fact that each conditional is tested in turn implies that the order in which the tests occur is a strategy.

This implicit approach to strategy is adopted by having one major rule which uses this predetermined order of testing. Thus the way in which the inference occurs is dictated by the first rule to be tested, which is thus a predicate form of a procedural algorithm. The way in which a rule may be used as a strategy manager is addressed in Chapter 7.

The frame system similarly requires some instruction for the order in which the frames are used (or whether they are used at all). Again, this can be determined by a 'master frame' or by the use of attributes and slots to call subsequent frames (as in the example above). Dreyfus and Dreyfus (1986) suggests that frames may also require some form of rule-based strategy to avoid the combinatorial explosion caused by the generation of frames as a result of changes in the course of the modelling of even a simple human decision process.

Tactical inference refers to the way in which one decision affects another. The outcome of one decision therefore determines which rule is to be fired next. Thus it can be seen that tactical inference is similar to strategic inference, but on a more microscopic scale. Tactical inference, like strategic inference, requires not only that the order of the elements in the conditional is determined, but that it is possible that the output of one element can be the input to another.

The global/local and strategic/tactical analyses of inference are not necessarily mutually exclusive. The latter determines the particular sequence of tests to be made by the inference process, while the former determines the degree by which the results of tests are available for inference purposes within the program as a whole.

4.4 THE TREATMENT OF UNCERTAINTY AND IMPRECISION

4.4.1 Introduction

The treatment of uncertainty and imprecision is one of the most critical areas of concern in the modelling of any decision process. If all issues concerning an outcome were certain and precise, the subsequent decisions would be predetermined, and the modelling of decisions, if required at all, would be simple. Most problems for which experts are required are, by definition, complicated by the difficulties associated with the lack of certainty and precision inherent in the generation of data (Gaines and Boose 1990). Some systems are further complicated by a choice of solutions even where the data are known with total certainty and precision. Since knowledge-based systems are perceived to be appropriate where such problems exist, it is necessary to approach the treatment of uncertainty and imprecision in order that all decisions take them into account.

There are three issues to consider. The first is the generation of the expert's appraisal of any uncertainty or imprecision inherent in the expertise. The second is the propagation of this in the light of data being provided by the user. The third is the analysis of any uncertainty and imprecision within the data itself. These issues apply to any decision modelling process, and will be discussed within the context of the discussion of the techniques used for their treatment within knowledge-based systems.

The use of the word 'uncertainty' within the expert systems field has been notable for the breadth of its interpretation. Spiegelhalter (1986) suggests that the term is used whenever strict logical reasoning is not possible, whether this is due to weakness in the knowledge base, unreliability in the data or because of stochastic relationships between propositions. In this thesis, the use of the term 'imprecision' is preferred when considering issues pertaining to data. This is to avoid confusion with the statistical meaning of uncertainty and its connotations with respect to probability theory.

4.4.2 Uncertainty and imprecision: the use of probability theory

4.4.2.1 Degrees of belief

Since its inception, probability theory has used real numbers to describe uncertainty and it is no surprise, therefore, to find that the early expert system builders also turned to numerical devices to describe uncertainty. Liu and Gammerman (1990) state that the theory of classical probability is usually weakened for such applications and cite the use of Bayesian formalism in PROSPECTOR (Duda *et al* 1978), where coherence is not required, as an example.

Propagation of probabilities is used in PROSPECTOR to determine the likelihood of mineral deposits given the analysis of cores obtained from the site under investigation. Probabilities are mixed with logical operators to calculate two likelihood ratios. One shows the increase in the odds of the consequent being true if the conditional is true and the other shows the effect if the conditional is false. This methodology does not follow all probability theory, however, and coherence is not required.

MYCIN (Shortliffe 1976) uses two measures to qualify propositions about which a level of certainty may not be total. Each proposition has a 'measure of belief' and a 'measure of disbelief' attached. The measure of belief is a measure of the increased belief in a hypothesis on the basis of particular evidence. Correspondingly the measure of disbelief is the increase in disbelief in the hypothesis based on the same evidence (Buchanan and Shortliffe 1985). Rule 4.3 above contains the value 0.8 which is the measure of belief in the consequent. The measure of disbelief is not written explicitly in the rule, and is taken to be zero. These values represent respectively the measure of belief in this hypothesis on the basis of the evidence established during the proof of the rule, and the measure of disbelief.

These values were obtained from the experts (the problems associated with the derivation of such values are discussed in Chapter 5). These are combined to form a certainty factor which gives, in a single number, a value for the evidential strength of this hypothesis which can then be compared with that for other competing hypotheses. Measures of belief and disbelief are constrained to be between 0 and 1, while certainty factors may have values between -1 and +1. The procedure for combining these measures is described in detail in Buchanan and Shortliffe (1985). Buchanan and Shortliffe also note that the experts were often willing to state degrees of belief in terms of conditional probabilities, but 'were unwilling to follow these assertions to their logical conclusion'. While considering that the confirmation of a hypothesis is close to a conditional probability (at least to the extent that this is sufficiently represented by a certainty factor) Buchanan and Shortliffe also state that this defies analysis as a probability measure.

Thus even in PROSPECTOR which makes some use of probability theory, it seems that it is necessary to waive some of the constraints of classical probability theory, and to introduce other methods in order to provide sufficient flexibility within the inference process. MYCIN has a whole calculus of belief and certainty which, although it has some similarities with classical probability theory, does not use probability calculus for its propagation.

Cohen *et al* (1987) suggest an approach towards the issue of combining evidence to provide support for solutions, and also a method for changing the resulting matrices to suit an expert's own interpretation of the same combination. The approach of Cohen *et al* is described probabilistically in a form which defines how the support for an hypothesis changes as evidence is accumulated. The obvious attraction of this type of methodology is that it fits within the confines of established methodology such as Bayes' Theorem. Although the methodology is described using degrees of belief and probability as alternatives with equal bases, there is no discussion of the validity of the use of probabilities to describe imprecision.

4.4.2.2 Imprecision

Imprecision has been quantified using probability theory, through the use of the concept of 'subjective probability' in which probability values may be determined on the basis of the assessment of chance (Ramsey 1926). Zwick and Wallsten (1990), however, note that the use of probability in this way carries an implicit assumption that all imprecision is due to randomness. Zwick and Wallsten suggest that there are in fact two sources of imprecision: stochastic uncertainty and linguistic inexactness. Stochastic uncertainty allows for the situation where the value of a variable is a result of a stochastic process. A proposition is linguistically inexact if it cannot be evaluated with respect to a given scale. Zwick and Wallsten give two simple examples which serve to illustrate the difference between these concepts. The first is an exact proposition which is uncertain: 'it will be 4 degrees Celsius tomorrow'. The second is a certain proposition which is linguistically inexact: 'it is warm now'.

4.4.3 Uncertainty and imprecision: the use of fuzzy set theory

4.4.3.1 Introduction

Engineers often use words such as 'high', 'fast', 'busy' 'a lot' and so on to describe aspects of the data available for the design process, and they are able to use such expressions in their inference process. Rather than attempting to convert these ideas into some numerical form (thus carrying a risk of error which may be propagated through the inference process) it may be preferable to perform the inference process using these qualitative descriptors.

4.4.3.2 Linguistic variables

A theory which attempts to describe vagueness, imprecision and uncertainty is fuzzy set theory which was introduced by Zadeh (1965). Fuzzy set theory is based on classical set theory which is centred on the concept of membership of a set. A set is a group of elements which have some common attribute upon which membership of the set depends. In classical set theory, an element is determined to be a member of the set if the attribute is present. In all other cases the element is not a member. A membership function is defined which takes the value of 1 if the element in question is a member of the set and 0 in all other cases.

Sometimes it is not possible to be precise about the membership of a set, for example when the attribute upon which set membership depends is qualitative or imprecise. Fuzzy set theory addresses this problem by allowing the membership function to take any value between and including 0 and 1. A fuzzy set is thus made up of elements, each of which has an associated membership value which indicates the fuzzy support for its membership of the set.

Thus the set is defined as a set of ordered pairs where the first element of a pair is the membership value and the second is the value of the particular element of the set:-

$$\Omega_A \;=\; \mu_1 \mid x_1(A),\ \mu_2 \mid x_2(A),\ \ldots,\ \mu_n \mid x_n(A) \qquad \textbf{(4.2)}$$

where

Ω_A	=	set of ordered pairs denoting membership of the set A	
μ_i	=	value of membership for element x_i of set A	
$x_i(A)$	=	element i of set A	
i	=	index	$\{i:\ 1, 2, \ldots, n\}$

A set can be used to describe a linguistic variable such as '*number*'. In this case the values of x would be real numbers. Definition of a descriptor, such as '*few*' is then possible, by defining a set which has elements x_i('*few*') with μ_i representing the evaluation of the membership of a value of x of the set '*few*'. In this way particular numbers can be described in terms of the descriptor '*few*'. Thus, for example, if '*number*' were defined to be the set of integers between 1 and 10, the following set of ordered pairs could be defined to describe '*few*':-

$$\Omega_{few} \;=\; .4 \mid 1,\ .8 \mid 2,\ 1 \mid 3,\ .4 \mid 4 \qquad \textbf{(4.3)}$$

The description of the linguistic variable '*number*' can be enhanced by the use of additional descriptors, such as '*several*' and '*many*':-

$$\Omega_{several} \;=\; \{\ .5 \mid 3,\ .8 \mid 4,\ 1 \mid 5,\ .8 \mid 7,\ .5 \mid 8\ \} \qquad \textbf{(4.4)}$$

$$\Omega_{many} \;=\; \{\ .4 \mid 6,\ .6 \mid 7,\ .8 \mid 8,\ .9 \mid 9,\ 1 \mid 10\ \} \qquad \textbf{(4.5)}$$

This simple example raises three issues:-

1 Values of the linguistic variable (x) may have supports for membership of more than one set. For example, the number 3 appears in the set definitions of both '*few*' and '*several*'. Fuzzy sets can contain element values which appear in other sets and set membership is therefore not exclusive.

2 Full membership of a set is transitory. The number 3 has a support of 1 in the set '*few*', while the number 1 has a support of only 0.4. It is quite possible that one number is too low to be considered to be '*several*', while another is too high.

3 The support values are assigned arbitrarily.

These issues will be discussed in more detail in Chapter 5. Set operations are exactly the same for fuzzy sets as they are for classical sets, thus set operators such as conjunction and disjunction are represented by minimum and maximum operators respectively (Zadeh 1965). Zadeh also suggests that a descriptor may be qualified, for example by the term '*very*'. Zadeh defines the qualification of '*very*' as concentration in which the support values for the descriptor (for example '*few*') are squared in order to obtain the support values for the qualified descriptor ('*very few*').

This is not altogether satisfactory, because when the support value for the descriptor is 0 or 1, the qualified descriptor has the same value. This implies, for example, that when a number is considered to be '*few*' to the extent that its membership value is 1, it is also considered - with equal emphasis - to be '*very few*'. This could be important if, for example, different decisions might be taken if a number were considered to be '*few*' rather than '*very few*'. This difficulty is discussed further, and a solution proposed, in Chapter 5.

4.4.3.3 Decisions

Bellman and Zadeh (1970), in showing the use of fuzzy logic in decision making, suggest that a broad definition of the decision concept is that a decision is the confluence of goals and constraints. This is represented logically by the intersection of the goals and constraints of the decision process. Thus with a set of goals ($G_{1 \dots n}$) and constraints ($C_{1 \dots m}$) a decision D could be represented by:-

$$D = G_1 \cap G_2 \cap \dots \cap G_n \cap C_1 \cap C_2 \cap \dots \cap C_m \tag{4.6}$$

Both goals and constraints are assumed to exist within the same search space and this suggests that logical formulae (which assume functional equivalence of propositions) can therefore model the combination of goals and constraints inherent in the decision process. Semantically, this formalism may be represented by the word 'AND', so a decision is represented as the result of:-

'goal 1 AND goal 2 AND ... AND goal n AND constraint 1 AND constraint 2 AND ... AND constraint m'.

The support for a decision is therefore the result of the equivalent logical operations performed on the supports of each goal and constraint. Thus the supports for the goals, constraints and decision in Equation (4.6) are defined:-

$$\mu_D = \mu_{G_1} \wedge \mu_{G_2} \wedge \dots \wedge \mu_{G_n} \wedge \mu_{C_1} \wedge \mu_{C_2} \wedge \dots \wedge \mu_{C_m} \tag{4.7}$$

where

μ_D	=	the support for decision D
μ_G	=	the support for goal G
μ_C	=	the support for constraint C

with the intersection operator represented by the minimum operator as in classical set theory. This form of use for fuzzy logic has also been used to model decisions in a multi-attribute problem, (for example, by Efstathiou and Rajkovic 1979).

4.4.3.4 Relations

It is necessary to consider how such a decision may be represented when the goals and constraints are themselves defined to be fuzzy sets. This can be achieved using the concept of fuzzy relations, and will now be described.

Zadeh (1965) defines a fuzzy relation, R, as a fuzzy set in the product space resulting from the consideration of a set of ordered pairs, where each element of the ordered pair is the value of a descriptor. Each ordered pair has an associated membership value. Thus, for example, the relation $x > y$ has associated membership values which depend on the particular values of x and y and which can be represented by:

$$R_{xy} \triangleq \mu(x,y) \mid (x,y) \qquad \textbf{(4.8)}$$

A relation can therefore be represented in the form of a matrix, where the rows refer to values of the first element (x in this case) and the columns refer to the second element (y in this case) and with each cell containing the membership value associated with the particular x and y values. Zadeh points out that it is not necessary for a relation to be limited to pairs of elements. The same principle holds for any *n-ary* fuzzy relation. In this thesis, however, only binary relations are considered.

4.4.3.5 Composition

Since relations represent descriptors, it is necessary to combine these in order to obtain a decision. In this sense a descriptor is equivalent to the goals in Equation 4.7. To combine binary fuzzy relations, the composition of two fuzzy relations is defined as a fuzzy relation whose membership function is related to the two constituent fuzzy relations by:-

$$\mu_{A,B}(x,y) = Max_v \, Min \{ \mu_A(x,v), \mu_B(v,y) \} \qquad \textbf{(4.9)}$$

where

A,B	=	fuzzy relations
(x,v)	=	ordered pair of elements in fuzzy relation A
(v,y)	=	ordered pair of elements in fuzzy relation B
(x,y)	=	ordered pair of elements in the relation between A and B

It should be noted for future reference that such a relation matrix is only a summary of the full relation between the elements defined by each of the two relations.

Jowitt (1984) discusses the use of fuzzy logic to assist in decisions about water quality. This often involves conflicting constraints (for example between the quality required for drinking and the quality resulting from deposits of chemicals). These constraints can be defined in terms of measurable quantities, but for decision purposes concepts such as '*poor quality*' and '*good quality*' are used by water quality engineers. Jowitt and Lumbers (1983) show that it is possible, using an adaptation of the Zadeh model, to obtain a linguistic description of imprecise data values which can then be used in the decision process. Jowitt and Lumbers suggest that composition may be used to calculate the result of the combination of an unary relation and a binary relation. Thus a fuzzy set (which is an unary relation) could be composed with a matrix. A fuzzy set of supports for values of an input variable (X) can therefore be composed with a binary relation where this input variable is related to an output descriptor (X,Y).

Since conjunction takes the minimum of the two inputs, this can be used to represent the worst possible case of combination of supports for X and the relation between X and Y as expressed in the matrix. Disjunction, being a maximum operator, is then used to select the best of the resulting supports to describe the output Y. Composition is thus a maximin analysis of the fuzzy supports of the input variable and the relation matrix. In particular, composition seems to be a reasonable framework in which to model decision making processes, since it is a logical representation of the effect of testing the set of supports for a data value against knowledge expressed as a fuzzy relation between data values and a solution.

The use of such a relation to represent the evaluation of data is useful when the outputs are in the form of linguistic values such as 'low' or 'high'. The relation can be used to evaluate actual data values in terms of the linguistic descriptors so that it is possible to obtain a qualitative description of the data set, together with the appropriate set of support values. The way in which this is used to represent expertise and to assist in the design process is described below in Chapters 6 and 7.

4.4.3.6 Conclusions with respect to fuzzy logic

The use of fuzzy supports is more appropriate for the treatment of imprecision than uncertainty since they are almost expressly designed for the purpose of describing concepts that are inherently imprecise (Zadeh 1975a, 1975b, 1975c) such as linguistic descriptors (for example, '*high*', '*low*', '*a lot*').

One of the principal difficulties with fuzzy set theory and its associated calculations, lies in the third issue identified in Section 4.4.2.2: that the actual values used for the membership functions are obtained arbitrarily. Fuzzy set theory is not alone in this respect: the measures of belief and disbelief used by Buchanan and Shortliffe (1985) and described in Section 4.3 are obtained from doctors who are required to state their feelings about the quality of a rule in terms of real numbers between zero and one. The propagation of these numbers eventually results in some decision being taken in preference to another. The difficulty lies in the fact that these values are at best crude approximations of a quantification of some qualitative estimate. This is particularly severe if there is more than one expert involved in the process because each one may have a different interpretation of the scaling of the numbers. Thus different sequences of rules may (if they have been derived by different experts) be qualified according to a different scale. This has the potential for inconsistency in the decision-making process.

One of the attractions of the use of probability theory to qualify rules is that the calculus exists, not only for the propagation of probabilities, but also for the derivation of initial probability values. Even when it is necessary to derive these intuitively Ramsey (1926) gives a quantitative method for establishing the probabilities from calculations of chance. It should be noted, however, that the initial subjective estimates of the odds may also be arbitrary.

In order to be consistent within a knowledge base, a method is required to establish the initial values (whether these are probabilities or fuzzy supports). This method should be used by all experts who contribute to the knowledge base and should seek to minimize any interpretative differences that may occur between experts. Such a method is presented in Chapter 5.

4.5 CONCLUSIONS

From Chapter 3 above, it is clear that the design of bus priority systems is a potential application area for knowledge-based systems. This is because the expertise involved with this design problem is heavily dependent upon the expert's interpretation of the existing situation and of the possible solutions. The discussion in this chapter suggests that such interpretation may be described in terms of logical relationships. The site-specific nature of the high capacity bus system design problem means, however, that it is difficult to induce rules from past experience since these are likely to be too brittle, and likely to change for conditions at different sites.

To attempt to model this aspect of the expertise using production rules would be likely to generate a very large rule base, possibly with arbitrary conditionals to define different consequents from the same data set. As discussed above, large rule bases tend to cause practical problems in the implementation of knowledge-based systems.

There is a place for some form of object definition in the knowledge representation. A frame system would help to maintain the concepts under discussion, making the associated differences within classes of object less likely to cause the system to fail. The use of a frame-based system does not, as has been shown in this chapter, reduce the need for rules (although it should reduce the number of rules necessary to make decisions based upon fine distinctions between different data values). There is, however, a difficulty with the use of either of these representations.

In the analysis of the expertise in Chapter 3 above it seems that the problem definition and solutions obtained are likely to be affected by the personal preferences and biases of the designer working on the project. Although there is a knowledge base which is common to all designers of bus systems (as will be discussed in Chapter 6), there is an element of preference in the expertise actually used by any particular expert. In order to represent the bus priority design process, therefore, it is necessary to use a knowledge representation and inference system that can capture this aspect of the expertise. Neither rules nor frames alone seem to have this capability, and therefore it seems that an appropriate knowledge representation and inference system needs to be devised.

Another point to be drawn from the analysis of high capacity bus system expertise is that the design process seems to be made up of three parts: data acquisition, data analysis and the design decisions. It would seem that while data acquisition is obviously an interactive process, particularly where data are vague, neither the analysis nor the design decisions themselves are particularly interactive, at least until any decisions have been reached. At this point in the process, interaction between the system and the user is important because the description and explanation of the decisions should make the whole process transparent. This has an important impact on the implementation of a knowledge-based system, and will be discussed in more detail in Chapter 6.

To facilitate interaction during data entry and the final explanation of decisions, the use of some aspects of blackboard architecture would be useful. The ability to record past decisions and their effects would mean that the user would have access to the entire decision process if required. Since the decision process itself is not interactive, the ability of blackboard systems to indicate possible future actions is not required. This helps to reduce the resources required to operate the system, but means that the use of the term 'blackboard' in these circumstances may be inappropriate.

It has been very difficult to find an example of an implementation of an expert system in which the decision process itself is explicitly modelled. In some cases the method has been to reformulate the problem to fit existing algorithms (for example, Brachman *et al* 1983). Some work simply fits a new knowledge domain into an existing knowledge-based system which has had its knowledge base removed (known as an 'expert system shell'. For example, Allen (1986), uses the inference system provided by a particular shell (M1). The criteria for choosing this shell relate to the type of computer available, the use of the same shell elsewhere in the relevant organization, its production rule basis, its explanation facility and the availability of training and support. Others use a particular computer language, such as LISP or PROLOG, to provide a knowledge representation system which seems appropriate (for example, Maher 1987). The approach in these cases seems to be to establish the knowledge representation system first, and then use the appropriate inference system.

One conclusion resulting from this review is that the decision process should be modelled explicitly if it is to be represented in a meaningful way. Such an approach models directly the inference process, which in turn determines the knowledge representation system to be used.

It seems, therefore, when faced with the type of bus priority expertise described in Chapter 3 and the methods used in knowledge-based systems introduced in this chapter, that the methods conventionally used in knowledge-based systems are not necessarily wholly appropriate for this problem. Certain facets of the existing work in this area are useful, however, and these will be used in the development of a suitable decision model.

The first of these facets is the separation of the expertise from the procedural part of the software. This is useful as it allows flexibility in the use of expertise obtained from different experts. This is discussed in more detail in Chapter 6. The second facet is the ability to trace the progress through the decision process. This is useful because it allows a full explanation about the reasoning for particular decisions to be given. Perhaps the most important facet, however, is the issue of the inference system.

The significance of the inference system is that, like the expertise, it is separate from the procedural control of the program. This means that it is dependent only on the particular inference being attempted at any particular time. The strategic control of the program can be based on the overall decision process and modelled explicitly. Inference, on the other hand, represents the process by which the expert makes the individual decisions which are required by this model. Inference can also dictate the subsequent activity of the model where such strategic changes of direction are suggested by specific decisions.

The representation of uncertainty and imprecision, discussed in this chapter, is important when the question of inference is addressed. This is because inference depends on the expert's evaluation of an existing situation and the range of solutions which may be possible from the particular point in the decision process. The discussion above has suggested that a method should be devised which can address the representation of such interpretation when the knowledge of the existing situation is imprecise and use this representation within the decision process. The derivation of a model to address this representation is the subject of the next Chapter.

CHAPTER 5
DERIVATION OF A MODEL TO REPRESENT EXPERT OPINION

5.1 INTRODUCTION

One of the conclusions reached in Chapter 4 is that there is a need for a method to represent the ability of an expert to interpret data, especially when these are not known precisely. Data which are known imprecisely can only yield imprecise decisions, yet as noted in Chapter 3 the decisions that are required are not imprecise. For example, it is possible to decide to locate a bus lane at the kerbside or to locate it in the median: at this decision level there is no compromise available - the option of locating the bus lane halfway between the kerb and the median is not usually feasible.

One of the functions of an expert in this domain, therefore, is to interpret the imprecise data in such a way as to obtain a precise decision, and to know to what extent this decision is likely to be robust. The representation of such an activity is the subject of this chapter.

In order to represent the knowledge of experts within a computer model, it is necessary to acquire the knowledge from the experts. In this research, the acquisition process was found to require a model to represent opinion and judgement. This is in addition to the requirement noted in Chapter 4 for such a model within the decision process. The role of this model within the acquisition process is discussed in Chapter 6. In order to facilitate that discussion, however, the derivation of this model is discussed first in this chapter.

This chapter has five further sections. The next section continues the discussion in Chapter 4 about the differences between imprecision and uncertainty and discusses the nature of imprecision when this is concerned with data. Section 5.3 gives the derivation of a mathematical model to represent expert opinion. Section 5.4 draws some conclusions about the use of such a model in the decision process. The use of this model as part of the knowledge acquisition process will be discussed in Chapter 6, and the use of the model in the design of high capacity bus systems is considered in Chapter 7.

5.2 UNCERTAINTY AND IMPRECISION

This section continues the discussion started in Section 4.4 about the relationship between uncertainty and imprecision. This section is in three parts, considering uncertainty in the first part, and imprecision in the second. The third part draws some conclusions from the discussion in the first two parts and, considering how an expert may treat these matters, sets out some requirements for a model to reproduce this process within a decision-making system.

5.2.1 Uncertainty

The concept of uncertainty is within the domain of the theory of statistics. Hamburg (1985), for example, considers statistics to be the body of theory and methodology for drawing inferences and making decisions under conditions of uncertainty. This body of knowledge can also be extended, according to Hamburg, to deal with extensions of these inferences beyond the particular set of the available data. Hamburg also states that statistical analysis does not replace the use of intuitive judgement, rather that it assists in structuring a problem and in bringing the application of judgement to it.

Underpinning statistical analysis, however, is the implicit assumption that there is a value (which may or may not exist within the available data set) which is the 'true' value. The other values are different from this 'true' value as a result of error - error in the measurement of the data or error as a result of some stochastic process.

5.2.1.1 Errors - data

One form of error arising from measurement is systematic bias. For example, if a large number of pieces of string were measured with a measuring device which was incorrectly calibrated, a systematic bias would be evident because every measurement would contain the same error. The other type of measurement error which arises is a result of differences in interpretation of the information given by the measuring device, even if this were perfectly accurate. This error is as likely to result in overestimates as it is to result in underestimates, and therefore is subject to a stochastic process.

Apart from measurement error, other errors appear as a result of the stochastic process which arises because the sample contains examples which are not exactly the same. For example, suppose that each piece of string within the set of a large number of pieces of string is supposed to be exactly one metre in length, and a perfect measuring device exists which has no bias, and its information is perfect. Using this device, it is possible to establish that not all pieces of string are exactly one metre in length - some are slightly more and some are slightly less. This error is not due to measurement error, but to the stochastic process which underlies the selection of the particular pieces of string.

5.2.1.2 Errors - models

The errors discussed so far in this section are all related to the data which describe an existing situation. Errors also arise as a result of the use of these data in models, for example in the prediction of future events. Alonso (1968) shows that there is a type of error that occurs as a result of the failure of a model to represent the real world exactly. A model is by definition a reduction of the real world problem, designed to facilitate analysis and prediction of future results. This type of error is therefore an inevitable result of the action of representing a real-world problem in the form of a model.

Alonso refers to this type of error as 'specification error' since it arises as a result of the failure of the model to specify the real world precisely. A very simple model of the real world might be just a constant: for example, to predict the ambient temperature a model could be derived which is simply a constant value. In this case, the specification is very poor, and therefore the specification error is great. The specification of the model could be improved if a variable were added, for example a variable to represent the distance from the equator: thus the greater this distance the lower the temperature. The specification is now better and the error resulting from specification is therefore smaller. Adding variables in this way can be expected to reduce the amount of specification error. For every variable, however, it is necessary to measure something (in the example above for instance, the distance from the equator) and, as discussed above, this measurement incurs a measurement error. Thus as variables are added the total measurement error increases.

The total error which arises as a result of these two error effects is therefore their sum. The effect of the combination of these two types of error as the number of variables is increased

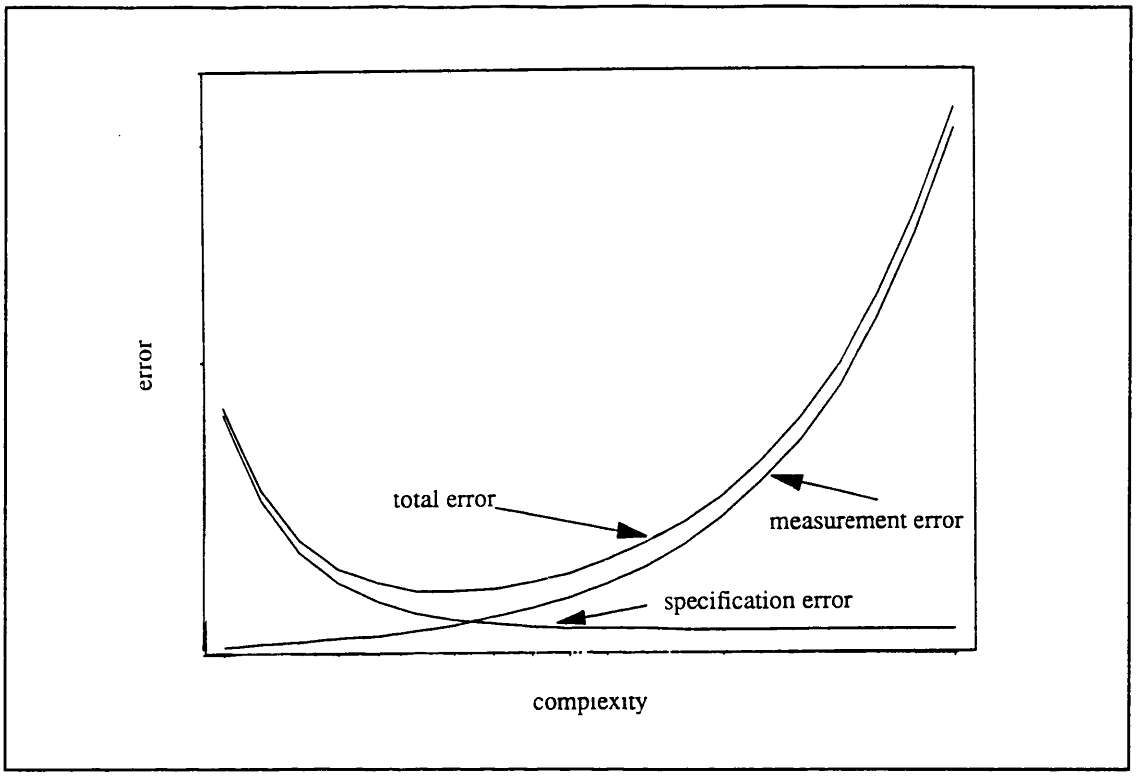


Figure 5.1 The changes in total error resulting from the changes in specification and measurement errors as the complexity of a model increases.

is shown in Figure 5.1, where the complexity of a model represents the number of variables used. It can be seen in Figure 5.1 that there is a point at which the total error is at a minimum, beyond which any increase in the number of variables is only likely to increase the total error. Since there is, in effect, a limit on the number of variables, it is important to consider which variables it is useful to include in the model, and which could be ignored. The Alonso model does not consider the possibility that the measurement error is not constant for all variables. Some variables might be extremely difficult to measure, for example, and the resulting measurement error could be very large. Other variables might incur a much lower level of measurement error.

Alonso shows, however, that the propagation of errors within a model can have significant effects on the outcome. These effects depend on the type of calculation performed with the data concerned. This does not alter the principle, but it may alter the rate of change of the error concerned. Thus the number of variables within a model is also affected by the complexity of the mathematical operations within the model itself.

Another aspect to be considered in the choice of variables to include in a model is that all variables do not convey equal information in terms of the output of the model. The Alonso model indicates that the number of variables is limited by the minimization of the total error. The resulting implication is that if the number of variables is to be limited, then the variables which should be included should be chosen on the grounds of their importance to the overall output of the model.

5.2.2 Data

5.2.2.1 Data and probability

Data might be considered to be a set of observations of some aspect of the real world. For example, a number of observations could be made of the commercial speeds obtained within a particular corridor. From these data, it is possible to assign probability values which describe the probability of the occurrence of each value. Jessop (1990) notes that a complete range of such values, having been assigned probabilities, may be considered to be a set of events. This set of events has two properties: the events are mutually exclusive, and they are mutually complementary. Thus only one of the events may occur, and one must occur. Data recorded to assess the commercial speed along a corridor, for example, may be considered to be such a set of events and a probability distribution obtained. The difficulty is that the database required to support these probability statements is likely to be large, and includes no specific reference to the errors described in Section 5.2.1.1.

Statistical analysis studies a data set in order to establish the extent to which the observations might be considered to be some form of 'true' value (with an appropriate amount of error). If the analysis shows that this is the case, it is possible to use the true value in subsequent models, as long as the error inherent in the distribution of data values is also considered. In this way, calculations are not necessary for every data value: for example, it is possible to calculate using parameters which summarize the distribution such as the mean and the variance.

When using such values, however, some thought should be given to their derivation. The error problem, for example, is greater if the data set is small, since there is less basis on which to estimate the 'true' value. It is therefore necessary to collect an amount of data that is sufficient to support the use of summary data. Some statistical analysis addresses this problem by using tests to establish whether it is possible to consider that a particular distribution describes a data set when this is small. Nevertheless, the use of summary parameters is weaker in such cases than it is when the data set is more substantial. This problem is even more marked when the data observations are spread over a wide range.

In some circumstances, it may not be reasonable to consider that a 'true' value exists. This may occur for two reasons. First, the population from which the data set is taken may not have a 'true' value which could be used for descriptive purposes. For example, if the distribution of data values has a wide range and where the frequency of occurrence of the values is more or less constant. The use of a particular value to describe these data would be difficult to justify in this case because the dispersion of the data outweighs the use of a location measure such as the mean. Normally the solution to this problem is to enlarge the data set, but this is not always possible, and may not have the desired effect.

The second consideration is whether the 'true' value is the most useful value to use in the subsequent calculations. This is particularly marked when the range of observations is large. The question of which data value to use is not only a question of its ability to describe the data, but also of the importance of the value within the data set. An important value in this context is one that indicates a change in solution. This arises where the decision outputs are discrete, but where the data distribution is continuous. For example, the number of lanes in a roadway is a function of the traffic flow. Of the data set of observations of traffic flow, the important values are those which imply a change in the number of lanes: these values carry the implication that if the traffic flow is greater then another lane is required with all the associated consequences such as increased cost, increased space requirement and so on.

These values have no systematic relationship with other locational values such as the mean which is a descriptor of the statistical distribution. On the contrary there is no way of establishing these values statistically - they are known only as a result of the knowledge of the relationship between the data and the eventual decision (in the case above this relationship is between the traffic flow and the capacity of a lane). Important values therefore carry more information than other values in the data distribution.

5.2.2.2 Information

Information theory is not unknown to statistical analysis. The work of Shannon in developing information theory has long been accepted. It is not proposed to explain the details of information theory, but it is helpful to consider one of the concepts which underlies the use of the theory. Shannon (1948) is concerned with the capacity of a telephone system to accommodate messages. As with any other capacity problem, two possibilities exist to increase the number of messages that it is possible to carry simultaneously: increase the capacity of the system or decrease the size of the messages. The former is expensive, and the latter carries the risk that the reduction in size has the risk that the message loses its content.

Shannon observes that a message consists of two parts, information and noise. Thus if noise can be reduced rather than information, then the message size can be reduced without loss of information. Shannon reports that statistical analysis of the English language shows that the frequency of occurrence of letters is not equal for all letters - there is a greater probability that the next letter is the letter 'e' than 'x', for example. Conditional probabilities can also be calculated: if the next letter is 'q', the probability that the subsequent letter is 'u' is much higher than that for the letter 's'.

The importance of this is that it is therefore possible to predict that, for example, if the next letter is 'q', the following letter must be 'u': the use of the letter 'u' after the letter 'q' within a message is therefore redundant and the message could be reduced in size without loss of information. Thus it is possible, with the appropriate knowledge, to convey information within a message of reduced size.

The parallel situation arises within a decision process: knowledge that a certain value is a threshold above which a subsequent decision has a different result means that it is important to consider whether the population represented by a data set is likely to include values above or below this threshold. Thus data values around a threshold value convey more information than other values, and therefore are more important. Correspondingly, values which are not likely to change a subsequent decision convey less information, and more noise.

5.2.2.3 Expert evaluation of data

As discussed in Chapters 3 and 6, a major part of the expert's activity is spent in the evaluation of data. This evaluation is conditioned by the knowledge of possible solutions, and therefore it might be expected that an expert is interested in the data around decision thresholds.

Where data are difficult to obtain, for example, it is better to obtain data which indicate the relation of the population to the thresholds rather than data which convey less information.

In the expert's evaluation of the data it is not usually possible to be definite about the eventual solution because this depends on the actual values which describe correctly the real world. The expert can, however, express an opinion about the various options, where this opinion is based on experience and the assessment of the existing data. Where the expert considers that more than one solution could result from the available data, two courses of action can follow. One is that more data could be collected with the aim of establishing more accurately whether the true values are above or below the relevant threshold. The other is to opt for the solution which appears to be the best given the data, always remembering that other possible solutions exist.

How this might be represented in model form is discussed in Section 5.2.5 below.

5.2.3 Imprecision

5.2.3.1 Introduction

Following the discussion in Sections 4.4 and 5.2.1, this section considers the role of imprecision within the decision process. Consider, for example, data concerning the bus flow along a corridor. Even if the time period over which the flow is measured is fairly long, the bus flow is likely to change from period to period. Especially where bus flows are high, the time-dependent nature of the queuing process at bus stops discussed in Chapter 3 suggests that the fluctuations which arise between relatively short periods of time become more important.

In order to establish the effects of the time-dependent queuing process at a bus stop it is necessary to know the distribution of the headways of buses arriving at the stop. The detailed data required in order to establish this are not easy to obtain. The implication is that unless data are available to this level of detail, the design process cannot proceed. The problem is that it may not be feasible to collect these data, or it may appear to be futile - especially if the design may not proceed. It would be useful to know, before embarking on an expensive survey, the extent to which the subsequent design of the bus stop is likely to be sensitive to fluctuations in the bus flow as it exists in the corridor.

For example, if the bus flow is about 50 buses per hour, it is reasonable to believe that a simple bus stop design with one berth would be sufficient. A bus flow of about 300 buses per hour would require a bus stop design which allows more buses to use the stop simultaneously - the Brazilian experience discussed in Chapter 3 suggests that two groups with two berths each would provide the capacity for such a flow if overtaking were allowed. In the first instance, a bus system designer needs to know at what bus flow the design of the bus stop should change from a single berth to a design with more capacity. For example, it might be that a flow of less than 20 buses per hour - whatever the headway distribution - could be accommodated with a bus stop with one berth. Knowledge that the bus flow is about 10 buses per hour is therefore sufficient at this stage in the design process in order to have at least a reasonable idea about the design of a bus stop: more precision is not necessary.

There exists a level of bus flow, however, where the flow is sufficiently high to suggest that a design with a higher capacity is required. This level is not an exact number, and depends on specific characteristics concerned with the arrival headways, passenger demand, stop spacing in the corridor and the vehicle characteristics (as noted in Chapter 3). It is therefore not possible to predict the exact bus flow at which the design should change before these characteristics are known. Nevertheless, this level exists and at this point, a marginal change in the bus flow is likely to result in a major change in the design of the bus stop. If the bus flow is about this value, the precision with which the data are known becomes important: an error could result in unnecessary investment or insufficient capacity.

Thus two characteristics are identified which are concerned with the interpretation of bus flow data when these are related to the question of bus stop design. First, for some large ranges of data values there is no significant alteration in the design strategy and second, for some small ranges a marginal change in the data value results in a major change in the resulting design.

It is therefore necessary to encapsulate these two concepts of the interpretation process within any interpretative model. The discussion in this section shows that these characteristics are related to the degree of precision with which the data are known. If the data imply that it is possible that a design threshold might be breached, then precise knowledge could be helpful. In other cases this is less important. Clearly, it is therefore necessary to know the approximate thresholds: a model which is intended to represent this process should be able to identify the thresholds and to relate these to the use of the data.

5.2.3.2 Decision thresholds - data values between thresholds

The first aspect of data with respect to precision is the data range where a subsequent decision is unlikely to change as a result of a marginal change in a data value. In this case, knowledge of the exact value of the data is not important, and a high degree of precision is not required in order to reach a decision. It is sufficient to know that the available data are located between two threshold values.

Nevertheless, a decision based on such imprecise knowledge of the data should be treated with more caution than one in which the data are more precisely known. The degree of caution necessary is not constant: if the data concerned are far from any threshold (for example near the centre of a range between two threshold values), then the degree of caution could be less than would be the case if knowledge of the data suggested that these could encompass values near or across a threshold.

The width of a bus lane serves as an example. During this research, design experts were asked about the ideal width for a bus lane. A common response to this question was that the bus lane should be 3.5 metres wide. The question was followed with another which asked how the designer felt about the situation where a bus lane of this width was not possible, and a bus lane - if it were to be implemented - would have to be 3.4 metres wide. Again a common response was that this would be possible, but the designer had a lower opinion of this width than with 3.5 metres.

This change in the opinion level could be represented in a graphical form as a unimodal distribution which is bounded at the lower and upper limits. Such a distribution is shown in Figure 5.2. Between the bounds of the data values, which are determined by the feasibility of such values (for example, a lane width of less than 0 is not feasible), the distribution reaches a maximum at a point between these limits. It should be noted that this maximum point need not be at the mid point of the data range. For example, if the range of possible lane widths were considered to be between 0 and 10 metres and the distribution describing the opinion of satisfactory lane width had its maximum at 3.5 metres, the mode of the distribution would not be located at the mid point.

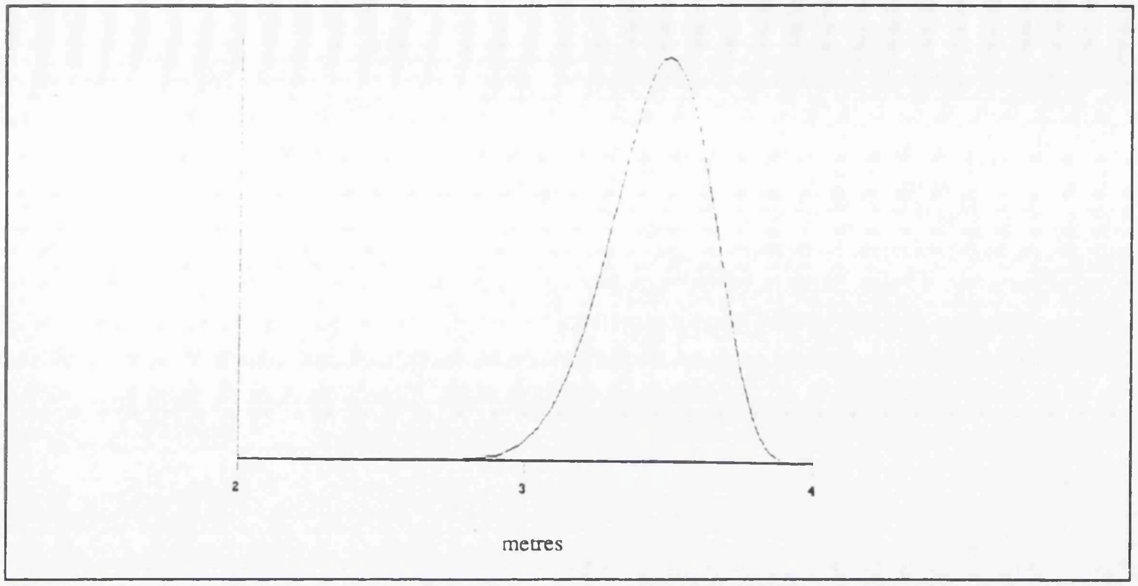


Figure 5.2 A graphical representation of the change in expert opinion of the satisfactory width of a bus lane.

5.2.3.3 Decision thresholds - data values near to thresholds

Consider the question in Section 5.2.3 above with respect to the width of a bus lane. This question could be repeated with different widths, until the opinion changed sufficiently to alter the decision from:-

"*this width of the bus lane is satisfactory*"

to:-

"*if this is the only way to have the bus lane then this width is possible, but it is not desirable*"

and from this to:-

"*this width is too narrow to be used as a bus lane*".

The first conclusion to be drawn from these statements is that the expert finds it easier to describe the points at which the decision changes than to differentiate between marginal changes between these points. The second conclusion is that there is a change in the expert's opinion between the thresholds: the opinion of a particular decision is at a maximum between the thresholds, and falls to a lower value as the data values concerned tend towards the thresholds. A third conclusion is that although the expert can identify that for different widths between thresholds there is a different level of opinion, this is not obviously quantifiable. Finally, the threshold is independent of its opinion value. It is therefore unlikely that a threshold could be identified mathematically from the data value alone.

Thus the experts find it easier to establish thresholds at which their chosen decision changes for a marginal change in the given data values than to quantify their opinion values. Near to these thresholds, the opinion level of each of the two competing decisions tends to equality. It is very difficult to choose between the two options, and knowledge of a slight change in the data could alter the decision. The opinion of the new decision is, however, only marginally different from the opinion of the previous option.

Thus it appears to be possible to use data which are only known imprecisely if it is also possible to know the relevant thresholds at which the decision options change, and around which the data carry more information for the decision process. The next section discusses how these thresholds may be identified.

5.2.4 Expert evaluation of decision thresholds

As has been discussed in Section 5.2.2, the design engineers involved in the design of high capacity bus systems appear to be able to use qualitative interpretations of data rather than actual values. The reasons for this have been discussed in Section 5.2.3 along with some of the issues surrounding the use of imprecise data. This section discusses how these qualitative interpretations might be modelled in order to represent this aspect of the design decision process.

Section 5.2.3 discussed the nature of imprecision and the importance of decision thresholds in the decision process. The use of imprecise data relies to a large extent on the definition of these thresholds. The first problem is therefore to establish the thresholds. This is in fact divided into two problems: first it is necessary to establish that the thresholds actually exist and second it necessary to establish the data values at which they occur.

Consider the question of the width of a bus lane, raised in Section 5.2.3 above. Figure 5.2 illustrates a distribution of expert opinion of the concept of '*satisfactory*' when related to the width of a bus lane. Obviously, there are widths for which the opinion of the descriptor '*satisfactory*' is very low, and thus where a different description might be better. The concepts identified in Section 5.2.3 above could each be evaluated by a distribution similar to that for '*satisfactory*', but with a different descriptor. The constraints on such distributions would be the same as those for '*satisfactory*', although they would have a different shape.

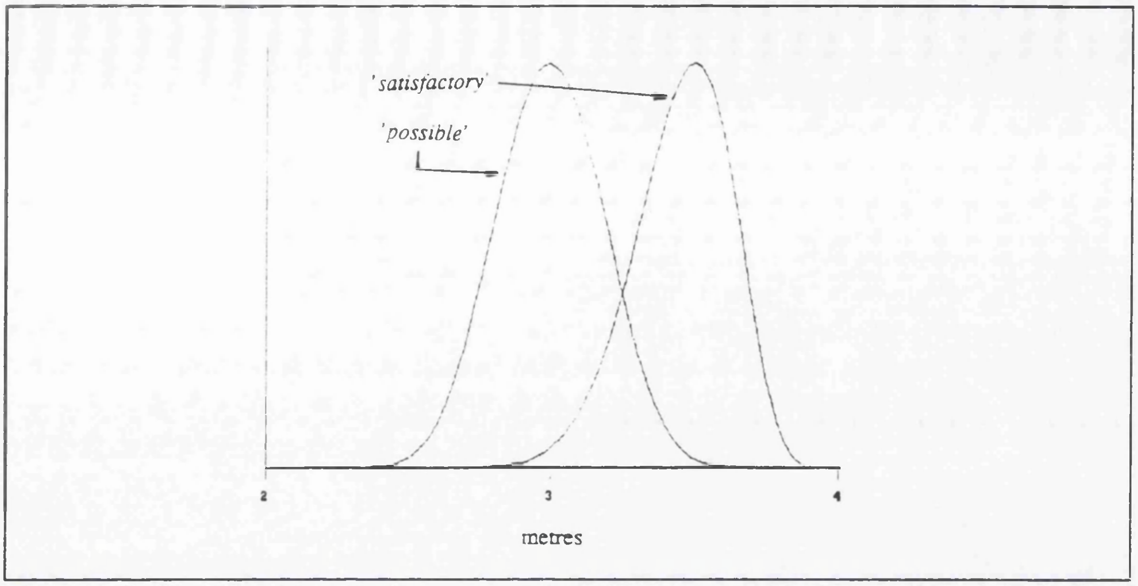


Figure 5.3 Two distributions of expert opinion about the width of a bus lane: '*satisfactory*' and '*possible but really too narrow*', showing the intersection which identifies the decision threshold.

Thus another distribution could be derived to describe another concept with respect to lane widths, for example where the lane width is not ideal, but if the choice were between having no bus lane and having a bus lane which is a little narrow it would be possible. Figure 5.3 shows such a distribution plotted with the previous distribution (shown in Figure 5.2). As expected there is a point at which the two distributions intersect. At this point a threshold is identified at which the chosen description changes as the width increases, from '*possible but really too narrow*' to '*satisfactory*'.

The relationship between these two distributions is a matter of some importance. If they are mathematically related, this could save some computational effort. On the other hand, even if they are mathematically independent, they are related semantically in the sense that they are merely describing graphically opinions of two different descriptions of the same object. There are two aspects to be considered concerned with any mathematical relationship between two distributions. The first aspect to consider is whether there is any commonality between the form of the distribution in each case. The second consideration is whether there is any mathematical dependence between the two distributions (whether of the same form or not).

In this case, however, there is more than a mathematical dependence to be considered. The distributions are not intended to describe mathematically some functional relationship between two parameters. Each distribution is a representation of the opinion held about a description of some values of a particular variable. Opinion is not measurable in the normal way, and therefore any mathematical function can only approximate the changes in opinion as the variable changes in value. This issue is discussed more fully in Chapter 6, and at this stage it is necessary to consider only that the mathematics is used to describe - rather than to define - opinion.

5.2.4.1 Distributional form

When considering the distributional form of opinion, it is helpful to consider the constraints that are required for its representation. Some have already been mentioned in Section 5.2.3 above. These are that there is an upper and lower bound to the values of the variable under consideration. This is intuitively correct: there is some value of a variable below which it is inconceivable to register an opinion of a particular descriptor. A similar condition holds for the upper limit. The range of data values between these limits which is described by a descriptor is known as the 'descriptor range'.

For example, consider the variable '*height*'. A descriptor '*tall*' could be used to describe particular heights. If considering normal adult humans, for example, a height of 0.5 metres would not be considered tall. If this were evaluated using a fuzzy set, the corresponding value of the membership function would be very low. In the sense that consideration of an even smaller height would be insignificantly smaller than this, the useful limit of the opinion distribution at the lower end of the variable range has been reached.

At the limits of the descriptor range, there exists an important issue which must be resolved before any particular mathematical distribution can be used to represent opinion. This is that the descriptor might be qualified with a hedge. Hedges were discussed briefly in Section 4.4.2, and this discussion is now continued. An obvious hedge in this case is the adjective '*very*', and for the purposes of this discussion the upper limit of the descriptor range is used for illustration.

A height of 2 metres could be described as '*tall*' (when considering normal human adults), and thus it could be expected that the value of opinion would be high in this case. If another descriptor were introduced, which is simply the descriptor '*tall*' modified by the hedge '*very*', and applied to the same height of 2 metres, two issues should be addressed. First, the availability of another descriptor allows a better definition of the original simple descriptor. This is because the descriptor '*tall*' no longer has to take into account very large values of the variable since there is now another descriptor which can do this more precisely. Second, there is a semantic issue to be considered. This is that the descriptor '*tall*' can be limited to describing a particular range, rather than describing all values above a certain level. Thus the opinion distribution which describes the opinion of '*tall*' applies up to a certain value of '*height*', beyond which it is felt that the descriptor '*very tall*' would be more appropriate. This implies that the distribution should be unimodal and bounded at each end of the descriptor range.

Whether a height of two metres is '*tall*' or '*very tall*' is a matter for personal preference, but if these are described by two mathematically independent distributions, the semantic relationship (as perceived by a particular individual) can be recorded.

Another constraint to be considered is that the distribution should be unimodal: that is it should have a mode, and that there should be only one. The existence of a mode is a matter of information. If there is no mode the distribution is uniform and no information can be obtained from knowledge that the variable has particular values. A multimodal distribution suggests that there is more than one descriptor being evaluated, and that a clearer definition should be obtained.

A third consideration is that if separate distributions are used to describe different descriptors, there should be an element of consistency so that comparisons can be made, particularly around the intersection points. Thus all distributions should be scaled equally along the vertical axis. This could be achieved by requiring that the mode is equal to 1. This provides a consistent scale for measuring opinion.

The fourth consideration is that the distribution should be smooth. There are no sudden changes in the evaluation of the opinion of any one descriptor over the descriptor range. This reflect the concept addressed above that the level of opinion does not change for marginal changes in the variable value. The question of changes in decision is addressed by the use of separate distributions, and is discussed in the next section.

Thus the distributional form of opinion can be summarized:

1 the distribution is bounded at the lower and upper limits;

2 the distribution is unimodal;

3 the mode of the distribution is equal to 1; and

4 the distribution is a continuous function.

Having considered the general constraints within which an opinion distribution should be defined, it is now appropriate to consider whether there is a formal relationship between the opinion distributions of different descriptors concerned with the same variable.

5.2.4.2 Distributional dependence

The previous section has discussed the possibility of having more than one opinion distribution to describe a variable. This section is concerned with the question of whether they may be related mathematically. Such dependence implies that once a distribution is defined for one descriptor, those distributions for the other descriptors are therefore systematically defined as a function of the first.

That there is a semantic relation is clear: the descriptor '*very tall*' implies a greater height than the descriptor '*tall*'. Thus there is a clear relationship between the two descriptors based on their meaning. The question remains, however, as to whether such a semantic relationship can be formulated in mathematical terms.

The first problem is the question of hedges which modify a descriptor which has already been described with an opinion distribution. Zadeh (1965) suggests some mathematical formalisms to describe hedges such as '*very*'. As noted in Chapter 4, however, these do not necessarily provide sufficient flexibility and there can be some potential confusion over their use. In this research, it has not been possible to identify any formal relationship to represent the effect of hedges in this way.

The second problem concerns the relationship between concepts which are not simply related by hedges. These concepts are less related to descriptions of the data under consideration (such as '*tall*'), rather they are descriptions of possible actions or states upon which these actions are to be based. Examples of such concepts are the descriptors used to describe bus lane width noted in the previous section: '*satisfactory*' and '*possible but really too narrow*'. The relationship between these two concepts is not hedged: these concepts are 'action descriptors' which are different descriptions which relate to different parts of the variable range, and which carry an implication of some action. Thus '*satisfactory*' implies that the subsequent bus lane is feasible, while '*possible but really too narrow*' implies that it would be better if the lane were wider. There is no semantic relationship between these concepts: the definition of one does not imply the definition of the other. Thus there is no apparent reason why the distributions should be related mathematically, although they could be defined by the same mathematical model.

5.2.4.3 Semantic considerations

The final question that needs to be addressed before deriving the mathematical model to describe expert opinion is the number of descriptors that should be considered. Consider first the action descriptors noted in the section above. The number of action descriptors to be considered in any one decision may be defined quite simply.

The first action descriptor is the one which defines the positive decision: the '*satisfactory*' bus lane width, for example. Every decision which leads to an action should by definition have one such positive possibility. The second descriptor to be considered is one for which the value of the variable is too low for a positive decision. For example, the descriptor '*too narrow for a bus lane*' suggests that a bus lane is not possible: the resulting decision would therefore be not to construct a bus lane. The third action descriptor to consider comes into effect because the value of the variable is too high for positive action. The descriptor '*this width is too wide for a bus lane*' suggests that a narrower bus lane should be implemented. After initial trials with the experts, it was felt that the use of three descriptors was not sufficiently flexible. Accordingly a fourth was added to soften the boundary between the positive and the negative actions. The descriptor '*possible but really too narrow*' falls into this category.

A fifth descriptor to soften the boundary between the two more positive actions was considered, but its inclusion was rejected on the grounds that such a descriptor would be redundant. The reason for this may be explained by an example. If the bus lane were described with the descriptor '*possible but really too wide*', this would provide a problem only if road space were limited. In this case, the limited space would result in other parts of the road space being required to be too narrow (or at least '*possible but really too narrow*') which would induce the model to attempt to reduce the bus lane width in favour of the other elements of the road space. Thus the problem would be identified during the consideration of the widths of the other elements. It was felt, therefore, that four descriptor distributions was sufficient for the decision process.

The descriptors which provide a purely descriptive function are less critical than the action descriptors which lead directly to some action. The descriptive descriptors serve to redefine a variable in some linguistic form so that a decision can be taken on the basis of such qualitative data. Thus a bus flow might be described as '*high*' or '*low*'. In these cases, the number of descriptors that is necessary is related to the requirement to describe the variable in more or less detail. In many cases, three descriptors are sufficient - for example, '*high*', '*low*' and '*medium*'. It was felt after initial tests with the experts, however, that three descriptors was insufficient. This is because the descriptor '*medium*' is rather bland and it was felt that it would be better to split this descriptor to define the variable more clearly. Thus it is possible to use the same model (with four descriptors) for both types of descriptor.

Thus it is concluded that in the case of the design of high capacity bus systems, all decision processes which make use of the opinion model may be modelled using four descriptors. Having discussed the semantic form of the opinion model, it is now necessary to define the mathematical form of the opinion model.

5.3 THE OPINION MODEL

The opinion model has two parts: the function which describes the opinion and the process with which this function is used in order to represent the decision process. The mathematical function will be derived first.

5.3.1 The mathematical representation of opinion

The mathematical function used to represent opinion is constrained by the four requirements noted in Section 5.2.4.1 above. Thus in order to conform to these requirements, the function should be continuous, unimodal, and bounded at both the upper and lower limits. Equivalence between the opinion distributions of different variables can be achieved by scaling the distribution so that the mode is equal to one. It is postulated by Tyler (1987) that the β distribution might be used for the evaluation of fuzzy supports, and this is examined further here. The β distribution is a function that is continuous, unimodal and bounded at both the lower and upper limits. This distribution can also be scaled so that the mode is equal to 1. The β distribution also has three major advantages within this context:-

1 **it is flexible**: as the distribution is used to describe the 'shape' of the opinion distribution (rather than just its numerical values), a single function which can provide many such shapes is computationally convenient;

2 **it is easy to compute**: the form of the function requires only one dependent variable;

3 **it has a small number of parameters to evaluate**: only three parameters are required which are mathematically related as described in Equations (5.2), (5.3) and (5.4).

The β distribution is now described. It is given by:-

$$\beta(x) = K\, x^{\alpha} \,(1 - x)^{\gamma} \qquad \textbf{(5.1)}$$

where

$\beta(x)$	=	the value of β given the value of x
x	=	independent variable
K	=	constant scalar
α,γ	=	parameters to be calibrated

The β distribution is defined only for values of x between 0 and 1, so values of a variable must be scaled so that they fall in the range [0,1]. The first problem is to define the parameters α and γ and the constant K.

A particular characteristic of the β distribution is that there is a relationship between the value of x at the mode of the distribution (x_m) and the parameters α and γ:-

$$x_m = \frac{\alpha}{(\alpha + \gamma)} \tag{5.2}$$

This equation can be rearranged to obtain γ in terms of α and x_m:-

$$\gamma = \frac{\alpha(1 - x_m)}{x_m} \tag{5.3}$$

As the value of β at the mode, $\beta(x_m)$, is constrained to be equal to 1, K is easily determined by:-

$$K = \frac{1}{x_m^{\alpha}(1 - x_m)^{\gamma}} \tag{5.4}$$

Equations (5.3) and (5.4) give values for γ and K for known values of α and x_m. In this case, the β distribution is used to define an opinion distribution, and it is possible to determine the value of x_m explicitly. This can be done with knowledge of the value of the variable at which the opinion value of the particular descriptor is at a maximum. The value of α may, however, be determined in terms of x_m, any other value of x (x_p), and the value of β at any other point, x_p, ($\beta(x_p)$). This relationship is given by Equation (5.5) where Z is given by Equation (5.6):-

$$\alpha = \frac{-\ln(\beta(x_p))}{\ln\left(\frac{x_m}{x_p}\right)Z} \tag{5.5}$$

$$Z = \frac{\left((1 - x_m)^{[\,1/x_m\,]} \right) - 1}{\left((1 - x_p)^{[\,1/x_p\,]} \right) - 1} \tag{5.6}$$

Thus once the modal value of the variable is established, only two further values (x_p and $\beta(x_p)$) are required in order to establish the values of all the parameters in the β distribution. As suggested above, the modal value x_m can be established easily. In practice it is not too difficult to establish the values for x_p and $\beta(x_p)$, since there is often some point (for example the threshold point where the opinion distributions of two descriptors intersect) that is quite easily identifiable, thus obtaining both values at the same time. The issues concerning the determination of these values are discussed in more detail in Chapter 6. At this stage, it is only necessary to know that a β distribution can be determined fully with knowledge of these values.

The definition of the parameters within the β distribution is unique for particular values of x_m, x_p and $\beta(x_p)$, Thus each descriptor for a particular variable can be defined with its own β distribution. In Section 5.2.4.3 it was noted that four descriptors are necessary to describe a variable. A variable can therefore be fully described over its whole feasible range by four β distributions where each defines a different descriptor. Such a set of distributions is shown in Figure 5.4.

In Section 4.4.3, the use of fuzzy set theory and some of the techniques used for the propagation of decisions were discussed. In that discussion the problem of obtaining the initial fuzzy support values was raised. Using opinion as the measure of support for a descriptor, this section has given a method which obtains the opinion distribution for a descriptor, given knowledge of the opinion values of only two points. Thus a major problem of fuzzy set theory raised in Section 4.4.3.2 (namely, the arbitrary evaluation of initial fuzzy support values) is effectively resolved. The next section considers the use of these distributions in the decision process.

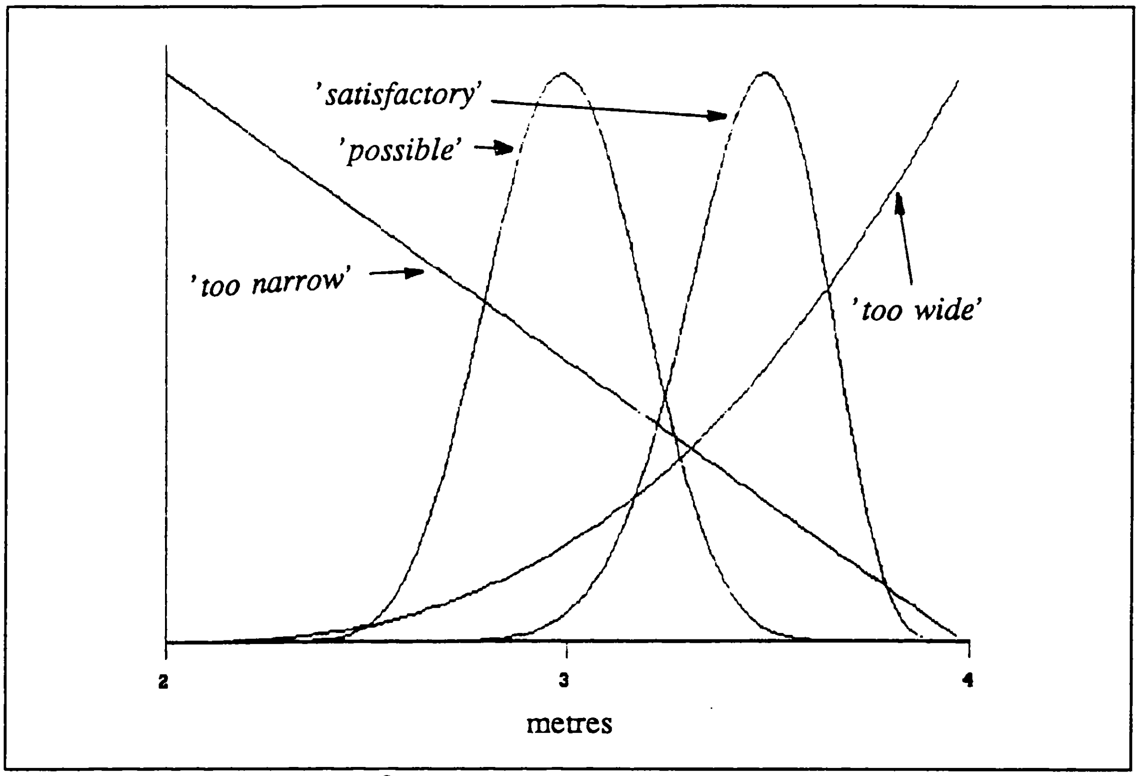


Figure 5.4 The set of four β distributions describing the opinions of four descriptors concerning the width of a bus lane.

5.3.2 The representation of opinion in the decision process

5.3.2.1 Introduction

This section discusses the mathematical representation of the use of opinion in the decision process. This involves the use of composition defined by Zadeh (1965) and adapted by Jowitt and Lumbers (1983) as discussed in Section 4.4.3.5. This section considers first the relationship between opinion values and the supports found in fuzzy sets. This is followed by a discussion on the formulation of a relation where this represents a decision matrix (as discussed Section 4.4.3.5). Next, the composition process is defined with respect to opinion values where these are considered to be the support values of a fuzzy set. Finally, examples are used to illustrate the difference between the decision process where exact values are known and that where the data are known approximately.

5.3.2.2 The base case

The opinion distribution defined in Section 5.3.1 provides an opinion value for each descriptor for each value of a variable. This may be denoted in such a way to suggest a link between the opinion value and a fuzzy support as defined in Section 4.4.3.2:-

$$\beta_j(\ x_v\ )_i \ \equiv\ \mu_{j,i}(\ x_v\ )_i \mid (\ x_v\ )_i \tag{5.7}$$

where

μ_j	=	fuzzy support for the descriptor j
$(x_v)_i$	=	the *ith* value of x for variable v
$\beta_j(x_v)_i$	=	opinion value of descriptor j at value $(x_v)_i$ of the variable v
i	=	index of values of x; $\{i : 1, 2, ..., n\}$

Thus the opinion values given by the β distribution are considered to be equivalent to a fuzzy support value for the representation of the decision process. Opinion values can therefore be represented in the form of a fuzzy set as defined in Section 4.4.3.2, Equation (4.2).

The β distribution defined in Section 5.3.1 is defined over the range of all real numbers between 0 and 1, and is continuous. Thus there is an infinite number of values of x_i to be evaluated. Since Δx ($\Delta x = [x_i - x_{i-1}]$) can therefore be infinitely small, this results in a correspondingly infinite number of opinion values for each descriptor. Thus the fuzzy set for each descriptor, and the corresponding relation matrices, are each infinitely large. This clearly presents a potential computational problem which is now considered.

5.3.2.2.1 Matrix columns

This problem can be resolved quite simply by considering the purpose of the exercise. This is to obtain a preferred decision based on the available data. It is reasonable to assume that, for a particular value of the variable, the decision output (or descriptor) with the highest opinion value is the preferred option. Although (as will be discussed in more detail below) this is qualified by the corresponding opinion values for the other descriptors, the chosen descriptor for a particular data value is that with the highest opinion value.

This can be seen in Figure 5.4. As the value of the variable increases, each distribution in turn has the highest opinion value of the set of four. The number of possible decision outputs is limited by the number of descriptors and, as discussed in Section 5.2.4.3, this is limited to four.

As there is a maximum of four possible outputs, the number of columns within the decision matrix is limited to four. It is therefore necessary to know only the evaluation of each descriptor within the ranges of the variable for which a particular decision output (or descriptor) has the maximum opinion value.

5.3.2.2.2 Matrix rows

Having reduced the number of columns in the matrix, the number of rows is now discussed. The same procedure applies in this case as with the columns. The entire range of the variable can be divided into sub-ranges which are defined by the change in the potential decision output. Thus, knowledge that the data value of interest lies within a sub-range of variable values for which a particular descriptor has the maximum opinion value carries the inference that the decision output is defined by the descriptor with the highest opinion value for that range.

As the number of descriptors is limited to four, the number of related sub-ranges of the data is also limited to four. Thus there are four rows in the decision matrix. The expertise concerning the relationship between the variable and the appropriate output (for example, the decision to implement a bus lane of a particular width) is contained in a decision matrix which is determined by the relationship between the β distributions which describe the four descriptors concerned in the decision. The decision process is therefore a question of comparing the opinions of the available data with the expertise.

5.3.2.2.3 Composition

Composition as defined by Zadeh (1965) and discussed in Section 4.4.3.5 is only used to process two fuzzy relations at a time. The adaptation of Jowitt and Lumbers (1983) allows composition between one fuzzy set and a relation. The result of this process is that a fuzzy support is obtained for the output given the support for the input and that for the relation between the input and the possible outputs. In the decision model presented here, there is one fuzzy set for each sub-range of the data, each consisting of the supports evaluated for each descriptor. Composition must therefore be performed on the relation between each sub-range of the data and each descriptor. The adaptation of the composition process in order to accommodate this is now described.

As described in Section 4.4.3.5, composition consists of two operations: the minimum of the supports for the input and those in the relation (decision matrix), and the maximum of these minima to determine the fuzzy supports for the outputs. Since a descriptor encompasses a range of data values, it is necessary to determine which of the set of fuzzy supports defined within the data range should be compared with the relation. The first operation of the composition process is therefore to take the minimum of the value of the opinion distribution (B) for each data value x_i within the descriptor range and the value of the relation at this data point:-

$$\mu_{(A,B)}(x_i) \mid (d,x_i) = \min\{\mu_A(d,v), \mu_B(v,x_i)\} \qquad \textbf{(5.8)}$$

where

$\mu_A(d,v)$ = the supports for the decision output d given the variable sub-range descriptor $v_{1 \dots 4}$ and decision descriptors (expertise) $d_{1 \dots 4}$

$\mu_B(v,x)$ = the supports for the variable sub-range descriptor v given the data range $x_{1 \dots n}$ and sub-range descriptors $v_{1 \dots 4}$

$\mu_{A,B}(d,x)$= the fuzzy set of supports for the decision output d (given the data range $x_{1 \dots n}$ and decision descriptors (expertise) $d_{1 \dots 4}$)

and the relations A and B are given by:-

$A(d_j v_i)$ = expertise relation between the fuzzy supports for the variable sub-range descriptor v_i and decision output d_j

$B(v_j x_i)$ = data relation between the fuzzy supports for the variable value x_i and the sub-range descriptor v_j

Having obtained the minima for all data ranges and all descriptors, the maximum support is obtained for each descriptor j:-

$$\mu_j \;=\; \max_{i,j}\{\, \mu_{i,j} \mid d_j,x_i \,\} \tag{5.9}$$

This results in a decision set, which is a fuzzy set of four ordered pairs, where each pair is the support for a particular descriptor:-

$$\Omega_{d\,,\,x} \;=\; \left[\, \mu_1 \mid d_1 \,,\, x_1 \,,\, \mu_2 \mid d_2 \,,\, x_2 \,,\, \mu_3 \mid d_3 \,,\, x_3 \,,\, \mu_4 \mid d_4 \,,\, x_4 \,\right] \tag{5.10}$$

The subsequent decision is then defined by the descriptor with the maximum support within this fuzzy set:-

$$\mu \mid \Omega \;=\; \max_{\,1 \le i \le 4} \left[\, \mu_i \mid (\, d_i \,,\, x_i \,) \,\right] \tag{5.11}$$

The whole process is given by an adaptation of Equation (4.8):-

$$\mu_{A,B}\ (d,x) \;=\; Max \left\{ Max_d\ Min\ \{\, \mu_A\ (d,v),\ \mu_B\ (v,x)\, \} \right\} \tag{5.12}$$

It can be seen that the relation A represents the expertise obtained from the design engineer with respect to the requirements for a particular decision. The relation B represents the qualitative interpretation of the data range in terms of linguistic descriptors. The fuzzy set $\mu_B(v,x)$ represents the supports for a particular descriptor for the value x. The fuzzy set $\mu_A(d,v)$ represents the supports of a particular decision given the support for a variable descriptor. The fuzzy set $\mu_{A,B}(d,x)$ therefore gives the resulting supports for each decision descriptor, given the preferred description of the available data values of x. The option to be chosen is that with the highest opinion value in this set. The other values in the set suggest the degree to which this choice is unique: if the second highest support is numerically very close to the maximum, for example, there is an apparent ambivalence about the chosen decision. Account should be taken of this in future use of this decision.

5.3.2.3 An example

It is helpful to describe this process linguistically, and with the aid of graphs of some relevant distributions, for a simple example. Figure 5.5 shows a distribution of opinion concerning a range of the ratio of peak commercial speed to that for the off-peak. First, the available data is

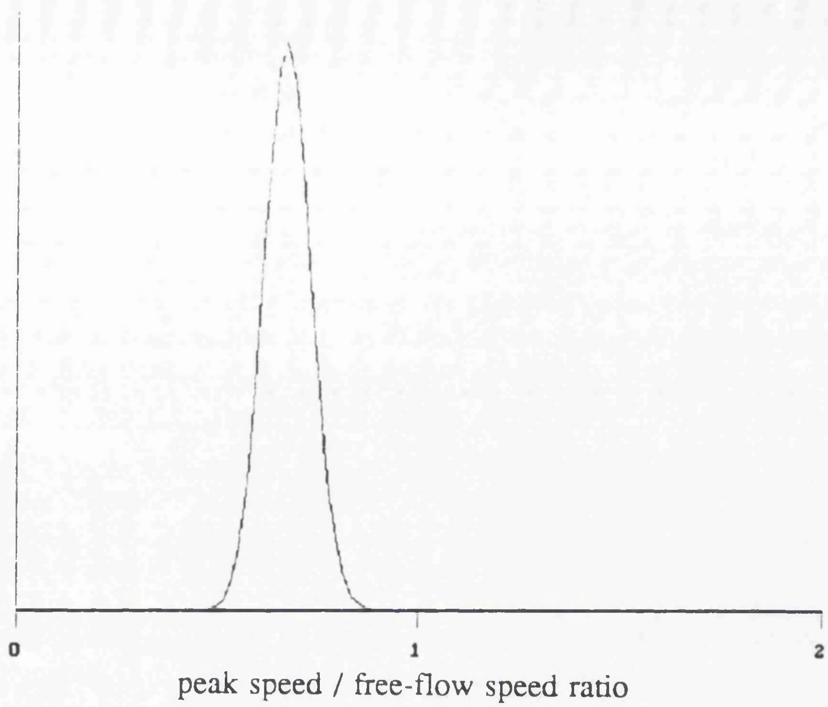


Figure 5.5 Data range for peak speed / free-flow speed ratio

categorized in terms of the decision descriptors which apply to various parts of the total range. Figure 5.6 shows the opinion distribution from Figure 5.5 superimposed upon a set of distributions obtained from an engineer which expresses the possible decisions resulting from various parts of the feasible data range. For each sub-range, the intersection between the opinion distribution for the data and that for each of the decision descriptors (defined by Equation (5.8)) defines the set of support values to be used in the next stage of the composition process. The resulting matrix is shown in Table 5.1.

Table 5.1 shows the supports for each of the decision descriptors given the opinion of the data values when these have been defined by data ranges resulting from the first stage of the composition process. The second stage of the process obtains the fuzzy set resulting from taking the maximum value for each descriptor. The maximum value within this set is then taken as the support for the preferred option. It can be seen from this simple example that the descriptor '*the bus lane is not needed*' has the highest opinion value, and that the decision would therefore be taken not to proceed with the design of a bus lane. Thus from an initial estimate of the approximate ratio between peak and off-peak commercial speeds, it is possible to derive a decision about the necessity for a bus lane. The method for calculating the opinion distribution of this ratio from opinion distributions of approximately-known peak and off-peak speeds is discussed in

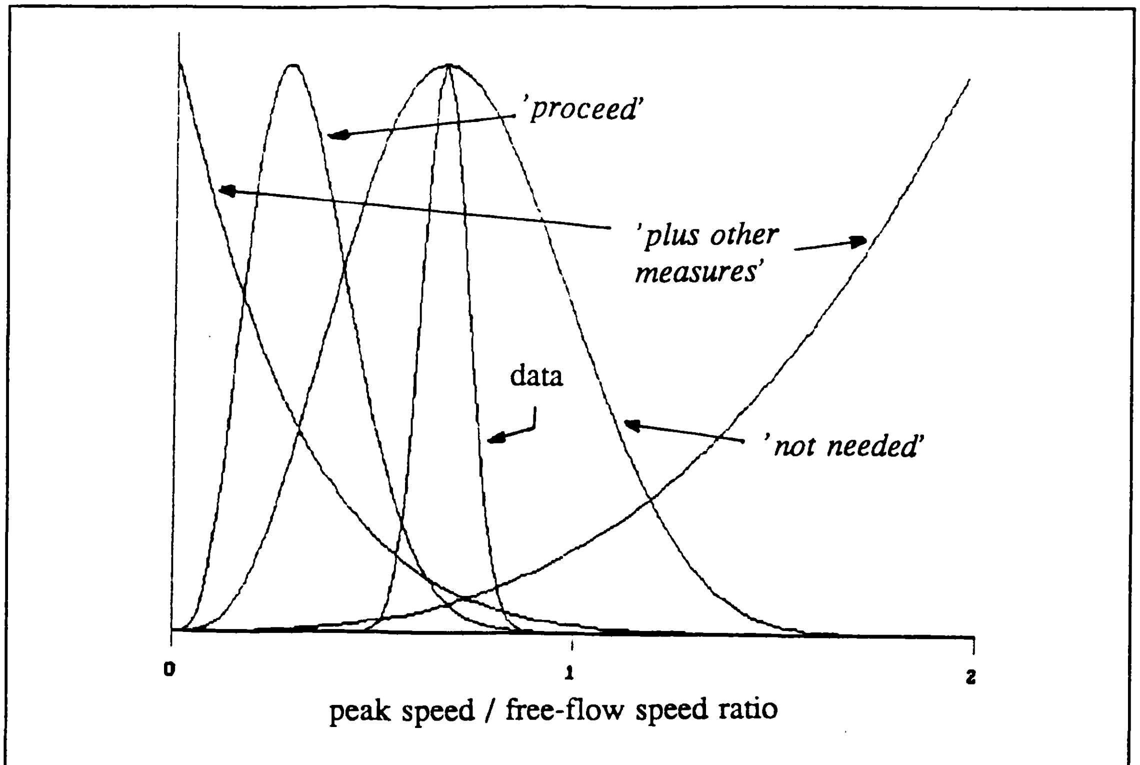


Figure 5.6 The data distribution from Figure 5.5 is superimposed on a set of distributions of opinion concerning various descriptors of particular ranges of the data.

Table 5.1 Decision matrix showing the support values resulting from the first stage of the composition process illustrated in Figure 5.6.

	plus other measures	*study should proceed*	*not needed*	*plus other measures*
1	0.00	0.00	0.00	0.00
2	0.00	0.00	0.00	0.00
3	**0.13**	**0.18**	**1.00**	**0.08**
4	0.00	0.00	0.00	0.00
max	0.13	0.18	1.00	0.08

Section 5.4. The opinion values are used as a measure to evaluate the opinion of the results of each part of the process. As noted above, the relationship between the various values within this fuzzy set provides an estimate of the reliability of this choice.

Thus the base case is described for the propagation of opinion values through the decision process. Two other cases should also be examined.

5.3.2.4 Other cases: alternative options

The first case examined is where the final decision is bifurcated because the final fuzzy set of opinions of the decision outputs contains elements from different sub-ranges of the data. This occurs when the distribution of opinion of the data intersects with the distributions of different descriptors in different sub-ranges of data values. This is most likely to occur if the data are known only very approximately (thus the variance is very large), but it can happen if the data range crosses a decision threshold even where the variance is small. An example of this is shown in Figure 5.7, with the corresponding decision matrix in Table 5.2.

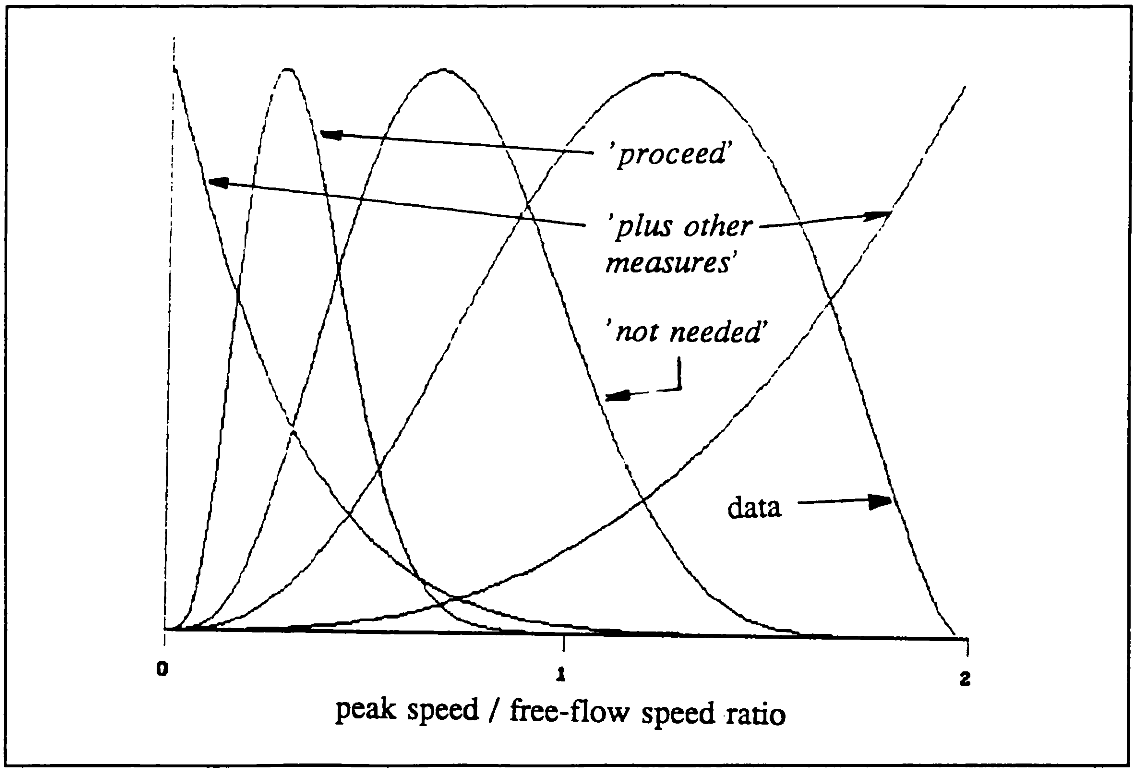


Figure 5.7 Superimposition of an opinion distribution and expertise where the decision option is not clear

There are two possibilities arising from this result. The first is that the data could be re-assessed in order to reduce the variance (and thus to reduce the possibility of this bifurcation). The second is to decide whether the appropriate sub-range of the data is above or below the threshold, and to act accordingly. Where data quality is poor it is extremely useful to have this check on the reliability of a decision based on poor quality data. In some cases, even very poor quality data may result in a unique decision set, while in others data which are known with a reasonable degree of precision may produce a bifurcation.

Table 5.2 Decision matrix obtained from Figure 5.7. Note that the resulting fuzzy set has values from different data ranges.

	plus other measures	*study should proceed*	*not needed*	*plus other measures*
1	0.02	0.02	0.02	0.00
2	0.17	0.17	0.17	0.00
3	**0.21**	**0.27**	**0.76**	0.24
4	0.00	0.00	0.23	**0.60**
max	0.21	0.27	0.76	0.60

5.3.2.5 Other cases: where data is known precisely

The second case to consider is where the data are known precisely. For the reasons stated in Section 5.2 above, this is unusual when the data refer to observations of the real world. Where the data on which a decision is based are generated as part of the decision process, however, they are more likely to be known precisely. This section discusses this possibility with a simple example.

The example considered is the allocation of road space to a combination of traffic lanes, bus lanes and footways. The width of each element of the road space is required to be acceptable (as defined by the design engineer). As one element is widened, at least one of the others must be reduced in width. An iterative process is therefore necessary in order to find a combination of satisfactory widths within the constraint of the total available space. Each iteration tests the widths for their acceptability and if necessary makes an alteration. The widths being tested are precise values, generated by the iteration process itself.

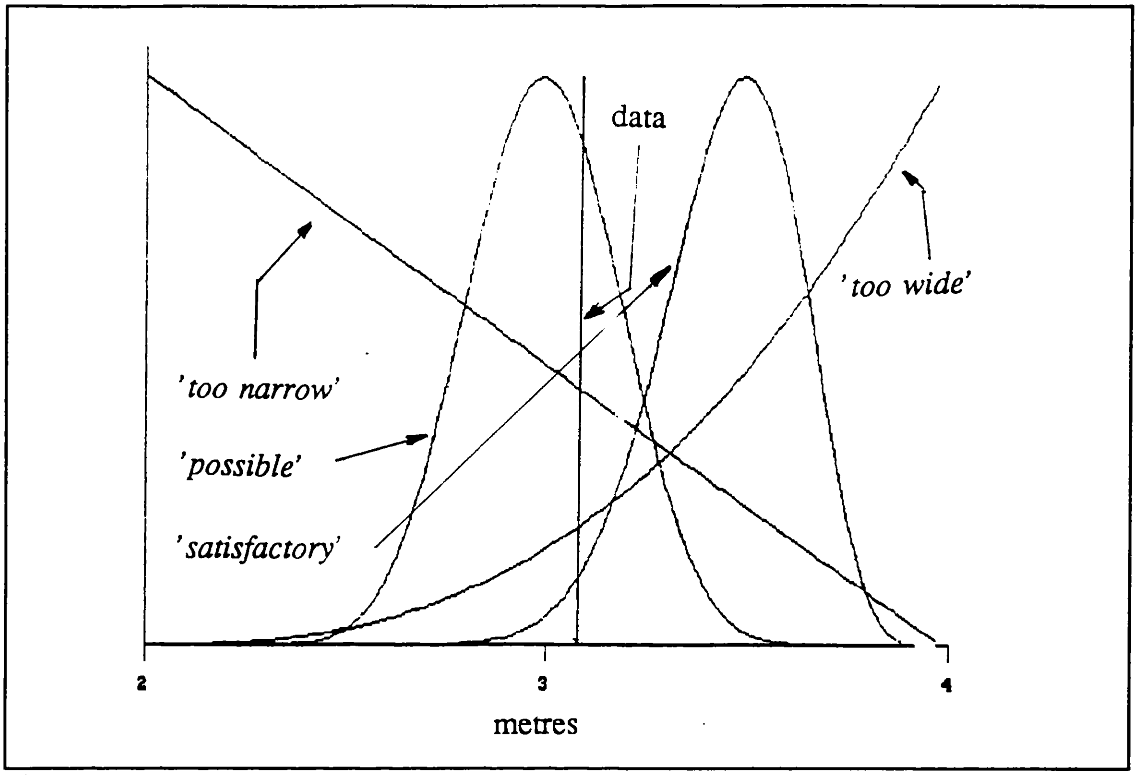


Figure 5.8 A precise data value is compared with a set of opinion distributions.

The composition process for a precisely-known value is a special example of the base case where the variance of the opinion distribution is infinitely small. Figure 5.8 shows an example in which it can be seen that the opinion distribution of the data appears as a straight line. The effect of this on the composition process is that while the number of columns in the decision matrix is still governed by the number of decision descriptors, the number of rows is reduced to one. The second stage of the composition process becomes trivial in this case since there is only one support value for each decision descriptor. This value is therefore the maximum. The final fuzzy set supporting the decision output is processed in the same way as for the base case.

5.4 CALCULATIONS WITH OPINION DISTRIBUTIONS

5.4.1 Introduction

The example used in this chapter to illustrate the use of opinion values in the base case (Section 5.3.2.3) involves the evaluation of the opinion values for the ratio of two values. The two values concerned (the commercial speed obtained during the peak and off-peak periods respectively) may not be known precisely. In order to produce the ratio with its associated opinion distribution, it is necessary to derive a method for calculating the opinion distribution which results from a calculation involving two (or more) approximate variables. In this example, the calculation to be performed is a division but the considerations are the same for any calculation.

If the data were known precisely, this calculation would be simple. If the peak speed were 10 kilometres per hour and the off-peak speed were 20 kilometres per hour, the resulting ratio would be 0.5. Two questions are addressed in this section. The first concerns the calculation of the ratio when the input values are approximate, and the second is the calculation of the opinion distribution of this ratio.

5.4.2 Calculations with approximate values

5.4.2.1 General issues

When data are known imprecisely, they are expressed in terms of a range; thus the peak commercial speed could be described as *'about 10 kilometres per hour'*. The concept of imprecision (*'about'*) may be defined in this case to be *'not less than 5 and not more than 15 kilometres per hour'*. A similar concept could be applied to off-peak speed; for example, *'not less than 17 and not more than 30 kilometres per hour'*. To obtain the ratio between these two speeds, it is necessary to consider every possible value of each speed range.

Each speed range is effectively a vector and the calculation is therefore the division of two vectors. Each value in the peak speed range should be divided by every value in the off-peak speed range in order to obtain every possible value of the ratio. The result of this exercise is a matrix, with the result of each calculation in its appropriate cell.

As the number of values in each vector is infinite, this matrix is potentially very large, and the calculation process may also be very intensive - even for a simple calculation such as that in this example. It would therefore be interesting to consider whether it is possible to generate an output sub-matrix which consists of a selection of results and which is sufficient to describe the whole matrix. The sub-matrix would be the result of calculations of selected values from each vector. This would mean that the number of calculations could be reduced.

The major difficulty is to choose the appropriate values in each vector. The choice of values is important because it is necessary to be satisfied that the resulting matrix is a representative subset of the whole matrix and not the result of calculations between unrepresentative values. This is achieved with the help of the opinion distribution.

5.4.2.2 Minimum and maximum values

Consider first the two values in each vector that are defined: the minimum and maximum. If these were the only values in each vector, the result of the division would be a four-by-four matrix where the rows correspond to the off-peak speeds and the columns refer to the peak speeds. The top left cell represents the division of the two minima and the bottom right cell represents the division of the two maxima. The other cells give the results of the division of the other two values. Thus the division of the two vectors described in Section 5.4.2.1 has the following result:-

$$\begin{pmatrix} 0.29 & 0.88 \\ 0.2 & 0.6 \end{pmatrix} \tag{5.13}$$

It is clear that there is a minimum value resulting from the division of the minimum value of one vector by the maximum value of the other and a maximum value for the matrix resulting from the division of the maximum value of one vector by the minimum value of the other. The minimum and maximum values are 0.2 and 0.88 respectively. It is important to realise that there can be no value within the whole matrix that is less than the minimum nor one that is greater than the maximum values in this sub-matrix. These values are therefore used to define the minimum and maximum of the range of the resulting ratio.

If the calculation is multiplicative, the minimum and maximum values of the result are simply the products of the two minima and maxima respectively. Addition, like multiplication obtains a minimum result from the addition of the two minima (and the maximum from the addition of the two maxima). Subtraction, like the division process described above, obtains the minimum result from the subtraction of the maximum of one vector from the minimum of the other. With subtraction and division, of course, the determination of which vector is to be subtracted from (or divided by) which is defined by the relation in question.

The next task is to define the opinion distribution which describes the opinion of this result.

5.4.2.3 The mode

The derivation of the minimum and maximum values in the way described above implies that there is a semantic equivalence between the minimum and maximum values in the two vectors. The equivalence is that there is no other possible value outside the range - these values constitute the limits. Therefore the minimum and maximum results appearing in the sub-matrix are the absolute minimum and maximum values it is possible to obtain from the calculation.

As mentioned in Section 5.3.1, in order to define an opinion distribution for the range of the ratio, it is necessary to know this range, the value of the mode, some other value of the variable and the opinion value at this point. Whereas during the collection of expertise (to be discussed in Chapter 6) the mode is relatively easy to ascertain since it is the preferred value, in the evaluation of a calculation (such as the division in this case) the mode must be inferred logically.

This is an extension of the principle underlying the use of the minimum and maximum values to determine the range of the solution. The mode is another value in the data range which can have a semantic equivalence with a similar value in another range. The mode represents the value of the variable at which the opinion is at its highest. Therefore it is not unreasonable to assume that a calculation involving the division of the two modal values also represents the division of two semantically equivalent values.

In this example, it is not unreasonable to consider that the speed value which generates the mode of one opinion distribution is equivalent semantically to the speed value in the other vector which generates the mode of the other distribution. This is because the mode of the opinion distribution identifies the preferred value in each case. Thus it is contended that the preferred value of the peak speed vector is semantically equivalent (although not necessarily quantitatively equal) to the preferred value of the off-peak speed vector. The calculation can therefore be performed using the modal values of each opinion distribution.

5.4.2.4 The other point

As shown in Section 5.3.1, it is also necessary to calculate another point in order to derive an opinion distribution for the results. Each speed value in a vector has an associated opinion value, and this can be used in order to identify certain characteristics of different speed values. In principle, any pair of values (one from each vector) would suffice. In practice, however, since both the value of the result (of the calculation) and its associated opinion value have to be calculated it is useful to establish a logical consistency in the matter of choosing appropriate values to use.

In the calculation of the modal value, this is not a problem because the value of the opinion distribution at the mode is by definition equal to 1 (and by definition is the same for each vector). In order to choose another pair of values which has semantic equivalence between the two vectors, the opinion values of the two variable values should be equal. This opinion value can be chosen arbitrarily, and for convenience the value of 0.5 has been used in this research. The reason for this choice is that 0.5 represents the idea of a null value of opinion - neither low nor high - and therefore it is a value with at least some semantic content that can be described (albeit in general terms).

The next task is therefore to calculate the value of the variable at which the opinion value is 0.5. The immediate difficulty is that since the β distribution is unimodal, there are two points in the variable range for which the opinion value is 0.5. Nevertheless, these values can be calculated and a rule devised to choose one in particular. In this case the rule is to choose the lower resulting variable value in each vector.

The calculation of the variable value is not trivial, and is derived in Allsop (1990). The results are obtained through the recurrence relation:-

$$x_{n+1} = x_n - \left(\frac{(\alpha \ln(x_n + \gamma \ln(1 - x_n) + C)\, x_n (1 - x_n)}{\alpha - (\alpha + \gamma)\, x_n} \right) \qquad \textbf{(5.14)}$$

In order to obtain the left-hand solution (that is the lower resulting variable in the vector), Allsop suggests the initial value of x could be obtained by:-

$$x_0 = \frac{\alpha}{2(\alpha + \gamma)} \qquad \textbf{(5.15)}$$

In this way a 'null' value of the variable can be obtained within each vector which can be used in the subsequent calculation. The result of this calculation produces the remaining value required in order to establish the opinion distribution of the resulting ratio.

This method permits the evaluation of the results of any calculation where there is an opinion distribution defined for the input variables. The result of using this method is a range for the solution and an opinion distribution for this range. These can be used for further calculations or can be evaluated by the appropriate expertise using the methods described above.

5.5 CONCLUSIONS

The relative importance of different values of a variable is a source of complexity within the decision process. This importance is caused by the mixture of a continuous input and a discrete output, and suggests that values of the input which are near to a threshold between two outputs convey more importance than other values and should accordingly be treated more cautiously.

If these decision thresholds are known, it is possible to use imprecise data as an input. This is because the effects of different values can be estimated by their relationship to these thresholds.

The opinion model presented in this chapter provides a way of identifying and evaluating these thresholds. Thresholds are determined by the intersection between the opinion distributions of different descriptors. This definition also provides that the values of opinion for each descriptor at the threshold are considered equal, but independent of their magnitude. Thus a decision threshold could exist where the associated opinion values are either high or low (or indeed any value).

The model measures a distribution of opinion of a range of data values against a set of distributions of opinion of descriptors which relate to particular parts of the data range. This represents the qualitative interpretation of data used by an engineer when faced with data which are not known precisely. Section 5.2 shows that decisions are often taken on the basis of thresholds around which the result of a decision alters depending on the particular value of data under consideration.

A reasonable function to describe opinion, and which fits all the constraints defined in Section 5.2, is the β distribution. A method is presented to obtain a decision based on the available data and the expertise, together with a measure of the opinion of this decision output. This is achieved through the use of composition, slightly adapted from the definition of Zadeh as presented in Chapter 4.

Section 5.3 shows that these thresholds can be established through the use of distributions of opinion of the various decision descriptors, since a threshold is defined by the intersection between two such distributions.

The opinion model defined in Section 5.3 has two elements: the definition of opinion in the form of a mathematical distribution, and the representation of the use of opinion in the decision process.

Section 5.4 presents a method for calculating with imprecise data. This uses particular values from the opinion distributions of the relevant variables in order to ascertain the associated data values. The calculation is performed using these values and a new opinion distribution is obtained which describes the opinion of the resulting data range.

A major problem with the use of fuzzy logic to represent imprecision, identified in Chapter 4, is the determination of the initial fuzzy supports for a variable. This can be resolved using the opinion model which defines such values in a systematic, robust and consistent way.

This chapter has dwelt on the derivation of the opinion model, and has left the discussion of the methods used for the derivation of particular distributions to be considered in the next Chapter. Chapter 6 discusses the acquisition of the knowledge for the design of high capacity bus systems, including the opinion distributions necessary to define the expertise for the decision processes.

CHAPTER 6
KNOWLEDGE ACQUISITION

6.1 INTRODUCTION

Chapter 5 has concluded that it is possible to represent opinion using a mathematical model. The use of this model depends on the ability to recognize the thresholds which exist between alternative outputs. The determination of these thresholds is one result of the knowledge acquisition process discussed in this chapter.

The knowledge used in any decision process originates from many sources, and the example of the design of high capacity bus systems is no exception. Chapter 3 has discussed some aspects of the existing knowledge concerning such systems. This knowledge has been obtained from the literature and from discussions with design engineers who have identified a number of variables and parameters of interest. The use of the opinion model presented in Chapter 5 relies on the availability of expert opinion about a number of these variables and parameters. As is shown in this chapter, the opinion model is also necessary in order to acquire the opinions from the experts. This chapter considers the process of obtaining this knowledge and opinion. The first knowledge necessary is the structure and sequence of decisions which make up the method used to design a high capacity bus system. It is then necessary to obtain the factual knowledge related to the requirements of a design for a bus system. Finally the interpretative models which form the basis of the opinion distributions used in the opinion model are established.

This chapter has five further sections. The next section discusses some general issues surrounding the knowledge acquisition process. Section 6.3 describes the acquisition of knowledge about the structure of the design process for high capacity bus systems. Section 6.4 considers the acquisition of the opinion distributions for use with the opinion models. Section 6.5 discusses the calibration of the knowledge base. Section 6.6 draws some conclusions about this process. Appendix C contains a list of the experts consulted during this research, their areas of interest, the selection criteria, and a basic list of the variables and parameters discussed with the experts and the way in which these are incorporated in the design model.

6.2 THE KNOWLEDGE ACQUISITION PROCESS

This section is divided into two parts. First, some general issues surrounding the process of knowledge acquisition are considered, and this is followed by a discussion about experts, and their role in the knowledge acquisition process.

6.2.1 General issues

It has been stated that the knowledge acquisition process is often the single most difficult problem in the construction of a knowledge-based system (Feigenbaum 1977). It appears that it is difficult to obtain the knowledge of an expert in a form that is appropriate for use in a knowledge-based system. Dreyfus and Dreyfus (1986) argue that this is because the form in which the knowledge is used by the expert is not necessarily convenient for encapsulation in a computer. As discussed in Section 4.2.2.1.3, it is contended by Dreyfus and Dreyfus that this is because experts use thousands of examples in order to determine the progress of the sequence of decisions.

6.2.1.1 Knowledge engineers

The process of knowledge acquisition is controlled by a 'knowledge engineer'. This is usually someone who is a specialist in the process of the encapsulation of knowledge and encoding it in a suitable form for use in a knowledge-based system. This specialist is normally independent of the process under consideration, and is determined to have no previous knowledge of the domain. The reason for this independence is that it is supposed to reduce any bias in the knowledge encapsulation which may exist through previous knowledge of the domain on the part of the knowledge engineer. In these circumstances, the expert is required to communicate and explain his or her knowledge to the knowledge engineer who is, in effect, the equivalent of a novice in the subject area. It is therefore important that the expert can indeed communicate at this level. This is not always the case: some authors (for example, Lamberti and Newsome (1990); McKeithen *et al* (1981)) suggest that the thinking processes of experts and novices are different. Experts are perceived to use abstract, conceptually-based representation, while novices represent their knowledge in a more concrete form. A simple example of this difference is given by McKeithen *et al* (1981), where computer programmers were asked to cluster the words of a programming language. Expert programmers tended to cluster the words in conceptual groups, whereas novices tended to cluster them alphabetically, indicating a different conceptualization of the problem according to the level of expertise. Thus it is not always the case that an expert (who is used to dealing with high-level concepts) finds it easy to consider the lower level concepts which may be more intelligible to a novice.

A knowledge engineer therefore has to obtain the 'raw' knowledge from the expert and to convert this knowledge into a form which is appropriate for the knowledge-based system in question. The knowledge engineer should have a wide knowledge of the various knowledge representation paradigms discussed in Section 4.2.2, and should be able to select the appropriate paradigm for the type of knowledge in question. The problems which may occur when inappropriate paradigms are chosen have been discussed in Chapter 4.

6.2.1.2 Knowledge acquirers

It may be preferable to allow the expert to operate at his or her normal conceptual level, but to use a knowledge engineer who has some previous knowledge of the domain. In this thesis, the term 'knowledge acquirer' is used to refer to a person with previous knowledge of the domain who acquires knowledge from experts. The term 'knowledge engineer' is therefore reserved for the domain-independent knowledge engineer (discussed in Section 6.2.1.1), responsible for encoding the knowledge in a suitable form for the knowledge-based system.

The knowledge acquirer is not necessarily a specialist in the process of knowledge encapsulation, but is capable of using his or her own knowledge of the domain to assist clarification of the expert's statements. This is not to say that the knowledge acquirer should have a similar level of expertise as the expert, rather that the level of knowledge should be sufficient to understand the expert without repeated simplifications. These are often irritating to an expert, and may tend to interrupt the flow of knowledge. The ability to discuss the details relevant to particular points of knowledge is also important as this may help to clarify any apparent inconsistencies.

Another reason for the advantage of previous knowledge of the domain that has been observed in this research, is that simplifications in the expertise (which may be necessary in order to help a knowledge engineer to understand complex processes) can be avoided - or at least explained in depth to a knowledge acquirer who has a basic grasp of the issues involved.

During the course of this research it has been found to be easier and more effective to obtain knowledge from an expert when the knowledge acquirer has a level of expertise such that the expert feels confident that there is a mutual level of understanding. The approach taken has been to attempt to model the expert's reasoning process rather than to force the expert to describe the expertise in unfamiliar terms.

6.2.1.3 Automated knowledge acquisition

Shapiro (1987) points out that an expert is generally involved in the interpretation of data rather than in the explanation of how these interpretations come about. Shapiro suggests that this is a major handicap in the knowledge acquisition process. When this is combined with a lack of knowledge of the inference system to be used in the proposed knowledge-based system, the knowledge acquisition process can become very slow. One possible solution to this problem is to model the inference system of the expert, thus the process becomes a matter of attempting to understand the thinking process of the expert rather than that of expecting the expert to understand the inference process of a knowledge-based system which is as yet unbuilt.

The knowledge acquisition process used in the construction of a particular knowledge-based system depends, at least to a certain extent, on the type of knowledge concerned. The knowledge acquisition process described in Allen (1986), for example, consists of a series of interviews, conducted by the developer of the knowledge-based system with a number of experts. These interviews attempt to obtain sufficient information from the experts so that rules may be derived. Other systems attempt to automate the process through the use of machine-learning techniques (Gaines and Boose, 1990).

Shalin *et al* 1990 suggest that the conventional knowledge acquisition procedure depends on the gradual modification of the knowledge-based system as performance failures occur. This is described as a process which attempts to acquire knowledge from experts but which is restricted by the reliance on ill-defined interactions between the experts and a knowledge engineer which continue over a long period of time. One major difficulty identified by Shalin *et al* is that it is difficult to know when the knowledge base is complete. Shalin *et al* define four ways to facilitate this process with machine-learning techniques:-

1 automated techniques may be used to generate an initial knowledge base which can then be refined by an expert;

2 they may be used to refine an existing knowledge base;

3 machine-learning techniques could also be used to adapt existing knowledge-based systems (to include an element of style for example);

4 they could be used to construct complete knowledge bases.

The fact that machine-learning techniques are used at all indicates the level of difficulty found in acquiring knowledge from experts.

6.2.1.4 The nature of expertise

As mentioned above in Sections 6.2.1.1 and 6.2.1.2, the acquisition process depends at least to some extent upon the type of knowledge involved. As has been indicated in Chapters 4 and 5, the expertise in the case of high capacity bus system design appears to be made up of three parts.

First there is the factual knowledge which consists of statements which are either true or false, and which are combined to form heuristics or rules in order to infer conclusions.

The second element of expertise consists of the use of models (either existing, or where necessary constructed) in order to establish particular parameter values, or to calculate results.

The last is the use of opinion and judgement in the process of the interpretation of both conclusions and of the available data.

The first type of expertise may be represented using existing methods discussed in Section 4.2.2, such as rules, frames or blackboards. The second is, in many cases, represented by various models (both within computer software and elsewhere) which are available to the expert. In other cases, the expert may construct a model where one is thought to be helpful. The last case, however, is not currently represented in the literature. It is therefore necessary to devise and use a model to represent the encapsulation and use of opinion and judgement. A model to perform these functions has been presented in Chapter 5. A considerable part of the knowledge acquisition undertaken during this research has involved the acquisition of the opinion distributions derived in Chapter 5 for this purpose. Before discussing the acquisition of expertise in this form, however, some of the issues concerning experts are now considered.

6.2.2 Experts

This section considers the issues surrounding the experts. A major consideration is whether to acquire knowledge from one or more experts. This is discussed in the next two sections. This is followed by the consideration of the image of the expert and the relevance of this to a potential knowledge-based system. Finally, the interactions between experts and their clients are compared with the possibilities offered by knowledge-based systems.

6.2.2.1 Multiple experts

Several knowledge-based systems (including MYCIN and PROSPECTOR) have used a number of experts to construct the knowledge base. Two reasons are given for this (see for example Buchanan and Shortliffe 1985). The first is that the knowledge-based system could contain more expertise about the domain, and that the interactions between the experts can help to clarify apparent inconsistencies in particular pieces of expertise. The second reason is that the use of multiple experts helps to offset any weakness or bias in the expertise of one expert. This has been questioned during this research, however, and a more refined definition of the term 'multiple experts' appears to be in order.

The use of multiple experts is similar to consulting an expert committee (where this is a collection of experts gathered together to consider a particular problem). For a committee to be successful, however, one person is required to coordinate the various activities of the other committee members. If the committee in question is a team of experts - each with different specialities - different areas of expertise would be required to deal with different parts of the problem. Decisions are needed to choose the best way to solve the problem, and the committee member responsible for such choices may change as different problem areas are tackled.

When this decision process is represented in a knowledge-based system, the selection of these 'superexperts' is critical to the outcome of the decision process. Weakness in this area could increase the likelihood of bad decisions. The problem with this method is that it is difficult to identify the superexpert, and the resulting danger is that the resulting decisions may be based on averages and compromises, taken over all the inputs from all the contributing experts. While an average or compromise may represent the 'common ground' between the experts, it is unlikely to be sufficient for the more difficult cases, or where the differences are caused by personal preferences on the part of particular experts.

The term 'multiple experts' suggests a number of individual experts who may be interviewed individually, or who may be brought together in order to elicit their expertise on some topic. Allen (1986), for example, uses the latter technique for knowledge acquisition. The difficulty arises when the experts disagree on fundamental aspects of the problem or its solution. Allen illustrates a problem where a number of purchasers within an organization were asked what price was recorded on the stock card for each item. Each expert had a different answer, revealing a major problem for the knowledge engineer who had to decide how to interpret this expertise.

Two alternatives exist in this situation: first, it might be possible to identify genuine reasons for these differences. This suggests that it is possible to obtain a coherent logical expression which differentiates between the contributions of individual experts. In this case it is possible to construct a rule base to identify these differences and thus to incorporate all the expertise.

The second alternative is that the knowledge engineer has stumbled inadvertently upon a process from which it is impossible to generalize a rule base from all the experts. One possible solution in this case is that it might be possible to impose a decision on the process (for example the type of price to be recorded could be determined by some higher authority within the organization). A second possibility is that this stage of the process could be left outside the domain of the knowledge-based system. The degree to which this is possible is dependent upon the purpose of the system: if it is intended for training purposes, for example, it is likely that such matters should be addressed explicitly.

Explicit handling of the problem may require a redefinition of the problem domain in order to exclude the difficulty. Alternatively, the use of a set of opinion distributions (as derived in Chapter 5), which address explicitly the differences between individual experts may serve to represent such differences. The methodology required for this is discussed below in Section 6.3.

In order to illustrate this problem, the position of a bus stop in the cross-section of the road is considered. As discussed in Chapter 3, a bus stop could be positioned in either the centre of the road or at the kerbside. While there are fundamental principles (discussed in Chapter 3) which indicate the circumstances under which one of these positions may be chosen, it is still possible for one expert to choose the median position, and another, basing the decision on the same data, to choose the kerbside. It is difficult to imagine how a compromise can be reached, or where is the average position.

The second reason that this use of multiple experts should be questioned is that some expertise seems to depend on the personal opinions and judgement of individual experts derived from their own experience (for example the preference for a median or kerbside position for a bus stop). It is therefore preferable to attempt to model the expertise with all the richness that such experience offers, with no loss suffered through the compromise inevitable when it is subjected to a form of committee system. In order to investigate this richness, it is necessary to approach experts individually. The question of the use of single experts is therefore now considered.

6.2.2.2 Single experts

It is important to note that a 'single expert' is not necessarily one individual: with a problem as complex as the design of a high capacity bus system it is unusual for the entire process to be carried out by one person. It is more usual in such cases, that the design is produced by a team of engineers. In such cases, individual members of the team perform particular tasks in which they are more expert than the other team members. Thus the team as a whole might be considered to be a 'single' expert. Knowledge acquisition therefore may take place with different members of the same team, according to their particular skills and depending on the particular aspect of the design process under consideration.

Knowledge obtained from single experts (or, as discussed above, teams of experts) can be stored in knowledge bases which can each be accessed as desired by the user. A by-product of this method is that the experts appear to like to feel that their expertise has not been subsumed into a knowledge base comprising the knowledge of many experts.

Accordingly, the methodology used in this research is to retain the individuality of the expert by maintaining separate expertise bases for each expert. However, it should be noted that further research may produce a methodology that would make the use of a knowledge base constructed using several experts, and that this should be actively pursued.

Whether using single or multiple experts, it is important to know (before any knowledge acquisition process commences) what role the proposed system is expected to play. This determines the scope of the knowledge that should be acquired and can influence the choice of experts to be consulted. This is now considered in two parts. First, the type of expert to be used for knowledge acquisition purposes is considered in terms of the various roles that a knowledge-based system can play. Second, the interaction between an expert and his or her client is considered. These two parts constitute the image of the expert.

6.2.2.3 The image of the expert

It is important to consider the image of the expert with reference to the proposed knowledge-based system because this helps to direct the knowledge acquisition process. If the knowledge acquirer has no clear idea of the relationship between the expert and his or her client, it is more difficult to extract the knowledge which is appropriate for the particular purpose of the knowledge-based system. It should also be clear to the expert during the knowledge acquisition process that the proposed system is intended to provide a certain level of decision. It should be considered, for example, whether the proposed system should provide general advice, or whether it should provide a detailed engineer's drawing of the proposed infrastructure. The decisions necessary for these two levels of system are different, and consequently each has a different requirement for data.

Four principal images for an expert are considered briefly here. First, the image could be that of a teacher passing on knowledge to others who have some defined level of expertise (including no previous knowledge of the problem domain). In this case it is necessary to obtain from the expert explanations of the reasoning underlying the decisions taken during the course of the design process. Knowledge acquisition from an expert of this type is much easier if the expert is used to communicating his or her ideas at a simplified level.

The second image is that of a decision taker. The expertise in this case is related to the bases upon which decisions are taken. This type of expert deals with the factors which are necessary to enable the system to be implemented at a particular site.

The third image is that of a designer. This type of expert is concerned with the details of how the final system will appear: the lane widths, bus stop lengths, bus stop locations and operating systems.

The fourth image is that of someone who conceptualizes. This type of expert conceives the type of bus system, and is not too concerned with current practice. What is important to this expert is solving the problem at hand, even if it is necessary to design a new type of system in order to achieve this.

The ways in which the image of the expert affects the development of a knowledge-based system may be discussed in two parts. The first element is the interaction between expert and client. Having established this, the possible relationship between a knowledge-based system and its user can then be discussed.

6.2.2.4 Interaction between the expert and the client

The knowledge-based system considered during this research is not intended to replace the expert, rather to suggest possible solutions to engineers who are not as experienced in the design of bus priority systems. In this context, the relationship between the expert and client could be viewed as being similar to that which would exist if the client were to telephone the expert to ask advice on possible solutions for a particular problem.

The expert could provide suggestions to try, give advice on the relevant data to collect and, given the appropriate information, an opinion of the data values provided as well as the resulting solution. At this stage in the process the expert is not committed to detailed design of the infrastructure, nor to any particular operating system. The result is simply a number of suggestions of how reasonable it might be to examine various approaches in more detail.

Bearing this in mind, the interaction between the expert and the client should be in terms that the client can understand. It is expected that the client is familiar with the engineering aspects of transport, and thus the information that such a system can give the user is geared towards describing the attributes of the proposed system. Typical information may comprise, for example, the number and width of traffic lanes; the position of the bus lanes or the location of bus stops and their design.

For the expert to reach such conclusions, however, the client must provide some information, and the expert can steer the client towards the provision of the correct data. The expert also knows the level of accuracy that is desirable or necessary for subsequent decisions and can advise the client when an improvement in the quality of the data might yield a more confident response. To this end, the expert can suggest suitable methods for data collection when required.

Especially during the more conceptual phases of the design process, experts may use sketches and diagrams in order to convey information to, or obtain information from, their clients. Such graphical images are able to convey a large amount of information, especially where it is difficult to describe this in words. A sketch of a design for a bus stop is much easier to understand than a detailed verbal description.

6.2.2.5 Interaction between the knowledge-based system and the user

In the case of the knowledge-based system, the part of the expert is taken by the system, and the client is in effect the user of the system. The types of interaction described above can be represented using a knowledge-based system. The use of graphics capabilities of microcomputers can be used to ease the interface between the knowledge-based system and the user. These capabilities can also be used to assist the user to communicate with the system.

As suggested in Chapter 5, data can also be described in terms of minimum, maximum and preferred values, with some idea of how the user rates their own estimates of the data values. This can be achieved graphically, so that the user presents a 'shape' to represent their opinion of the particular data range. Solutions can be sketched on the screen to amplify or replace a verbal description of the proposal. This capability extends the image of the expert suggested above (being able to give advice by telephone) to the ability to communicate suggestions, including diagrams where these are considered helpful, by FAX machine.

The interface should also allow the user to ask why various decisions have been taken, and explanations in various levels of detail should be available. Thus the experienced user, who simply wants a quick opinion should not have to read through large amounts of explanation, and an engineer to whom these concepts are relatively new can obtain a detailed explanation of the complete decision process.

6.2.2.6 Representation of the image of the expert

In this research, the chosen level of expertise for the knowledge-based system is that of preliminary advice. Thus, for a given corridor a type of system can be suggested, together with sketches of the appropriate infrastructure to show how the system would appear. The level of detail should not go beyond this because of the difficulties of defining the necessary details. Besides if the user is an experienced engineer who simply does not have particular expertise in the possibilities which are available for bus systems, the engineering diagrams can be generated by the engineer in the normal way from the data which is provided by the knowledge-based system.

Having decided on the level and type of expertise it is desirable to collect and the relationship between the knowledge-based system and the user, the issues surrounding the acquisition of this knowledge are now discussed. The next two sections discuss the three areas of knowledge acquisition mentioned above. The next section considers the issues concerning the acquisition of knowledge about the structure of the decision process and the factual information required to design a high capacity bus system. Section 6.4 discusses the methods used to acquire the opinion distributions.

6.3 ACQUISITION OF FACTUAL KNOWLEDGE

This section discusses the process used to obtain factual knowledge about the design of high capacity bus systems. The first aspect of factual knowledge acquisition is the development of a model of the structure of the design problem. On a detailed level, the decision structure may be defined differently by different experts. It is therefore important to analyze the structure at a level which is low enough that there is a reasonable degree of agreement between the experts. This makes it possible to have a basic structure for the computer program which is appropriate for all the experts consulted. The higher-level differences between the experts can then be represented using the opinion model derived in Chapter 5.

The second aspect of factual knowledge acquisition is the recording of the factual knowledge itself. This includes the details of the models preferred by each expert at each stage of the process outlined in the decision structure. The first activity in the knowledge acquisition process is therefore to obtain the basic structure of the decision process. This is followed by acquisition of the more detailed facts which constitute the design process.

This section is therefore divided into two parts. The first considers the structure of the design problem, and the second discusses the acquisition of factual knowledge.

6.3.1 Problem structure

In order to determine the actual course of the decision process, some form of model is required which identifies not only the various decision points, but also the inputs and outputs necessary for each decision. At some points in the decision process, it may be necessary to calculate values for a number of parameters before the subsequent decision may be taken.

If the parameters are not inter-dependent, the order in which these are calculated is unimportant. In other cases it may be necessary to calculate one value before it is possible to calculate another. Where a decision is dependent on the output of such a model (or one model is dependent on the output of another), it is clear that the appropriate calculation should be performed first. Any data necessary for this calculation should therefore be available at this point. In terms of the structure of the problem, these data are considered to be direct inputs for the decision, and are therefore regarded as necessary prerequisites for the relevant point in the problem structure.

Consequently, it is important to establish the basic structure in terms of those parts of the decision process which may be represented in a parallel structure (that is, they can be performed in any order) and those that are sequential in nature. Thus with some knowledge of the data requirements of each decision, together with any models and parameters used for its solution, it is possible to determine a sequence of decision points and their inputs which forms the basic structure of the problem.

The methods by which factual knowledge can be acquired are now considered. This is considered in six parts. First, the process of acquisition is considered in the next section. Section 6.3.1.2 discusses the overall decision structure for the design of high capacity bus systems. The following four sections consider the four principal elements of this structure, including the sub-structures which underlie each decision point in the overall process: the initial assessment of the current system, the preliminary hypothesis of a possible design, the (theoretical) application of this design to the corridor as a whole and the initial evaluation of this application.

6.3.1.1 Acquisition process for the decision structure

This section considers some important general issues surrounding the methods used to acquire from experts the knowledge concerning the decision structure for the design of high capacity systems. Five general issues are discussed. The first is the location for the interviews with the experts. The decision structure used by the expert is affected by the necessity of establishing values for, and using, parameters. In many cases these are treated as constants. This may not always be sufficiently general, and the second general issue to be considered is the analysis of these constant values and their effect on the decision structure. Since the time available for working with experts is usually heavily constrained, the length of time taken for knowledge acquisition is another issue discussed here. The fourth issue considered in this section is the checking of the model of the decision structure with that of the expert. Finally, the fifth section discusses briefly the nature of the output of this process.

6.3.1.1.1 Location for the interviews

Where factual knowledge is concerned, the principal means of knowledge acquisition used in this research is the personal interview with the expert. The environment in which these interviews take place is extremely important. In order to acquire different parts of the knowledge, four types of environment are used:-

1 **the expert's own office**. Here the expert has more or less immediate access to his or her own references (which may be in the form of reports, working papers, notes, drawings, computer models and, of course, other members of the design team);

2 **sites at which the expert has been involved with the design and implementation of a high capacity bus system.** This permits easier explanations of difficulties encountered (for example, particular constraints such as junction design and capacity requirements, limited available width or exceptional demand characteristics). Visits to these sites enable the knowledge acquirer to see at first hand the solutions devised for the problem (whether these are successful or not).

3 **sites at which no high capacity bus system has been designed or implemented, but which might be candidates for such treatment.** This is particularly useful for acquiring knowledge about the initial assessment of the corridor in terms of its potential for the operation of a high capacity bus system.

4 **the office of the knowledge acquirer.** This has the advantage that the interview is less likely to be interrupted than one undertaken at the expert's own office. The knowledge acquirer also has access to such equipment (for example, computers, models and tape recorders) as may be useful to facilitate the knowledge acquisition process.

In the process of knowledge acquisition, all of these environments may be used. In the case of this research, for example, for one part of this process a Brazilian engineer (Szász) was brought to London in order to determine the way in which he or she set about the task of designing a high capacity bus system in a new environment. The interviews were undertaken in environments which could be listed under the last two categories above. For other stages in the acquisition process, interviews were conducted in Szász's own office and at relevant sites in Sao Paulo.

6.3.1.1.2 Consideration of constant values

It should be emphasized once again that the knowledge acquirer should have a reasonable knowledge of the design problem: it is important to know when the expert is oversimplifying the problem. For example, it is sometimes the case that an expert, in presenting a model to the knowledge acquirer, includes some constant values. It is essential to establish the source of all values which are treated as constants in the models used by an expert. These may be categorized in three ways:-

1 **variable-constant**: where the value should be established for a particular environment, and is then used as a constant. An example of this is the marginal boarding time, which is (as discussed in Chapter 3) critical for calculating the capacity of a bus stop.

2 **variable**: where the value is in fact the result of some form of calculation which should be evaluated as appropriate. An example of this is the optimum distance between bus stops. The value of this variable could, of course, be the result of either a quantitative calculation or a qualitative assessment.

3 **global constant**: where this is generally acceptable for all environments. An example of this could be the value of walking speed.

It is important to establish into which category a constant should be placed because this can affect the decision structure. This may alter the structure in the sense that some variables would be required at an earlier stage in the computer model than in the engineer's own methodology. Where necessary, the decision structure suggested by the expert may also have to be adapted to allow for the inclusion of a model to calculate a variable or variable-constant at the appropriate point.

6.3.1.1.3 Time requirements for knowledge acquisition

A major practical difficulty to be resolved in the knowledge acquisition process is the amount of time taken to obtain this knowledge from the expert. The expert is unlikely to have unlimited time to devote to the knowledge acquisition process so it is important to use the available time to the best advantage. Previous knowledge of the problem domain is helpful in that it enables the expert to talk freely, without the need to translate his or her thoughts into 'layman's' language. If necessary, this can be done later by the knowledge acquirer. It is interesting to note that a knowledge engineer (as defined in this thesis) is unlikely to be capable of achieving this as it requires a fair degree of knowledge of the problem and of the reasoning underlying the expert's interpretation of it. In fact, the rationale behind the use of a knowledge engineer is that this forces the expert to undertake this activity.

It is also useful for the expert to be able to represent the structure graphically where this is helpful. Once again, with a reasonable understanding of the problem, the knowledge acquirer can translate these sketches into a more formal representation at a later stage.

6.3.1.1.4 Checking the model

Having obtained and understood the structure from the expert, it is necessary to check that the version to be used in the proposed knowledge-based system is sufficient. For the reasons stated in Section 6.3.1.1.2 above, the structure of the model and that of the expert may be different. Sufficiency in this context is therefore concerned with the end product of the decision process.

It is important that the model generates output that is similar to that of the expert when given the same data. Checking this necessitates further interviews with the expert until the final version is agreed. These checking sessions mainly involve the evaluation of the logical formalisms generated by the processes described in Chapter 4 and earlier in this chapter. Checking the thresholds within the opinion model itself, while this may be performed at the same time as the checking of factual knowledge, is considered in more detail in Section 6.4 below. This also has an effect on the time taken for the knowledge acquisition process.

As mentioned above, Szász was brought to London for part of this process. This has the advantage that both the acquisition and the subsequent checking process can be undertaken without the interruptions which are inevitable when busy experts are involved. The use of the time is therefore much more productive than in the more normal case where experts are interviewed in their normal environment.

6.3.1.1.5 Outputs

The output of this process is a model of the decision structure of the problem, as perceived by the expert. This enables the structure of the proposed knowledge-based system to be designed. As mentioned above, the structure of the knowledge-based system may not match exactly that of the expert. There are two reasons why this may be acceptable. First, this could be for the reasons stated above concerning the question of constants. Second, this could be because there may be constraints brought about by the use of a computer, for example where memory limitations dictate that a single process is divided into two or more parts. Once again, a prior knowledge of the problem can help the knowledge acquirer to identify these problems and to present to the expert a suitable adaptation of the given structure. It is important, however, that the basic structure of the decision process is the same in the computer model as it is in the expert's mind.

The next section considers the basic decision structure for the design of high capacity bus systems.

6.3.1.2 Basic decision structure

In order to illustrate the knowledge acquisition process, an example is considered in this section. The example to be considered is the basic structure of the decision process. After a brief consideration of the basic structure, this section is in four further parts, where each part is concerned with one decision point in the basic structure.

A problem with as much complexity as the design of a high capacity bus system has many underlying structures. The first structure to be defined is the overall decision 'management' structure. This includes the basic sequence of activities necessary to design a bus system. This management structure may be identified by the expert as a list of questions, for example:-

1 is the current bus system encountering problems which are likely to be solved by the implementation of a high capacity bus system?

2 what type of system is most likely to resolve this problem?

3 are the physical constraints of the corridor in question likely to preclude the implementation of such a system?

4 do changes to the system (for example, required by the physical constraints) invalidate its use?

5 what are the potential benefits and disbenefits of the proposed system?

These questions can be structured in such a way so that the logic of the decision sequence appears reasonable and so that the data collection process can be minimized. A simple representation of this structure, showing the data and expertise requirements at each stage of the design process is shown in Figure 6.1.

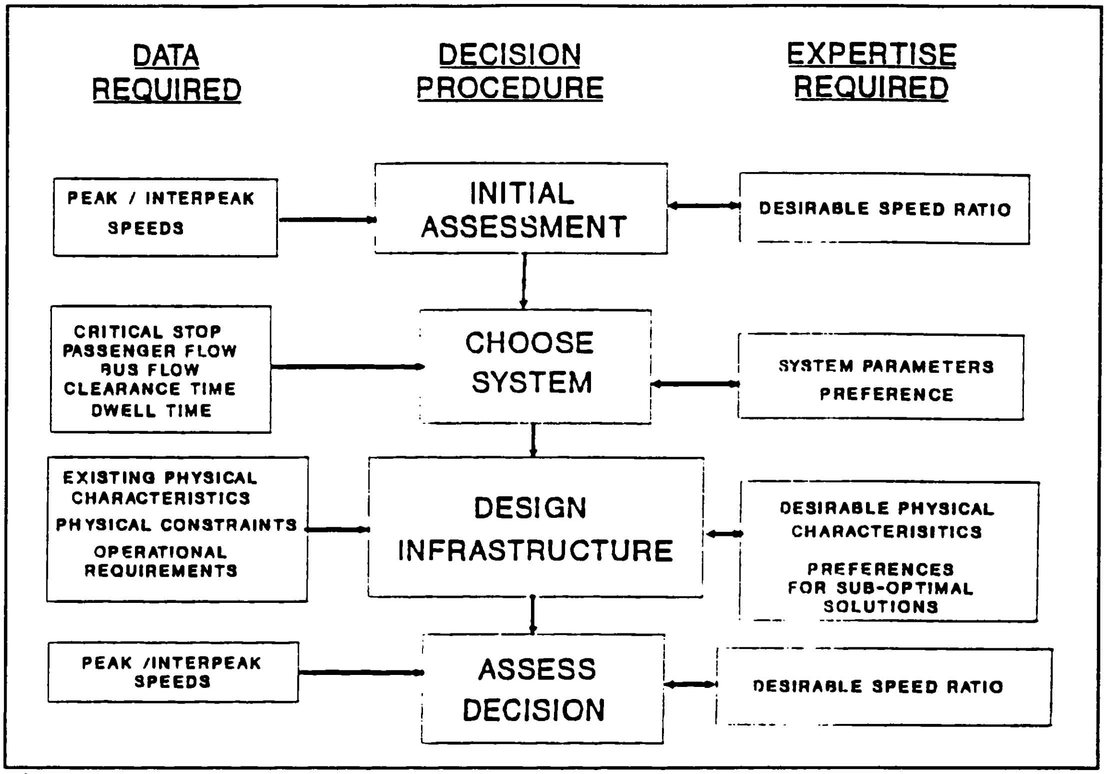


Figure 6.1 A simple representation of the decision structure of the "management" of the design process.

Each of the elements of this structure may have a separate underlying decision structure. It is possible, however, that the decision point consists of just one decision. An example of this is the first stage in the decision process: the initial assessment of the current conditions of the corridor in terms of the usefulness of proceeding with the design of a high capacity bus system.

6.3.1.2.1 Initial assessment

The first decision to be taken in the design process is whether to proceed with the design process at all. This is important, as an incorrect decision here may involve unnecessary expense in obtaining data for impossible or unnecessary infrastructure. Thus it is important to be able to make such a decision on the basis of a limited amount of data. Interviews with Brazilian bus system designers indicate that this decision can be based on a simple assessment of the commercial speeds obtained in the corridor at various times of day. Two conditions need to be evaluated:-

1 the absolute commercial speed during the peak; and

2 the relationship between this and the off-peak commercial speed.

If the corridor under consideration has very low peak speeds, it is possible that a bus system could be designed to increase the commercial speed (and therefore reduce the journey time). On the other hand, a high capacity bus system is envisaged as being able to remove the difference between the commercial speed observed in the peak, and that observed in the off-peak. Thus a corridor where the peak commercial speed is already the same as the off-peak speed is unlikely to obtain a great deal of benefit from the implementation of a high capacity bus system. If this is combined with a low off-peak speed, however, a bus system may be able to benefit both peak and off-peak users, and therefore the design process should continue. In some cases, the peak speed is greater than the off-peak speed. This may be because passengers using the system in the peak move faster, or because there is less interference from parking in the peak. In such cases, the use of a bus system could still be justifiable: the consideration here is that the bus system should not be subject to such differences.

This decision might be expressed by an expert in the following way:-

If the peak commercial speed (PS) is low, a bus system (BS) could be designed to improve its performance. If the peak commercial speed is high, then it is probably unnecessary to implement a bus system. If the peak commercial speed is low and approximately equal to the off-peak speed (OPS), then a bus system may be able to help, but it is probable that other measures (OM) will be required as well in order to benefit the bus system. If the peak speed is low and the off-peak speed is high, it seems that a well-designed bus system may be able to achieve similar speeds in the peak as are currently being obtained in the off-peak.

The first consideration is the logic which underlies this expertise. As described in Chapter 4, this expertise can be reduce to logical formalisms:-

$$
\begin{aligned}
PS\ [\ low\ ] &\supset BS \\
PS\ [\ high\ ] &\supset \neg BS \\
PS\ [\ low\ ] \wedge OPS\ [\ high\ ] &\supset BS \\
PS\ [\ low\ ] \wedge OPS\ [\ low\ ] &\supset BS \wedge OM
\end{aligned}
\qquad \textbf{(6.1)}
$$

Using the methods described in Chapter 4, these formalisms can then be converted into predicates:-

```
BS(PS,OPS) :-
  PS(low),
  OPS(high);

  PS(low),
  OPS(low),
  OM.

BS(_,_) :-
  QUIT.
```

In the predicate formulation, the NOT operator is evaluated as a default: if all the other combinations do not succeed, the predicate is terminated. Since this is a desired result (in the sense that it is valid and does not constitute a failure so backtracking is not required) the predicate still succeeds. In this case the success results in the termination of the design of the bus system (since this is considered unnecessary). It should also be noted that for the purposes of this illustration, all the terms are represented in this predicate. When this is encoded in PROLOG, for example, the logic may be taken one step further by controlling the consideration of each clause with the value of the parameter PS:-

```
BS(PS(low),OPS(high)).

BS(PS(low),_) :-
  OM.

BS(_,_) :-
  QUIT.
```

This example also shows a typical interaction between the three types of knowledge identified above. The knowledge that the commercial speeds need to be evaluated and the ratio between them calculated in order to make a decision is factual knowledge. This type of knowledge also includes the rule that links the results of these calculations to the eventual decision. The model of the peak speed/off-peak speed ratio, although simple, is an example of the use of a calculation to assist the decision process. The decision is not, however, taken only on the basis of the results of this calculation. It is the interpretation of the result that is used. There is an element of qualitative assessment of these results (for example, '*high*','*low*'), and the establishment of which qualitative description best describes a particular speed is an example of the use of opinion and judgement based on experience.

This example also indicates which data are required for this decision. As suggested above, in this case it is desirable to have a minimal data requirement so that a preliminary decision can be taken at a minimum cost. Data which are relatively easy to obtain (especially if the level of precision required is not too high) are the commercial speeds along the corridor at various times of day. As the data relating to the commercial speeds are evaluated in the decision model according to linguistic descriptors, it is of course necessary to convert the actual data into this linguistic form. For this conversion, the opinion model derived in Chapter 5 is required (using expertise collected as described in Section 6.4 below).

The result of this element of the decision structure is binary: either the design process should proceed or it should not. This element of the decision process is therefore very simple. If the result is not to proceed, the decision process stops. If the decision is to continue, the next element of the decision process is tackled.

6.3.1.2.2 Preliminary assessment of possible designs

This element of the decision process considers the possible system designs that may be applicable for the corridor. As this is a preliminary assessment, and in order to reduce the data requirement, this decision is taken on the basis of a critical point in the corridor. The rationale for this is that the basic system is dictated by the point at which it is most difficult to operate in the corridor. It is therefore useful to investigate the appropriate solution for this point and then test the hypothesis that this system is suitable for the rest of the corridor. For the reasons discussed in Chapter 3, this assessment is usually concerned with the critical bus stop.

With information about the bus and passenger flows, it is possible to estimate the required capacity of the stop. As described in Chapter 3, this carries the implication of certain designs of stop which in turn has further implications for the design of the system. At this stage it is possible to identify whether the system should be located in the median or at the kerbside, and whether overtaking is desirable at the bus stops. Basic consideration can also be given to the particular design of the critical stop, for example whether the stop should have parallel platforms, the number of berths and any requirement for grouping. These design considerations may have implications for the whole system. For example, the grouping of buses is likely to be the case for the whole corridor, rather than just for the critical bus stop.

The result of this process is a suggested operating system for the corridor, and the next stage is to evaluate the possibility of designing this system in the corridor as a whole.

6.3.1.2.3 Corridor design

This is perhaps the most complex decision structure within the design process. It can, however, be defined in three parts: the location and design of stops, the design of junctions and the design of links. For these processes it is necessary to have a large amount of data about the physical nature of the corridor. In addition, data concerning the passenger, bus, and traffic flows are required. Since the principles involved are the same for each of these elements of the corridor, the structure for one of them - the location and design of bus stops - will be described.

As discussed in Section 3.3.5.2, the location of bus stops takes the demand for boarding and alighting along the corridor and distributes it amongst the various access points. It is therefore necessary to establish the relationship between the current location of bus stops and the current passenger demand. Once this has been established, the redistribution of this demand as a function of any new stop locations can be calculated using a model preferred by the expert. For this reason the relocation of bus stops is considered first.

Section 3.3.5.3 discusses the variety of models available for the calculation of the location of bus stops. There is a common basis (the minimization of cost or time) for them all, however, and the most significant difference between experts appears to lie in the estimation of parameters rather than in the choice of model itself. It must be stressed of course, that there are some types of bus stop which cannot be relocated and must remain in their current location: an interchange at a metro station serves as an example.

Once the stops have been relocated using the location model, it is necessary to calculate the redistributed demand. This establishes the capacity required for each stop which in turn (as mentioned above) determines the design. The next operation is then to attempt to fit the stops into the corridor, taking into account any physical constraints. This is a two-dimensional process: the stops must fit in terms of both their length and width.

In terms of their length, the major constraint on the location of stops is the location of junctions. It is possible that the stop location may need to be adjusted in order to maintain sufficient capacity at the junction. The effect of the width constraint is that a stop may be reduced in capacity (for example because it is not possible to include an overtaking lane) as a result of a restricted amount of available width. Thus if a parallel bus stop is required to provide sufficient capacity, but the corridor is too narrow at this point to accommodate such a bus stop, then some other alternative should be found. For this reason, it is necessary to test whether it is possible to provide a bus stop with the chosen capacity at the stop location suggested by the location model. In some cases it may not be possible to locate a bus stop in the desired position. In extreme cases this restriction may be so severe that it may render the feasibility of the chosen bus system itself open to question.

Therefore the first decision point in this part of the design process is to establish the current passenger demand, together with an indication of its distribution along the corridor. Using the chosen location model, the theoretical location of bus stops is then established. The demand can then be redistributed to the new stop locations and the new saturation levels calculated. In some cases, this process is sufficient to resolve the problem, and the design process could therefore be terminated at this point with a recommendation to relocate the stops. More often, however, it is necessary to redesign the bus stops, and the next part of the design process concerns this problem.

For the purposes of bus stop design, the roadway is considered to comprise four elements: bus lane(s), traffic lanes, footways and bus stop platforms. The allocation of space to these elements is a question of finding the best mix of widths. The exact allocation of space to these elements is a question of the preferences of particular experts, and is therefore a matter of opinion. A diagram of the structure of this process is shown in Figure 6.2.

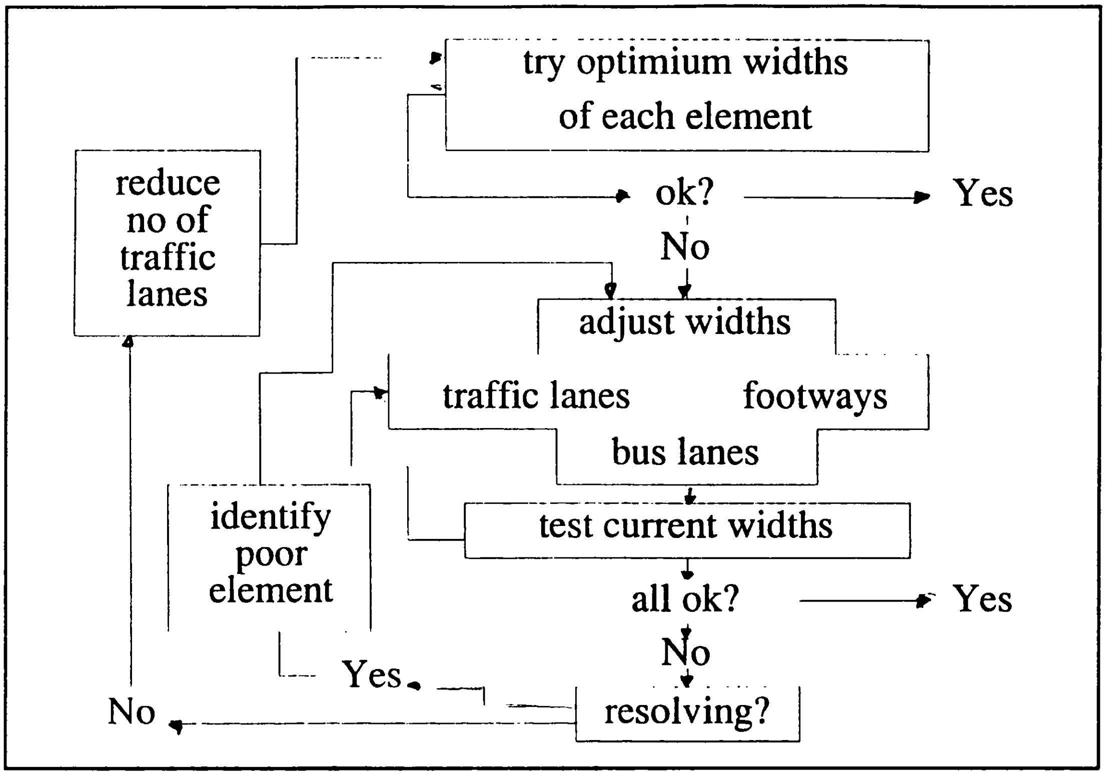


Figure 6.2 Decision structure for the assessment of physical constraints in the design of a bus stop.

As can be seen in Figure 6.2, there is a complex arrangement of iterative processes involving the estimation of the width of each element. These iterations are controlled by the opinion of the expert in considering the feasibility of the width of each element. These constraints may be described as follows:-

If the widths of the bus lane, the traffic lanes and the footway are all satisfactory, then the design is satisfactory. If the width of any one of these elements is unsatisfactory, then the design should be altered to increase the width of the relevant element. This may require a reduction in the width of the other elements or a more drastic change such as a reduction in the number of traffic lanes. If this is the case, the widths of all elements should be reassessed. If any element is too wide, then this suggests a poor use of the available space, and therefore other elements could be increased in width (or the number of traffic lanes could be increased). If the situation arises that the design is impossible without some *element(s) being narrower than the ideal, it may be necessary to use the design even with suboptimal widths in order that the wider benefit of the system as a whole may be realised.*

Although this expertise is more complex than that illustrated previously in this section, the corresponding logical representation is simpler. This apparent paradox arises because the conditions surrounding the acceptance or rejection of particular sets of widths is a matter of the interpretation of the results rather than the results themselves. Thus while the logical representation could be as follows:-

$$\begin{gathered} BL(ok) \wedge TL(ok) \wedge FW(ok) \supset DESIGN(ok) \\ \neg\, BL(ok) \vee \neg\, TL(ok) \vee \neg\, FW(ok) \supset \neg\, DESIGN(ok) \end{gathered} \qquad \mathbf{(6.2)}$$

this does not capture the richness of the expertise as represented above. In fact the heuristic represented in this logical formalism is merely the skeleton of the decision process - it is necessary to encapsulate the reasons behind the definition of what is meant, for example, by '*NOT (ok)*'. This richness is provided by the use of models to calculate the widths, and then to generate the opinions of these widths using the opinion model.

The predicate formalism equivalent to the predicates given above for the initial assessment process is not therefore informative because it relies on the inputs from the opinion model in order to determine whether a particular width is acceptable or not. Since these opinions are specific to particular experts, a deterministic heuristic is not possible in this case as the constraints may change when a different expert is consulted. The logical formalism is therefore encoded as such in the computer model, but the definition of '*ok*' is left as an input provided by the opinion model. In this way the logical formalism in Equation 6.2 is the controller of the iteration process, using variables obtained from the use of the opinion model and the associated expertise. Thus there is a structure which underlies the decision process at this stage, which is illustrated in Figure 6.3.

This structure shows how the opinion model is used to evaluate the results of other models so that the decision process can proceed. In this way, the opinion model lends flexibility to the logical process without loss of the logical rigour required of the decision process. The logical formalism which represents this process is therefore more complex than that shown above and is represented as follows:-

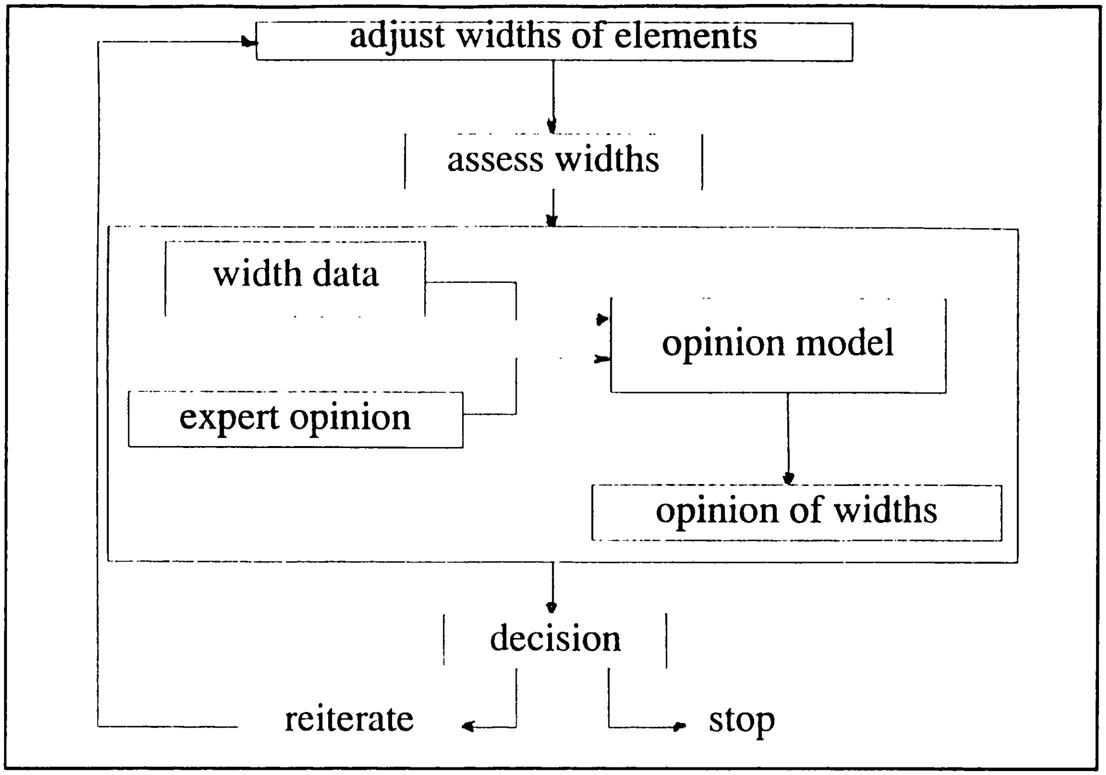


Figure 6.3 Representation of the control of the iteration process used in the determination of the widths of the various elements in the design of a bus stop.

$$\begin{array}{l} \forall e; \\ CALCULATE\ (\ element_e,\ element\text{-}width_e\ ) \\ \wedge\ OPINION\ (\ element\text{-}width_e,\ (status_e\ =\ ok\ )\ ) \\ \supset DESIGN_e(ok) \\ \vee\ \neg\ DESIGN_e(ok) \supset REDESIGN\ (\ element_e\ ) \end{array} \qquad \textbf{(6.3)}$$

In order to evaluate the design, it is necessary to calculate the width of each element and then to evaluate the opinion of this width. This is repeated for each appropriate element. If the output obtained from the opinion model (status) for all the elements is satisfactory, then the complete design is considered to be satisfactory. The fourth line of Equation 6.3 allows for the alternative if there is any element which is not considered to have a satisfactory width: the relevant element must be redesigned.

The factual knowledge (which determines the elements to be assessed at a particular stage of the design process), the use of models to calculate the values for these elements and the calculation of the opinion values for the elemental values are thus combined to provide the basis for the logical control of the iteration process.

The use of the opinion model to evaluate the results of other models is also found in the process of evaluation of the design.

6.3.1.2.4 Evaluation

Evaluation of the design requires the estimation of some performance parameters. The results of these estimates may be that the proposed design should be changed in order to provide a better overall performance. The structure of the evaluation process is therefore based on the repetition of testing and amending designs until a satisfactory result is achieved.

The tests are generally models which calculate the values of the performance parameters, although in some cases the evaluation may be expressed linguistically. In these cases, the evaluation parameters are measured according to the experience and opinion of the expert, and so this is another area where the opinion model presented in Chapter 5 can prove useful.

In order to amend a design, the iteration process described above is modified only in that the choice of action is more open. Thus instead of simply altering the widths of critical elements, the result of an unfavourable evaluation could be that the form of the bus system should be changed. This requires a second attempt at the design process. this can be achieved relatively simply by calling the appropriate routines for the design, but excluding the possibility of the current result.

As discussed in Chapter 9, however, in the case of the model generated in this research, it is difficult to evaluate the effectiveness of the designs produced by this decision model in terms of the bus operation. There are three reasons for this. First, it is unlikely that it is possible to evaluate the design in reality until the system is built. Second, there are too few examples actually in operation to ensure that the results are obtained from the system and not from other characteristics of the particular locations. Third, simulation models do not exist that are capable of representing the subtleties of the designs produced by this model. This is clearly an area for urgent further research. For these reasons, this element of the decision structure is not represented directly in the decision model generated by this research.

6.3.2 Factual knowledge

6.3.2.1 Introduction

The decision structure discussed in Section 6.3.1 determines the sequence (or pattern) of events in the decision process. This section considers the acquisition of knowledge concerning these events which comprise the factual knowledge possessed by the experts. Factual knowledge in this context is considered to be the actual heuristics and the various models used by an expert during the design decision process. The factual knowledge acquired during this research has been presented in Chapter 3; this section is concerned only with the acquisition process.

The factual knowledge used by the experts in high capacity bus system design is concerned with the use of particular models, the data required for their use, and the inferences which may be made as a result of either calculations or evaluations of these data. The acquisition of factual knowledge therefore aims to identify which models are used (together with the necessary input variables) and the inferences that may be drawn from their results. The interpretation of these inferences and the evaluation of imprecise data are discussed in section 6.4 below.

The acquisition of factual knowledge is a different type of process from that used to ascertain the structure of the problem. This does not mean that the two types of knowledge are never acquired simultaneously. Particularly where the knowledge concerned in the form of heuristics, the structure of the problem and the factual knowledge used to solve it are often perceived to be the same.

Nevertheless, in order to model either the factual knowledge or the structure it is necessary for the knowledge acquirer to perceive the two elements separately. Thus an interview which discusses the capacity of a bus stop, for example, includes discussions over the problem structure (including where this calculation fits into the structure of the problem as a whole as well as the structure for calculating the capacity itself). This interview could naturally extend to the use of a particular model for the calculation, together with the strengths and weaknesses of this model in this situation.

As discussed in Section 6.3.1, the structure of the problem is determined from the experts and modified in order to ensure the maximum generality and suitability for representation within the constraints imposed by the use of computers. Factual knowledge, on the other hand, may be categorized in two ways. The first category comprises the heuristics and models which are more or less generally agreed amongst the experts. The second category consists of heuristics and models which are used specifically by particular individuals. Thus in the acquisition of factual knowledge, it is necessary for the knowledge acquirer to establish into which of these two categories a particular piece of knowledge falls.

This section is divided into two parts. First, the type of model about which there is a general agreement is discussed. Section 6.3.2.3 then considers how models used by one or two individual experts may be incorporated into a knowledge-based system.

6.3.2.2 Knowledge that is generally agreed

This section considers the issues surrounding the acquisition of knowledge which is generally agreed amongst experts in the domain. As mentioned above, this section is concerned with the acquisition of knowledge rather than with the knowledge itself (which has been discussed already in Chapter 3).

One characteristic of such general knowledge is that it is generally at a low level. Thus experts may agree on how to calculate the value of a passenger car unit, or how to measure the saturation flow of a junction. They may be less likely to agree on the higher level issues within the domain: the calculation of the capacity of a bus stop is an interesting example of this which will be discussed in Section 6.3.2.3 below.

In order to collect factual knowledge it is necessary at first to establish what is generally agreed, and what is peculiar to a particular expert. This should be done by the knowledge acquirer before meeting the expert.

Knowledge that is generally agreed is identifiable from the general literature. A literature survey therefore provides a good grounding for the knowledge acquirer both in the development of knowledge in the domain and also in establishing the areas of commonality between the experts. This can save a lot of time with the experts: it is not productive to ask an expert about something which can be easily discovered from the published literature. Indeed, the time with the expert should be used to tease out the ways in which he or she differs from the general consensus (to be discussed in Section 6.3.2.3).

It is also useful to have a clear idea about the various implementations of high capacity bus systems, so that discussions may take place about their design and effectiveness without too much reliance om any expert's own interpretation of these. Previous work of this nature facilitates the objectivity of the knowledge acquirer, and thus the quality of the knowledge within the system should be enhanced.

6.3.2.3 Knowledge that is specific to particular experts

Having prepared a reasonable bank of knowledge, the knowledge acquirer is ready to embark on the second part of the acquisition process. This takes the form of a series of direct questions about each issue under consideration. This has the advantage that the acquisition session is directed towards the subject matter of interest to the knowledge acquirer. The ability of the knowledge acquirer to know which matters are of importance is dependent on the amount of work performed before the session, and comprises part of the previous knowledge of the domain which has been recommended above.

Thus it is likely to be more productive if the knowledge acquirer were to ask a question of the form:-

Can you explain why your model for the capacity of a bus stop differs from the one used by XXXXX?

than if the question were like this:-

What is your model for the capacity of a bus stop?

The former question will obtain a more explicit answer than the latter because it is asking, not for the model itself, but for the reasons underlying its construction. The request for a comparison with some other model helps to focus the expert's attention on these differences and the reasons for them. By concentrating on the expert's reasoning about these differences, the knowledge acquirer is able to obtain an insight into the basic principles upon which the model is based. As discussed in Chapter 4, the reduction of the problem to its basic principles is extremely important when attempting to ensure that a model is as general as possible.

The latter question may obtain a quicker response: it is much easier for the expert to describe the appropriate formula than it is for him to explain why the formula is constructed in this way.

It is also important that the knowledge acquirer is quite clear about the expertise being acquired. If necessary it may be necessary to return to the expert in order to ensure that the recorded knowledge is correct. This implies that the knowledge acquirer should check all formulae and models to ensure that they are consistent: any inconsistencies should then be checked with the expert.

Such inconsistencies may arise for two reasons. First, the knowledge acquirer may have recorded the information incorrectly or incompletely. Second, the model concerned may have some internal inconsistencies. The expert may feel that within the scope of his or her own activities these inconsistencies are containable (that is, the results are within a reasonable margin of error), and therefore that they are not important. It is extremely important that the knowledge acquirer is fully aware of these inconsistencies (often they are in fact limits of operation of the model) so that if the model is to be included in the knowledge-based system, the knowledge about the inconsistencies is included as well.

6.4 KNOWLEDGE ACQUISITION FOR THE OPINION MODEL

The acquisition of knowledge for the opinion model involves the derivation of the opinion distributions described in Chapter 5. Unlike the acquisition of factual knowledge, however, the definition of an opinion distribution is not generally familiar to the experts, and detailed explanations by the knowledge acquirer are often required in order that the expert understands what is required.

Using the formulae derived in Chapter 5, it is possible to construct a distribution which can be altered by the expert in order to obtain the expertise. This can be achieved interactively with a computer using a graphical display of an opinion distribution. This distribution can be altered by redefining the necessary points through which the curve should pass. These points are defined with reference to the variable under consideration (for example the minimum, maximum, preferred and any other value (together with its associated opinion value)) and can be identified using a cursor in such a way that the opinion distribution is forced to pass through these points. Thus the expert can effectively draw the opinion distribution on the computer screen.

This process is repeated until the expert is satisfied with the distribution. At this point the parameter values which define the shape of the particular distribution can be calculated and recorded for future use. Thus the calibration of these opinion distributions (discussed in detail in Section 6.4.2) takes place interactively with the expert, the knowledge acquirer and the computer. This process is repeated for each parameter in which the expert uses opinion in the decision process, and for the set of distributions required in each case, as discussed in Chapter 5.

As mentioned above, there may be a problem for the expert in providing a numerical estimate of a particular opinion value. This value may be necessary in order to calculate the distribution. Early attempts at knowledge acquisition required the expert to provide a numerical estimate of the opinion value of the 'other' data value. This was found to be difficult because the experts found it difficult to quantify their opinions in this way. It was, however, observed that once a distribution was displayed on the screen, the expert found no difficulty in modifying the distribution until satisfied with the resulting shape.

Accordingly, the knowledge acquisition process was modified so that no numbers were required as input for opinion values. The expert is now shown an initial distribution, calculated from a constant set of parameters within the program, and is asked to modify the distribution using the cursors to locate points through which the distribution should pass. The opinion distributions are therefore perceived in terms of a 'shape' rather than a precise mathematical formulation. The parameter values required in order to generate the distribution are calculated by the computer from the definition by the expert of selected points through which the curve should pass using Equations (5.2), (5.3), (5.4), (5.5) and (5.6). There is no necessity for the expert to be aware of the parameter values. This process can be repeated until the expert is satisfied with the results.

It should also be noted that the reasons underlying the expert's preferences (as opposed to the reasoning) are not encapsulated in these distributions. These reasons affect the logical formalism in the sense that they determine the actual logic. They are not an inherent part of the representation of opinion except in the sense that they may affect the location and variance of particular distributions within the set (as in the example above). Nevertheless, it is important to represent these reasons as they form a major part of the explanation of the decisions that are taken during the design process. Accordingly these reasons, together with any other comments, are also obtained from the expert at the same time as the opinion distributions. These may include aspects of the decision, such as the impact on construction or operating costs of a particular design. The exact comment is dependent on the expert.

The initial set of distributions, which are presented to the expert as described above, are accompanied by a corresponding set of statements which describe the expert's interpretation of each distribution. These can also be modified by the expert in order to convey a more comprehensive explanation to the user.

The acquisition of a set of opinion distributions is now described using an example from the knowledge acquisition program used in this research. The issue considered in this example is the position of a bus stop in the cross-section of the roadway. As mentioned in Chapter 3, there are two possibilities: either the stop can be positioned at the kerbside or in the median.

6.4.1 Acquisition process for the opinion model

This section illustrates the acquisition of a set of opinion distributions concerning the question of the position of a bus stop within the cross section of the roadway. This is an area in which there is a considerable amount of disagreement between the experts. Some consider that the bus stops should always be at the kerbside, since this provides easy access for passengers. Others suggest that bus stops (and therefore any associated bus lanes) should be positioned in the median in order to avoid problems with parking and to help provide a reserved track for the bus system. It is necessary therefore to have a method for the representation of this expertise that is able to reflect these differences and to provide a suitable degree of flexibility within the knowledge-based system.

6.4.1.1 Use of opinion models to represent differences of opinion

Using the kerbside position as a default, it is then a matter of defining at what point an expert feels that it is necessary or desirable to move the stop to the median. Those experts who consider the median position tend to cite parking and enforcement as a major reason for this choice. Since these experts have considered that there is a choice of position, it is important to analyze their decision methodology. This may be described as follows:-

If the land use is such that the level of parking could be expected to be high (for example a large amount of shopping activity), it may be better to position the bus system in the median in order to reduce the interference from parking and loading or unloading activities. If the enforcement system is adequate and observed, this may reduce the necessity of moving the position to the median. If the expected parking level is low, and/or the enforcement is good, the system could be positioned at the kerbside. If enforcement is poor, it may be necessary to have the bus system in the median even where the parking level is expected to be quite low.

This suggests that some measure is needed to describe the levels of parking and enforcement. If these can be derived, the keenness or reluctance of different experts to position the bus system in the median can then be represented using different sets of opinion distributions. It is important to consider that the logical formalism, and the possible outputs are separate issues. The logical formalism underlying this decision is constant for all experts, but the differences between experts can be represented using their opinions. The logical formalism may be written:-

$$\begin{aligned}
&PARK_{low} \supset POS_{kerbside} \\
&PARK_{vlow} \supset POS_{kerbside} \\
&PARK_{vhigh} \supset POS_{median} \\
&PARK_{high} \wedge \left(ENF_{high} \vee ENF_{vhigh}\right) \supset POS_{kerbside} \\
&PARK_{high} \wedge \left(ENF_{low} \vee ENF_{vlow}\right) \supset POS_{median}
\end{aligned} \tag{6.4}$$

It is clear from this formalism that the level of enforcement is only taken into consideration when the level of parking is considered to be high. If this is considered to be very high, or low, the position of the bus stop is determined solely by the parking level. Clearly, the actual decision about the bus stop position is taken on the basis of the expert's evaluation of these descriptors.

It is therefore necessary to seek expert opinion about these descriptors. Two sets of distributions concerning parking levels are illustrated, which indicate how these opinions may vary. These distributions attempt to describe linguistically the concept of the likelihood of parking along the relevant stretch of the corridor. Figure 6.4 shows a set where the implied preference is for a

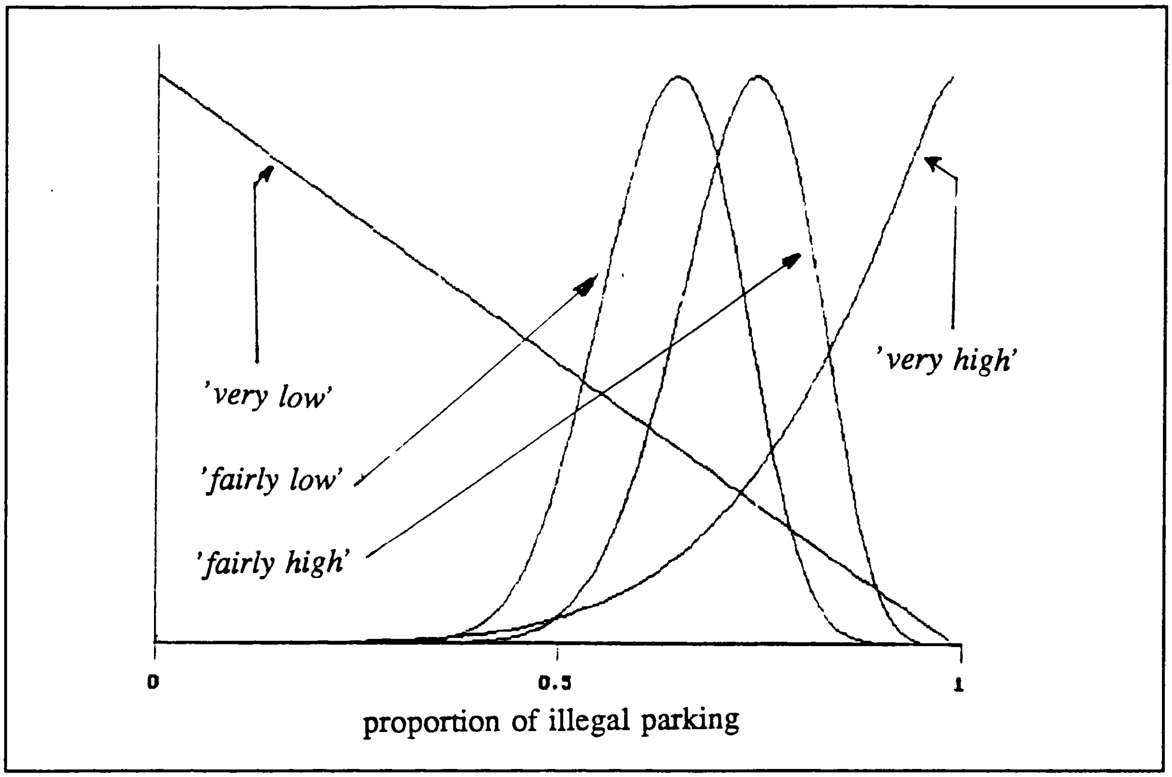


Figure 6.4 Set of opinion distributions concerning the evaluation of various descriptors of parking levels, implying a preference for a kerbside position.

kerbside position. This is reflected in the location (at the extreme right-hand end of the x-axis) and the small variance of the distributions describing the descriptors *'high'* and *'very high'*. The location and variance combine to reduce the chance that a parking level will be described with one of these descriptors. Reference to the logical formalism (6.3) above shows that it is therefore highly likely that a bus stop would be located at the kerbside if this expert were to be consulted.

Figure 6.5, on the other hand, shows a more even spread of the distributions and therefore the expert represented in this example is more likely to consider the possibility of a median position for the bus stop. Clearly, an expert who insisted on a median location could be expected to locate the distributions describing *'low'* and *'very low'* at the extreme left hand end of the graph and to define a distribution with a small variance. This would have the corresponding effect to that illustrated in Figure 6.4. In this way, the opinion model is able to produce different decision outputs based on the same data, depending on the particular preferences of the expert being consulted.

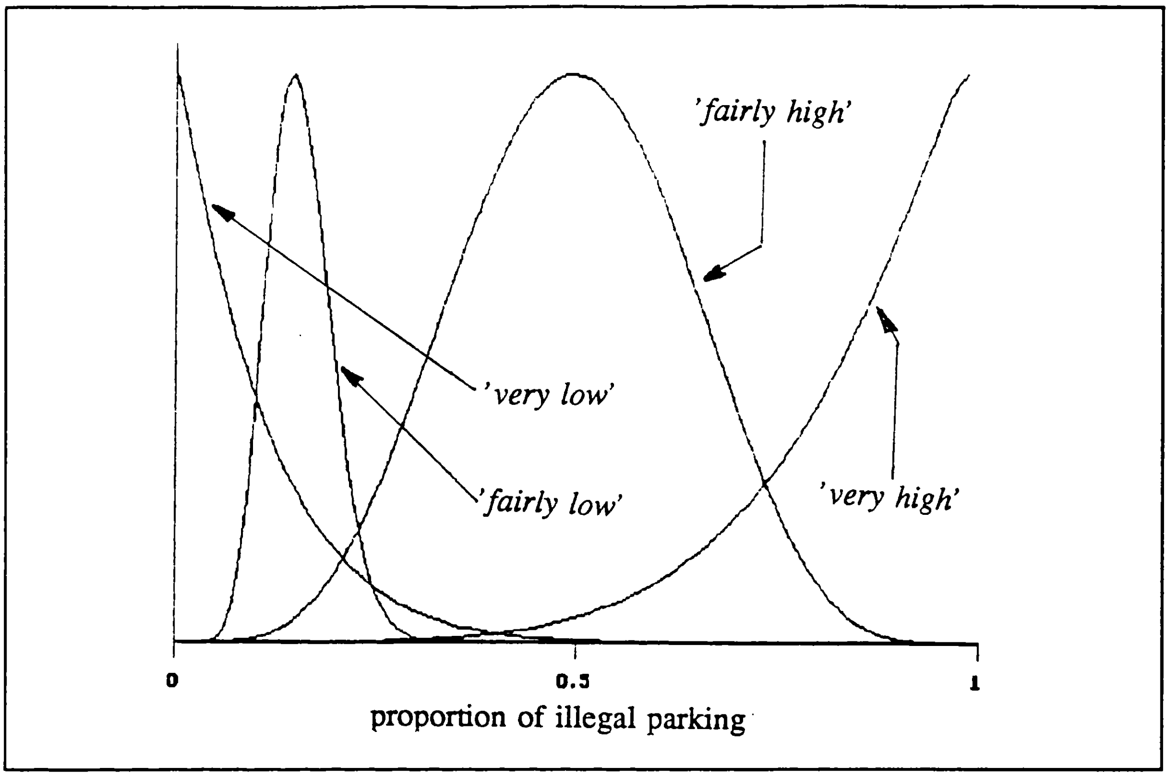


Figure 6.5 Set of opinion distributions concerning the evaluation of various descriptors of parking levels, implying a preference for a median position.

6.4.1.2 Representation of logical control using the opinion model

In order to maintain a logical control over the decision process, it is necessary to place a small constraint on the expert. This is because the opinion model, in order to be as general as possible, and as explained in Chapter 5, has a fixed number of distributions. Where the expert wishes to consider a lower number of descriptors, this may result in an apparently arbitrary difference between the descriptors of some of the distributions in a set. This does not appear to affect the distributions which describe the major alternatives (such as whether or not to proceed with the design for a bus lane). It may, however, affect whether a parking level is described as '*low*' or '*very low*'. The disadvantage of this is that it may be difficult for an expert to conceive of these distinctions when asked to derive their distributions. Once this has been explained to the expert, there is not usually a problem: there is always the possibility of making the variance so small that the chance of the option being considered at all is remote (see for example Figure 6.4).

As can be seen from the logical formalism (6.3) above, this is unlikely to have a major effect on the subsequent decision. This may result in a situation where two different descriptors result in the same decision, and it may be felt that it should be possible to reduce the logical formalism to remove such redundant elements. While this is clearly correct in purely logical terms, the practical advantage of being able to use the same model form throughout the decision process has been felt to outweigh the slight disadvantage which may arise during the knowledge acquisition process.

6.4.2 Calibration of the opinion distributions

It is important to be sure that a decision reached by the expert may be reproduced by the decision model. Thus for given data, the decision outputs should be the same whether the expert or the model is used. In order to ensure that this is the case, it is necessary to check that the model provides output that is in agreement with the expert.

This is achieved during the interactive parts of the knowledge acquisition process. With both factual knowledge and the acquisition of opinion distributions, the expert is actively involved with the generation of the computer model. The generation of opinion distributions, since it is less familiar as a process to the expert, is more likely than the acquisition of factual knowledge to require further substantiation. It is important to realise, however, that the model should be tested with the results for specific values, rather than for ranges of data. This is because it is necessary for the test to establish whether the thresholds identified during the acquisition process are correct. This part of the calibration process is therefore concerned with the intersections which occur between the distributions of two descriptors, and the evaluation of particular data values below and above the threshold value.

6.4.2.1 Calibration of thresholds

Calibration is undertaken at different acquisition sessions to those in which the original distributions are acquired in order to test the expert's continued acceptance of the distribution obtained. The procedure is to display the current set of opinion distributions. The expert is then given the result which occurs if a particular data value is chosen for analysis by the opinion model. If this produces a result with which the expert disagrees, the expert can alter the offending distributions in order to improve the agreement between the expert and the model. Any such changes are then tested for other values to ensure that the distribution remains valid for other values.

It should be emphasized, however, that the process is testing the expert as much as the model in the sense that it is also asking for evidence of consistency in the expert's responses. For this reason, the separation of the testing process from the acquisition process helps to ensure that the distribution that is finally adopted in the model is a representation of the expert's opinion of the values concerned. This is not the only criterion to calibrate, however, and it is also necessary to ensure that decision sequences produce equivalent output in the model and the expert.

6.4.2.2 Calibration of decision sequences

If each set of opinion distributions (gathered for each decision point) is calibrated so that the model agrees with the expert, then the question of decision sequences is largely a question of an appropriate representation of the decision structure.

This is also tested during the knowledge acquisition activity, since the structure is determined with the expert. The combination of the opinion distributions with this structure may produce unforeseen results, and it is important that the expert is satisfied that this combination in the model is likely to produce appropriate results. Because of the potential complexity of the possible results which may arise from a given set of data, it is best to compare the entire decision sequence of an expert with that generated by the model for the same set of data. An attractive idea is to take a project which has been implemented, and to input the 'before' data to the model. The design produced by the model should be the same as that actually implemented. This process of validation of the model is described in Chapter 8.

6.5 CONCLUSIONS

This chapter has considered the principles underlying knowledge acquisition and the processes used during this research. Some conclusions may now be drawn about this topic.

6.5.1 Knowledge types

In the case of high capacity bus systems, the knowledge is extremely specialized and exists in two forms. First the generally accepted knowledge which can be found in the literature, and second, the expertise which is more specific to particular experts which is often only available directly form their source.

Expertise may also be categorized in another way. One type of expertise may be defined as factual knowledge from which heuristics may be derived. This knowledge tends to be binary in decision terms (either true or false), although different experts may disagree about the validity of any particular piece of factual knowledge.

Another type of expertise is concerned with the use of models. While the knowledge that a model exists is factual, the use of the model - and the choice of model for a particular circumstance - is a fundamental choice for an expert.

The knowledge acquired which might be categorized into these types forms a large part of the knowledge discussed in Chapter 3. The third type of expertise is the judgement and opinion discussed in Chapters 4 and 5, and which underlies the requirement for an opinion model for its representation.

6.5.2 Expert types

Two types of expert have been considered and defined. A single expert is considered to be either a single person, or a team of experts with different specialities brought together to solve a particular problem.

Multiple experts, on the other hand, are considered to be separate experts who are consulted in order to obtain knowledge about the problem domain.

It has not been possible to devise an algorithm which is capable of differentiating between the different heuristics and judgements of separate experts. It is concluded that to perform this function, it is necessary to obtain some form of 'superexpert'. It is therefore considered better to encapsulate the expertise for each expert in a separate knowledge base and to allow the user to choose the expertise. If the results from the use of different knowledge bases differ, these can be evaluated by the user in the light of the particular circumstances which prevail.

6.5.3 Decision structure

The decision structure consists of two types of process: those which must be undertaken sequentially (and for which the order of this sequence is important), and those which can be undertaken in any order (or in parallel). This has a profound implication on the data acquisition process of the decision model.

The opinion model is necessary for the knowledge acquisition process because it lends flexibility to the logical process with a minimum loss of logical rigour. This flexibility simplifies the logical representation. As the opinion model operates within a logical framework, it has the advantages of the simplicity and directness of a formal logical representation. It also allows, however, the complexities which arise - and which are resolved in the real world by expert judgement - to be represented within this framework.

This has important implications for the knowledge acquisition process. The knowledge concerned is therefore seen in two parts: structure and judgement. The knowledge acquisition process should be guided by this distinction.

6.5.4 Knowledge acquisition process

Knowledge acquisition is generally considered to be a difficult process in the development of knowledge-based systems. A considerable amount of time and effort has been spent by the research community in attempting to reduce the scale of this problem and the time taken to acquire knowledge. As a result, a number of automated systems are available to assist the knowledge acquisition process.

Knowledge acquisition proceeds more easily and efficiently if the knowledge acquirer has some knowledge of the domain before the process commences. This is at variance with the conventional view that the knowledge engineer should be independent of the knowledge domain.

The knowledge acquisition process includes the calibration of the expertise in order to ensure that its representation is correct with respect to the expert concerned. This calibration is required for each parameter and variable considered by the expert.

Validation of the expertise, however, is required, and this should consider the decision sequences of the model, rather than the consideration of individual parameters. This is considered in the next chapter.

CHAPTER 7
A COMPUTER MODEL FOR THE DESIGN OF HIGH CAPACITY BUS SYSTEMS

7.1 INTRODUCTION

The knowledge acquired using the techniques described in Chapter 6 is now combined with the opinion model derived in Chapter 5 in order to attempt a representation of the knowledge itself described in Chapter 3. This chapter considers the representation of this decision process in the form of a computer model. Detailed discussion of the construction of particular predicates is not included here: the issue of importance is the representation of the design decision process and not the technical details of the source code.

This chapter is divided into seven further sections, where each section is concerned with one module. The first five of these sections discuss the design process. The sixth discusses the management module.

7.2 MODULES

7.2.1 Introduction

With a decision process as complex as that for the design of a high capacity bus system, it is unavoidable that the computer model is similarly complex. The approach taken in its construction is to divide the problems into modules. This has four major benefits.

First, the model is more understandable, as it is divided into functional parts, each of which is responsible for a particular activity. If this activity should change, it is possible to alter the model in this respect without affecting other parts.

Second, the memory requirements are much less, since the entire process is not loaded into memory at the same time. This facilitates the use of the program on machines with limited memory capacity (as may commonly be found in developing countries).

Third, this simplifies the problems of redesign of certain aspects of the proposed bus system as it is necessary only to define which module should be reactivated rather than having to find the appropriate part of the program.

Finally, a process may be modelled using the most appropriate computer language. This is normally fraught with difficulties, but since each module is completely separate from the others, the language in which it is written is of no concern: the best language for the circumstances can be used in each case. In fact in this model, only one language (PROLOG, referred to in Chapter 4 above) is used. Nevertheless the possibility exists for other modules to be added in different languages.

Thus the modular approach used here is to ensure that each module is completely separate and independent. Each therefore consists of an executable program which, once started proceeds to its termination at which point, if necessary, another module may be called (the 'management' of the modules is discussed in Section 7.7). There is, however, a necessity for some form of communication between the modules. This is because the results of one module may have a function as inputs to another module: it is therefore necessary that the latter module is made aware of the results of the former. Rather than using specific memory addresses for this purpose, this is achieved using a set of data files. This follows the principles of blackboard architecture discussed in Section 4.2.2, but uses disk files rather than memory for the communication process. The way in which this works is now described.

7.2.2 Communication between modules

The method used to make information available amongst the various modules is by access to files held on disk. Four data files are generated by the program as a whole:

1 input data concerning the existing physical system;

2 input and calculated data concerning the flow variables;

3 a 'design file' containing calculated data concerning the potential design(s); and

4 a 'notepad' file which contains records of the decision processes undertaken during the design process.

A fifth file is also required in order to use the program: the expertise file which includes all the expertise distributions obtained during the knowledge acquisition process. This file is not generated by the program, but its identity is recorded in the notepad file.

The first three file types are conventional and contain the data described above. These are read and updated by the various modules as necessary during the course of the design process. It is necessary therefore for each module to know the names of these files. It is also important to have some record of the current opinions held of the various stages in the process. These data are recorded in the notepad file.

The notepad is a form of blackboard (discussed in Section 4.2.2), but on a much smaller scale. The notepad has the information which allows each module to access the correct data files, ensures that each module makes use of the same expertise, and maintains a record of the results of decisions which have been generated by the opinion model (as discussed in Chapter 5). It does not, however, contain any information about possible future strategies. Where strategy is dependent upon expert opinion, details of strategic choices are recorded in the notepad file. Otherwise strategy is implicit within the decision structure and encoded within each module (including the management module to be discussed in Section 7.7). The notepad file is therefore the key to the inter-module communication system.

7.3 MODULE 1: INITIAL ASSESSMENT

This module performs the function of assessing the corridor for its suitability for a high capacity system. It also sets up the various databases for use in other parts of the program. The module consists of five linked programs. Three of these programs are 'service' programs: one contains the opinion model, another sets out and controls the menu system, and the third contains the graphics procedures. Of the two programs which are functionally concerned with the initial assessment, the first processes the initial assessment, and the second assesses the possibilities for a particular stop in the corridor.

7.3.1 Initial assessment

Being the first module of the program, there is a number of necessary functions to perform, which are not directly related to the design process. These include the definition of the various data files described in Section 7.2.2, the establishment of the correct expertise file and the identity of the corridor (this is important so that the user can always check subsequently that the corridor on which each module is to work is in fact the corridor intended).

This module follows the decision structure outlined above in Chapter 6. Thus, after obtaining some reference information noted above about the corridor to be analyzed, it sets out to perform this assessment on the bases of the commercial speeds, land use and parking enforcement using the minimum of data. It asks whether the user has access to the data it needs, and if not provides advice on how these should be obtained, and if possible, asks for other data from which the required data can be calculated.

7.3.2 Commercial speeds

Thus commercial speed data are requested. If these are not available, the module explains how to collect the appropriate data so that the program can proceed and then asks for estimates of the time taken to travel along the corridor at the appropriate times of day. The module then calculates its own estimates of the commercial speeds from these data.

As indicated in Chapters 3, 5 and 6, the data are collected in the form of opinion distributions. The user is asked for the minimum, maximum and preferred values for the data. These are then used to generate a suggested opinion distribution which is displayed on the screen. The user is then invited to alter the shape of this distribution so that it represents more closely his own opinion. This is achieved by placing two cursors in the appropriate positions (one for the mode, and another for any other point through which the distribution should pass). The user may also alter the minimum and maximum values if so desired. If the initial assessment is being undertaken on a corridor which has been analyzed before, the initial distribution displayed to the user in each case is the one previously recorded.

At this stage in the initial assessment, the user is informed if the implementation of a bus system should be proceeded with, or is so unlikely to produce any benefits that the process should be curtailed at this point.

7.3.3 Parking and enforcement

As described in Chapter 6 above, these data are required in order to establish the position of the bus system in the cross-section of the roadway.

7.3.3.1 Parking

It is neither necessary nor desirable to require at this stage a large amount of detail about the land use along the corridor. Accordingly, the data used to describe land use are obtained by asking about the level of parking expected along the corridor as a whole. These data are requested in the form of an opinion distribution since it is not possible to be precise about such a question.

The level of parking is requested for two reasons. First, the parking level indicates that aspect of land use that is most likely to affect the decision about the position of the bus system (this issue has been discussed in more detail in Chapter 3). Therefore the decision will be made on the basis of the factor most likely to influence a change. Second, the use of this measure to describe land use does not require the exact relationship between land use and parking level to be established. Such a relationship is likely to be inaccurate at best, and would require some form of calibration to be performed for every city (and perhaps for each corridor) in which the model is used. The level of parking is therefore assumed to be a result of the land use and the relationship which applies in the particular environment concerned. Thus this relationship is implicit in the consideration of the level of parking and does not need further elaboration for design purposes.

The user is asked for an opinion distribution for the probability of parking along the corridor. Clearly this is not likely to be constant along the whole corridor. This parameter is used in the decision to position the bus system at a particular position on the cross-section of the roadway, and this is considered to be a system-wide decision. It is influenced by the more influential of two aspects: the parking at a point in the corridor which is critical for some other reason (for example the available width is limited) or the parking at a particular point in the corridor which attracts a high level of parking (for example a shopping centre). The probability of parking for the appropriate point is used.

7.3.3.2 Enforcement

As discussed in Chapter 6, enforcement is only required for the decision process if the level of parking is either very low or high. It could be argued that therefore it is only necessary to obtain data from the user about the level of enforcement in these circumstances. It is felt, however, that it is better to establish all the available data at once to separate the data collection process from the subsequent design process.

The reason for this is that, as concluded in Section 4.5, the design process is not treated interactively: once the data are established, the design process can be left to proceed unattended. If these data were collected at a later stage because a subsequent design decision altered the perception of parking and thereby required information about enforcement, the collection of these data would interrupt the design process. This would be particularly inconvenient if the design were being calculated overnight, for example.

Enforcement is considered to be important especially when it is weak. For this reason, the measure used to describe the enforcement available in the corridor is the proportion of illegally-parked vehicles which are subjected to the appropriate form of enforcement procedure. Again, this is requested in the form of an opinion distribution.

The data obtained in the case of land use described in Section 7.3.3.1, and enforcement described in this section, are converted using the opinion model and the appropriate expertise, to a qualitative form (for example, '*very low*', '*low*','*high*' and '*very high*') which is then appropriate for the decision to be taken using the expert heuristics identified above in Chapter 6. These use the concept of '*low*' parking and '*high*' enforcement, for example, in their conditionals.

The effects of land use and enforcement are thus tested using the opinion model, and the results presented to the user. As with all uses of the opinion model, where there is an alternative which is considered to be possible within the given data ranges, this is also indicated to the user. Such difficulties are, of course, clearly defined in the recording of the decision set for each decision in the notepad file.

At this stage in the process, the user is therefore informed of the choice of position of the bus system in the roadway.

7.3.4 Testing a critical point in the corridor

Before embarking on the complete design process, it is useful to consider a particular critical point in the corridor, using the current opinions about the design (such as its position in the roadway). As mentioned in Chapter 3 above, it is considered that the critical point in the corridor for these purposes is likely to be a bus stop, and therefore a bus stop is considered.

This program therefore asks for approximate data with respect to the physical attributes of the roadway, the bus flows and the passenger flows applicable at the critical point. These are then used to assess the current and potential degree of saturation at the bus stop. On the basis of the proposed degree of saturation, the design is reassessed and may be altered in order to increase or to decrease the capacity of the stop, as appropriate. The resulting design (this may of course be unchanged from the original estimate) is then checked to see if it is likely to fit into the available space. If necessary a further alteration to the design may be made at this point and the new saturation calculated. Indications are given about the ways in which a particular design may be altered to fit the space. For example, a median positioned bus stop with overtaking lanes could be staggered with that for the opposite direction which would reduce the space required without reducing its effective capacity. It is this altered design that is therefore used in the subsequent design process (described in Section 7.5.3) as the initial hypothetical design for the system as a whole.

Approximate data are required for this process - including data describing the physical characteristics which are normally considered to be obtainable as exact numbers. The use of approximate numbers at this point removes the necessity of having complete knowledge about the physical characteristics of this point in the corridor. As with all uses of approximate data within this model, any differences in a decision resulting from different parts of the given data range are identified. This allows the assessment to continue even with inexact knowledge of these characteristics. Because of this it is possible using this module to make initial assessments very quickly - choosing several points in the chosen corridor or even choosing different corridors.

Once this process has been completed, and the decision to retain or change the original design has been made known to the user and recorded in the appropriate files, the module terminates. The next module is concerned with the more detailed data input necessary for the chosen design process to proceed. It is also possible at this point to sketch the plan for the given critical point, using another module.

7.4 MODULE 2: DATA INPUT

This module consists of six linked programs. Like Module 1, however, three are service programs and are not considered here. Of the remaining three, two are concerned with the acquisition of data and the third calculates the degree of saturation of the existing bus stops, and assesses the effects of redesign of these stops. This section is therefore in three main parts. First, the acquisition of physical data is discussed. This is followed by a discussion concerning the acquisition of the data concerning the flows of buses, passengers and other traffic. The third part considers briefly the testing of the current bus stops.

This module (like all others except the initial assessment module) starts by reading the notepad file in order to identify and locate the data files. The acquisition of the physical data then commences.

7.4.1 Physical data

In order to put the other data in a suitable context it is first necessary to obtain data about the physical characteristics of the corridor. This process starts with the design of junctions, and proceeds with the characteristics of the links between the junctions.

7.4.1.1 Junctions

Junctions are described in terms of each arm. The arms along the axis of the corridor are considered first, and are given particular index numbers in order to be identified easily during the design stage. Street names are also recorded for easier identification by the user. Data regarding the number of lanes and the width of each half of each arm (the lanes approaching the junction and the lanes leaving it along each arm). One way streets are considered with the appropriate half of the arm recorded with zero lanes and zero metres. The approach length is also recorded. This is the distance from the kerb line to any change in the number of lanes, or the part of the link that may be considered to store vehicles for the junction if there is no change in the number of lanes. The kerb line is chosen for this measurement in order to ease the measurement calculations to be performed during the design process. The existence of a central reservation (if there is one) is also recorded, along with its width.

If the junction is controlled by traffic signals, this fact is recorded, together with the cycle time and the available green time. While this level of detail is not sufficient for calculations of saturation or delay, it is an indication frequently used (see for example the use of these data in the Highway Capacity Manual referred to in Chapter 3) to ascertain the amount of time available to traffic wishing to traverse the junction. Finally, the possible turning movements are obtained. All data are obtained as precise numbers.

7.4.1.2 Links

Links are considered to be the part of the corridor between the junctions. Therefore, once the information about junctions is known, the program asks the user about the corresponding links. Each link is identified to the user by name and the index numbers of the junctions at either end.

The length of the link between the two points identified as the end of the approach length of the junctions is obtained, together with the width of the roadway, footways, and any central reservation. The number of lanes is also recorded. The user is also invited to add particular constraints to be considered during the design process such as the imposition of a central reservation of a certain width.

7.4.1.3 Bus stops

Bus stops are considered to be located between two junctions. The lack of a more precise location is considered to be important. This is for two reasons. First, it is almost impossible to know the exact location of a bus stop unless there is a well-established and consistent method for choosing the point of the bus stop to use as a measuring point. Second, especially in developing countries, as mentioned in Chapter 3, bus stops are perceived more as an area within which buses stop rather than a particular point in space: in some cases there may be no bus stop *per se* at all. The exact location of a bus stop is therefore unlikely to be known, and since it is only necessary for calculation purposes later (in Section 7.5.2) in the design process (where the distribution of demand over the proposed locations is considered), it is not necessary at this stage to know to the nearest metre the location of a bus stop.

Accordingly, the user is asked to state the junctions between which a bus stop is situated, and an approximated location for them is calculated. For this purpose only, bus stops are assumed to be equidistant from each other and the associated junctions. While this may be an oversimplification, it is only used for the calculation of the current distribution of demand over the existing stops, which is in itself an approximation (see Section 7.4 below), and precision is not required at this stage.

Having established the location of the bus stops, it is possible to add the flows of passengers and buses to the data base.

7.4.2 Flow data

As has been suggested earlier (in Chapter 5 for example), flow data are considered to be known imprecisely. Therefore, the user is requested to construct an opinion distribution to describe the data in the manner described in Section 7.3 above. This section considers the three types of flow of interest to the design process. First the bus flow is discussed in Section 7.4.2.1. This is followed by a brief discussion of the acquisition of passenger flows. Finally traffic data are discussed in Section 7.4.2.3.

7.4.2.1 Bus flows

It is felt that the user should be able to furnish the points at which the buses enter and leave the corridor. Accordingly, details of each bus line using the corridor are obtained. These details include the entry and exit points for the bus line in the corridor, and the approximate flow of buses on each bus line during the appropriate time of the day. The program then calculates the approximate flow of buses at each bus stop in the corridor, the corresponding total flow of buses along each link, and estimates the number of buses making each turning movement at the junctions.

The number of passengers in the bus as it joins the corridor is also obtained in order to allow calculations of loadings along the corridor.

7.4.2.2 Passenger flows

Passenger flows are recorded for each bus line at each bus stop in the form of an opinion distribution in the manner previously described. The approximate number of boarding and alighting passengers per bus at each stop are obtained. These are then combined with the bus flow data in order to estimate the passenger flows along the various parts of the corridor. The technique used for calculating with data which is obtained in the form of an opinion distribution is given in Section 5.4. While this is not an estimate of trip lengths, it gives an indication of the loadings along the corridor, and therefore where the bus flows may be usefully changed.

7.4.2.3 Traffic flows

Traffic flows are slightly different from the flows recorded for buses and passengers in that traffic consists of more than one type of vehicle. It is therefore necessary to record the various vehicle types as well as their flows. A further difficulty is that for subsequent calculations it may be necessary to have an evaluation of the passenger car unit value (pcu) for each vehicle type. If this level of detail is required at a later stage, it is better to obtain it during the data collection phase of the design process (for the reasons stated in Section 7.3.3.2 above with respect to parking enforcement).

It is important that such equivalent values are consistent with their use in other aspects of the traffic engineering situation in a particular city. It is therefore possible for the user to enter the pcu value for any vehicle type. This is then checked with the current knowledge of pcu values and vehicle types. This allows vehicles to be considered which may not be known to the experts being consulted (for example, the opinion of an Indian traffic engineer is likely to be more relevant to the evaluation of the pcu value of a rickshaw than that of a Brazilian expert in high capacity bus systems). If the vehicle type is unknown to the expert, this is added to a special file which contains information which is 'learnt' as the program is used. Such pcu values are also recorded in the form of an opinion distribution. If the user has absolutely no idea of a pcu value for the vehicle type, it is possible to describe the vehicle in terms of the relationship between the power available to it and its weight. An opinion distribution os then shown to the user which can be used until such time as the more information about the vehicle becomes available. It should be stressed that the pcu values so derived are available for use, but it is not usual for them to be used.

Traffic flows are recorded in the form of opinion distributions of the flows making each turning movement defined during the acquisition of data about the current design of the junctions. This indicates the flows along each link and their direction.

Having obtained the data necessary for the design process to commence, the module then considers the current relationship between supply and demand at the bus stops. This relationship is quantified by the degree of saturation.

7.4.3 Current saturation at bus stops

It is useful to ascertain the saturation of the bus stops under the existing conditions while the user is still able to exert an influence. This part of the module aims to check that the model used within the program to calculate the saturation of bus stops (discussed in Chapter 3) agrees reasonably with the expectations of the user. These are calculated and the results of the calculation presented to the user in the form of an opinion distribution which can be altered if desired.

7.4.3.1 Calculation of capacity

In order to calculate the degree of saturation at a bus stop it is necessary to estimate the capacity of the stop. As discussed in Section 3.3.3.2, this can be estimated using a negative exponential function of the number of boarding passengers per bus. The parameters vary as a result of the varying designs and bus characteristics. Figure 3.2 shows two of the curves which result, and shows the difference between one berth and two berths at a bus stop.

Using the appropriate parameters, the capacity can be estimated. The calibration for the particular corridor can be checked by the user after the degree of saturation is calculated.

The degree of saturation is calculated by dividing the flow of buses by the capacity. The user can check that the opinion distribution for the degree of saturation is reasonable, and alter it to some preferred shape and range if desired.

7.4.3.2 Comments on this method

This process is not considered to be ideal because the user is often unable to estimate the current degree of saturation. It is therefore possible that the value used in the subsequent calculations is not correct. One improvement could be to consider the relationship between the degree of saturation and the resulting queuing characteristics which would be more easily identifiable for the user. This is very difficult to achieve, since as discussed in Section 3.3.3.2, this relationship is a time dependent queuing function dependent upon the relationship between the arrival headways at the tail of the queue and the departure headways at the head of the queue. This level of information is not likely to be available at the level of knowledge assumed by this program. The decision was therefore taken to retain the approximate (and possibly erroneous) capacity model described above.

7.4.3.3 Results of this process

The program then suggests the type of bus stop which would be necessary at the current locations in order to achieve a reasonably satisfactory degree of saturation (according to the expert consulted). This allows the current saturation flows to be compared with those that would be obtained if the only change being made to the system were a redesign of the bus stops.

At this stage in the process, the data acquisition is complete and the module terminates after recording all the data obtained in the appropriate data files, and the appropriate decision sets for the use of the opinion model in the notepad file. It is now possible to commence the design process.

The design process is contained in a large number of modules, some of which are dependent upon the basic design parameters decided during the initial assessment and the data acquisition stages described in this section and Section 7.3. These modules are now considered.

7.5 MODULE 3: BUS STOP DESIGN

7.5.1 Introduction

This section considers Module 3, which locates and designs the bus stops for the whole corridor. This module is entirely non-interactive. This module contains four linked programs. Each program performs a different part of the stop location and design process. The opinion model is, in this case, contained within one of these programs rather than in a separate program.

The first program calculates the location of bus stops along the corridor. The second assesses the demand for the existing stop locations, and calculates that for the locations proposed by the first program. The third program calculates the resulting degree of saturation, and decides upon the design of the stop in order to achieve a satisfactory (in the expert's view) degree of saturation. The last program examines the corridor, and attempts to fit the stops designed by the previous program into the corridor defined in the previous module at the locations calculated in the first program of this module. This section therefore considers each of these programs in turn.

7.5.2 Bus stop location

This program calculates the existing stop locations using the methods described in Section 7.3.1.3 above. The first calculation to be performed with respect to the new stop locations is to derive an optimum stop distance. This is derived for each link independently using the model given in Chapter 3 (Equation 3.20), with the number of passengers inside the vehicle derived from the loadings calculated from the passenger and bus flow data. Cost values are, as a default, set at the values suggested by Szász (1990).

Having obtained the optimum stop distance, this is recorded and the module moves on to calculate the demand.

7.5.3 Passenger demand

This program obtains estimates for the passenger demand at each existing stop, and then allocates this demand to the new stops at their new locations.

The assumption made in order to achieve this is that the demand observed at a bus stop is generated as a linear function of the distance between the stops. Thus the demand observed at a bus stop comprises that which is proportional to the distance from the point midway between the bus stop and the previous stop, and that proportional to the distance between the bus stop of interest and the midpoint between this stop and the next. In this thesis the length of corridor from which the passenger demand is allocated to a particular stop is called the demand range of that stop. The linear assumption is that the demand generation is constant per unit of distance. Thus it is a simple matter to calculate the demand generation per unit distance for either side of the bus stop for the existing situation, where the passenger demand is known and the stop location is assumed (as described in Section 7.3.1.3), since this is the demand divided by the appropriate distance. The generation rate may be different on either side of the stop. This allows for fluctuations in demand along the corridor.

The demand for the new stops is then a question of allocating the current demand to the new stop locations. It is assumed that the demand generation rates remain the same after implementation. These, of course, are actually quite likely to change as a result of improvements in the service. It is impossible at this stage, however, to ascertain what form these changes may take. Thus it is reasonable to assume that no change takes place as this is in effect a 'worst case' position.

The demand for the new stops is allocated on the basis that the demand generation rates calculated as described above are uniform per unit distance, and therefore the passengers might be considered to be effectively generated at the distance concerned. Thus if the generation rate were, for example, to be 2 boarding passengers per metre for a particular existing bus stop, this can be taken to mean that two passengers are generated in each metre within the demand range of the stop (whether the stop is an existing one or whether it is newly located).

The reallocation process is illustrated in Figure 7.1. From this it can be seen how passengers may be expected to use the bus stops after the change in location. This is not to say that passengers are actually perceived to come from the exact distance along the corridor. A better explanation is that the distribution of passengers using an existing bus stop is comprised of two uniform distributions, one relevant to the distance from the previous midpoint and the bus stop itself, and the other relevant to the distance from the bus stop to the next midpoint. The distribution of passengers using a new bus stop is exactly the same as that using the existing bus

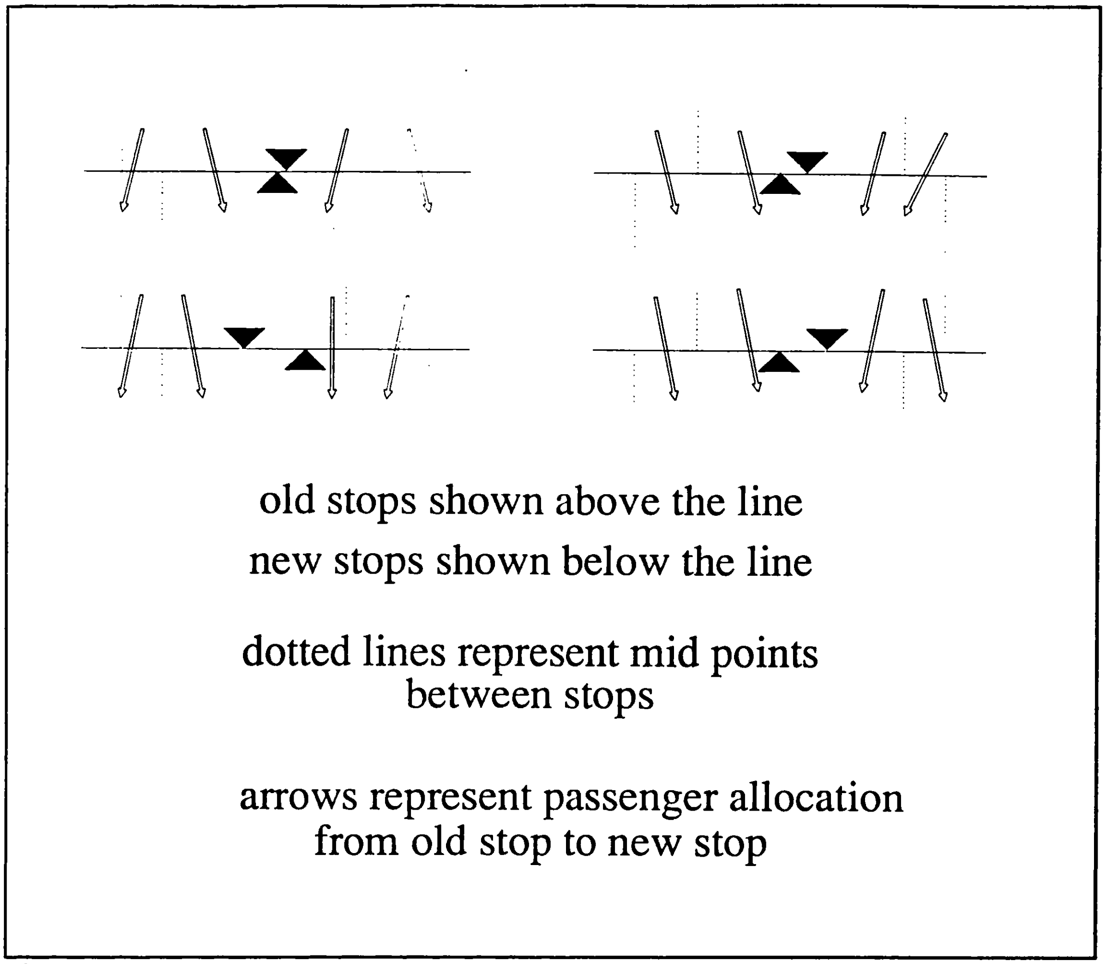


Figure 7.1 Allocation of passenger demand from existing bus stops to proposed bus stops at new locations in the corridor.

stops whose midpoints fall within the relevant midpoints before and after the new stop.

It is not argued that this is the best method to use in order to predict the demand at bus stops once the locations have been changed. It is, however, a method which takes into account the weakness of the data available. Since all the data are approximate and expressed in the form of opinion distributions, the result is an opinion distribution of the demand at the new stops.

The demand having been calculated for the stops at their new locations, it is necessary to design the stops in order to cater for this demand.

7.5.4 Stop design

This program calculates the capacity of each bus stop, given its demand (calculated as described in section 7.5.2), and the initial hypothetical design derived as described in Section 7.3.4

above. Using the expert's opinion distribution, the acceptability of a bus stop design (with the calculated level of demand at the new location) can be assessed. If this is not acceptable (according to the expertise), the design is changed and the process repeated until the result is satisfactory in terms of the expertise.

Having obtained a design which is considered satisfactory for the stops in the corridor, it is then possible to attempt to fit them into the physical constraints of the corridor. In this module this is only attempted with respect to their length, and this is described in the following section.

7.5.5 Fitting new bus stops

This program attempts to fit the newly located and designed bus stops into the corridor, given the constraints which apply as a result of the available length in the links.

Bus stops are initially located at their optimum points determined by the optimum stop distance. The length of the platform determined in the previous program as part of the stop design is then used to determine where the bus stop platform ends. This is then tested to ascertain whether it breaches another constraint, such as a junction. If it does, the bus stop is moved and the test repeated. If the bus stop, when located at the optimum point in the corridor, protrudes into a junction, it may be split so that the platform for one direction is on one side of the junction, and that for the other is located on the opposite arm. This is a common design for a bus stop near to a junction and is considered, for example, in Kraft and Boardman (1971) and Levinson *et al* (1976).

This may result in the optimum distance being impossible to achieve for a number of bus stops in the corridor. However the physical constraints cannot be broken.

At this stage, this module stores all the details about the current design - the stop designs and locations - in the relevant data files. The decision sets are recorded in the notepad file, and the module terminates.

7.6 MODULE 4: CORRIDOR DESIGN

7.6.1 Introduction

This section describes Module 4, which is concerned with the design of the links and junctions in the corridor in order to accommodate the bus system as proposed in Modules 1 to 3. For the purposes of the design of bus systems, links and junctions fall into two categories: those with bus stops and those without. The former category includes (in most cases but dependent on the length of the link, or the approach to a junction) a part of the link or junction arm occupied by the bus stop and other parts in which the special conditions required for the situation of the bus stop are not required. The latter category only requires the reallocation of space to include a bus lane.

This module considers the allocation of space within the roadway to the various elements of the system for both categories of each link and junction. In order to facilitate alterations in the design at a later stage (for example the relocation of a bus stop from one link to another) the allocation of space is calculated for both situations in all links and junctions in this module. The calculations are performed for each link in turn. Once the links have been processed, the process is repeated for each junction.

Module 4 consists of sets of pairs of submodules. Each set of submodule pairs calculates the design of a different type of bus stop. This is to enable easier consideration of alternatives at a later stage: to redesign a corridor, it is necessary only to call the appropriate submodule set. The first element of the set calculates the allocation of space for the bus stop areas. The second element performs this function for links (or junctions) where there is no bus stop. Since this is independent of the bus stop type, this element is in fact the same submodule for each bus stop design. The construction of each element of a submodule is similar: each has three programs. The opinion model occupies one program, and the other two are concerned with calculating the allocation of space. Since the basic structure and principles underlying the methods are the same in each submodule, these will be described in detail in Section 7.6.2 which considers the submodule which allocates space for links and junctions with no bus stops. Section 7.6.3 then discusses the additional aspects considered in the submodule which allocates space for bus stop areas.

7.6.2 Without bus stops

This submodule is constructed from three programs. The first performs an initial assessment, and checks the necessity for involving the more detailed second program in the design process. The third program contains the opinion model. The two design programs are considered separately in Sections 7.6.2.1, which considers the initial attempts to allocate the available width to the various elements of the bus system infrastructure, and 7.6.2.2 which considers the process of calculating and assessing solutions which are suboptimal in terms of available space.

7.6.2.1 Initial space allocation

The concept which underlies this program is that the first preference of the expert is to allocate an optimal width to each of the elements, and to minimize possible congestion effects by maintaining a similar number of lanes for all traffic - even if one is now allocated exclusively to buses.

This program therefore obtains the current overall width from the data file describing the existing corridor. The current number of lanes is also available from this source, as are the current widths of the existing components. The first action is therefore to reduce the number of traffic lanes by one, and to allocate this to buses. The opportunity is then taken to ensure that each element is of an ideal width. The term 'ideal' in this context is considered to be the mode of the opinion distribution acquired from the expert which describes the expert's opinion of a 'satisfactory' (or its semantic equivalent) width. Thus the footway, traffic lanes and the new bus lane are set at their ideal width, and the desired width for the central reservation obtained from the user (as described in Section 7.4) is added. With links, it is considered that for the purposes of this activity the roadway is symmetrical about the absolute centre. Thus the design is the same for both directions. Including a small constant value which allows for segregation infrastructure (such as the eccentric kerbstones described in Chapter 3), the total width required for the new design is calculated. Junctions, however, are assumed to retain any differences between the stop and entry lines that were recorded in the data input process.

This is then compared with the space actually available. If this can accommodate the new design using the optimal widths, then the design stands and the module proceeds to the next link (or junction). If the proposed design requires more space than is actually available, it is necessary to choose a suboptimal combination of widths, or a suboptimal design, in order to have any bus system at all. This process is considered in Section 7.6.2.2.

7.6.2.2 Suboptimal space allocation

This Section considers the methods used to obtain a set of widths which is adequate, but not optimal for the high capacity bus system. Extensive use is made of the opinion model in the manner described in Chapter 5 in order to control the iterative processes involved.

The major difficulty with selecting suboptimal widths is to know which is the best suboptimal combination. In this sense, the experts consulted suggest a few simple heuristics:

1 It is generally better to have narrower traffic lanes than fewer lanes.

2 It is important to maintain the width of bus lanes as far as possible because the vehicles are generally larger than the majority of other vehicle types (it should be remembered that the experts are from Brazil where buses have a standard length of 10-12 metres), and also because the segregation of the bus lane by definition restricts the amount of manoeuvrability available to buses.

3 It is important that footways do not become too narrow as this has a safety implication, particularly where pedestrian flows are high.

These heuristics are represented by the order in which suboptimality is sought. Thus the first strategy is to reduce the width of the traffic lanes, and to maintain the width of the other elements. The reduction is undertaken initially in steps of 1 per cent of the width. This reduction is then apportioned to the other elements in equal parts. Once the width of a traffic lane has been reduced by 1 per cent and the other elements increased accordingly, the width of each element is then tested with the opinion model to check its acceptability according to the chosen expert. The result of this test is a decision set which indicates which elements are unsatisfactory, and the appropriate action to take.

If the offending element is very unsatisfactory, its width is increased by 2 per cent (if it is unsatisfactory, the increase is 1 per cent), and the elements tested again. If more than one element is not satisfactory, the adjustments are made in the order indicated by the heuristics above. After every change in width, all elements are tested for the expert's opinion of each width.

If it proves impossible to obtain a set of widths which are all regarded as satisfactory with the current number of lanes, and the number of traffic lanes exceeds one, the number of traffic lanes is reduced by one. In this case, all the widths are set at the optimum again and the testing process is repeated. If there is only one traffic lane in a current design and therefore it is not possible to reduce the number any further, it is not considered possible to have a bus lane at this point. This fact is recorded in the design data file.

Thus, as described in Chapter 6, the opinion model controls the progress of the iteration, not only in terms of whether it should be terminated but also in terms of which element to tackle next, and by how much.

Once a combination of widths is obtained which is considered to be acceptable, the iteration stops and the next link or junction is tested. After the last item has been processed, the submodule saves the data relating to the designs in the design data file. The decision sets are all recorded in the notepad file and the program terminates. Data are maintained regarding each stage of the iteration. This is to allow the user to examine the way in which the model has arrived at its conclusion.

This part of the design process having been completed for the whole corridor, it is necessary to consider the allocation of space to the bus stop areas.

7.6.3 With bus stops

As indicated in Section 7.6.2, the way in which the space is allocated and iterations controlled is the same in this module as in that for links and junctions without bus stops. The description of this process will therefore not be repeated here. This section considers the extra elements necessary at a bus stop, and the way in which these are taken into account.

This submodule is different from the others because each possible design for a bus stop is accommodated in a different version of the submodule. Which submodule is used initially for any particular design is determined by the design chosen by Module 3. If this turns out to be unsuitable for some reason, the management module calls another design module.

The extra elements required at a bus stop depend upon its design. The two elements concerned are: the platforms and the overtaking lane. These are considered in the following two sections.

7.6.3.1 Bus stop platforms

The amount of space required for bus stop platforms depends in the first instance upon the position of the bus stop in the cross-section of the roadway. The effect of the position of a bus stop is simple, but important. If the bus stop is in the median, space for a platform must be allocated from the road space. If the bus stop is located at the kerbside there is no need to allocate road space to the bus stop platform. It is necessary to check, however, that there is enough space on the footway to accommodate waiting passengers and pedestrians. Nevertheless, the space necessary for waiting passengers is less at the kerbside than in the median because the platform in the median is surrounded on each side by vehicle lanes (either for buses or for general traffic).

As mentioned in Chapter 3, the width of an island platform is considered to be 3.5 metres. The Brazilian experts feel that safety problems may arise if the platform is narrower than this.

7.6.3.2 Overtaking lanes

As discussed in Chapter 3, overtaking lanes are required for most high capacity bus stops, as this allows buses to be independent of the stopping patterns and activities of others. Overtaking lanes require extra space which must be accounted for in the space allocation process. This adds an extra constraint to the iteration process described in Section 7.5.1 above.

This extra constraint applies to all links and those junctions which are expected to be symmetrical about the corridor axis. Asymmetrical junctions are already asymmetrical by virtue of the fact that there is a different number of lanes for each direction. The actual processing of symmetry is calculated as appropriate in each case, by ensuring that the number of traffic lanes in each direction is accommodated and then calculating the space for the appropriate number of bus lanes and platforms. The space taken by the overtaking lanes depends on whether the bus stop is staggered, or whether the bus stop for each direction is located at the same point in the corridor axis. In the latter case, the width required is that occupied by the two extra lanes (one for each direction). The roadway also retains its symmetry in this case, with the centre line separating the two directions of bus lanes lying on the centre axis of the corridor. It may be necessary, however, to reduce the amount of space taken by the bus stop by staggering the platforms for the two directions, and this is indeed normal in implementations of high capacity bus systems (as discussed in Chapter 3).

When a bus stop located in a link is staggered, the width required is reduced by the space taken by the platform and the overtaking lane. The design of the roadway is therefore no longer symmetrical about the centre of the available space. In order to allow the same procedures to be used in the calculation of this type of stop as with others (in order to simplify the source code) each half of the road space is calculated, and the widths apportioned about the (theoretical) centre line.

Thus the available space is apportioned amongst the widths of the given number of traffic lanes, the footways (one for each side of the roadway), one and a half bus lane widths and half the bus platform width. This allocates the correct amount of space to each half of the available space. When the design is presented, the direction served by the part of the bus stop with the platform has, of course, two bus lanes (one for overtaking) and a complete platform. The opposite direction then occupies only one bus lane (plus the segregation infrastructure as described above). The reverse situation applies for the other half of the bus stop. The overall effect is to produce an asymmetrical bus stop similar to that illustrated in Figure 3.7 in Chapter 3. A computer printout of such a design produced by the computer model and allocated to the available space by this module is shown Figure 7.2.

If the available width is such that it is not possible to have overtaking lanes without unsatisfactory compromises being made to the widths of one or more other elements, the program will redesign the bus stop without the overtaking lane. This reflects the action taken in Sao Paulo (and noted in Chapter 3) where such conditions apply: it is felt that it is better to have a high capacity system with a suboptimal bus stop than no system at all.

In such a case, this program records the fact that, according to the expert's opinion, the overtaking lane should be removed. After this program has terminated, the management module

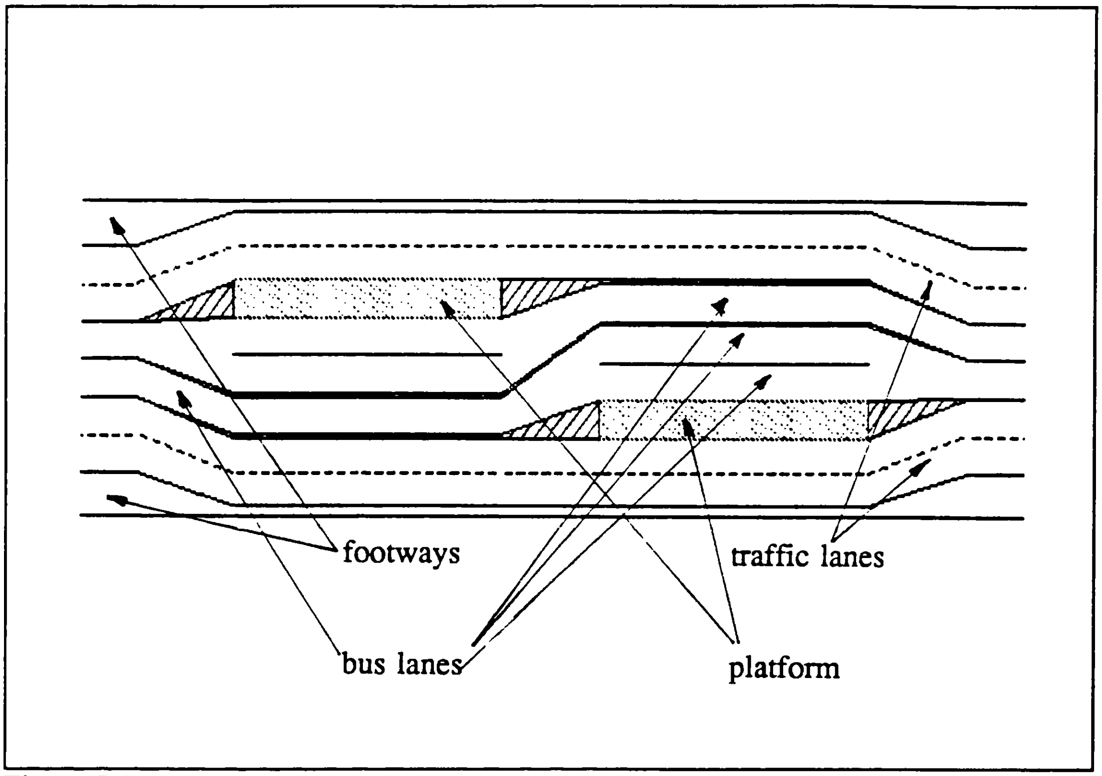


Figure 7.2 A computer-drawn representation of a bus stop with overtaking lanes.

(described below in Section 7.6) detects this comment and calls, for the appropriate links, the module which allocates space for bus stops without overtaking lanes. For these links, therefore the problem is then perceived to be one of fitting a bus stop without overtaking lanes and, as with a reduction in the number of traffic lanes discussed in Section 7.6.2 above, the element widths are returned to their optima and the re-allocation of space is calculated.

If the bus stop is staggered, whether it has overtaking lanes or not there has to be an interchange between the platform for each direction. This is because the layout at the bus stop is asymmetric: Since the platform occupies the equivalent of a bus lane width, the bus lane is usually diverted from a straight line in order to approach the platform. Since this occurs in each direction, the centre of the bus stop between the two platforms contains the lane deviation to accommodate the change in side of the platform. This is accommodated in the bus stop length calculations described in Section 7.4. In terms of the width required for the bus stop, this part of the stop requires no especial calculation: it occupies the same width as the bus stop itself. It is, however, necessary to ensure that the lanes are clearly marked (especially where this transition not only includes the change in platform but also in the direction of the overtaking lane). This point can be seen clearly represented in Figure 7.2.

When bus stops are located at junctions, the iterative process proceeds in the same way as for the links. However, it is not necessary to accommodate the transition between directions within the infrastructure itself. This is because the transition occurs across the junction itself.

It may be, of course, that it is impossible to fit even a suboptimal bus stop in a particular link. In this case, this fact is recorded in the design file and the design for the corridor adjusted accordingly.

Once the process has been completed for all links, the program records all the data describing the progress through the design process in the design data file (and the notepad file where applicable). The program then terminates.

7.7 MODULE 5: MANAGEMENT

7.7.1 Introduction

In this chapter several references have been made to the possibility of calling modules, submodules or programs for particular purposes. This activity is undertaken by a module which is specifically allocated to the management of the other modules. This section considers this module. Section 7.7.2 discusses briefly the actual structure of programs included in this computer program and the way in which the management module tackles its role within this structure. The implications of this form of construction are considered in Section 7.7.3.

7.7.2 Management

Having discussed each module in the previous sections within this chapter, it is instructive, before discussing the management module, to consider the way in which these modules interact, and the overall structure in which the design process is represented in this program.

The nature of the design problem, as discussed in Chapter 3 above, requires a large degree of flexibility in the procedure by which a design is reached. As mentioned at the beginning of this chapter, the modular structure is intended to represent this flexibility. In order to maintain a coherent procedure, however, it is necessary to have some form of management control of the process. This is undertaken by the management module described in this section. In Section

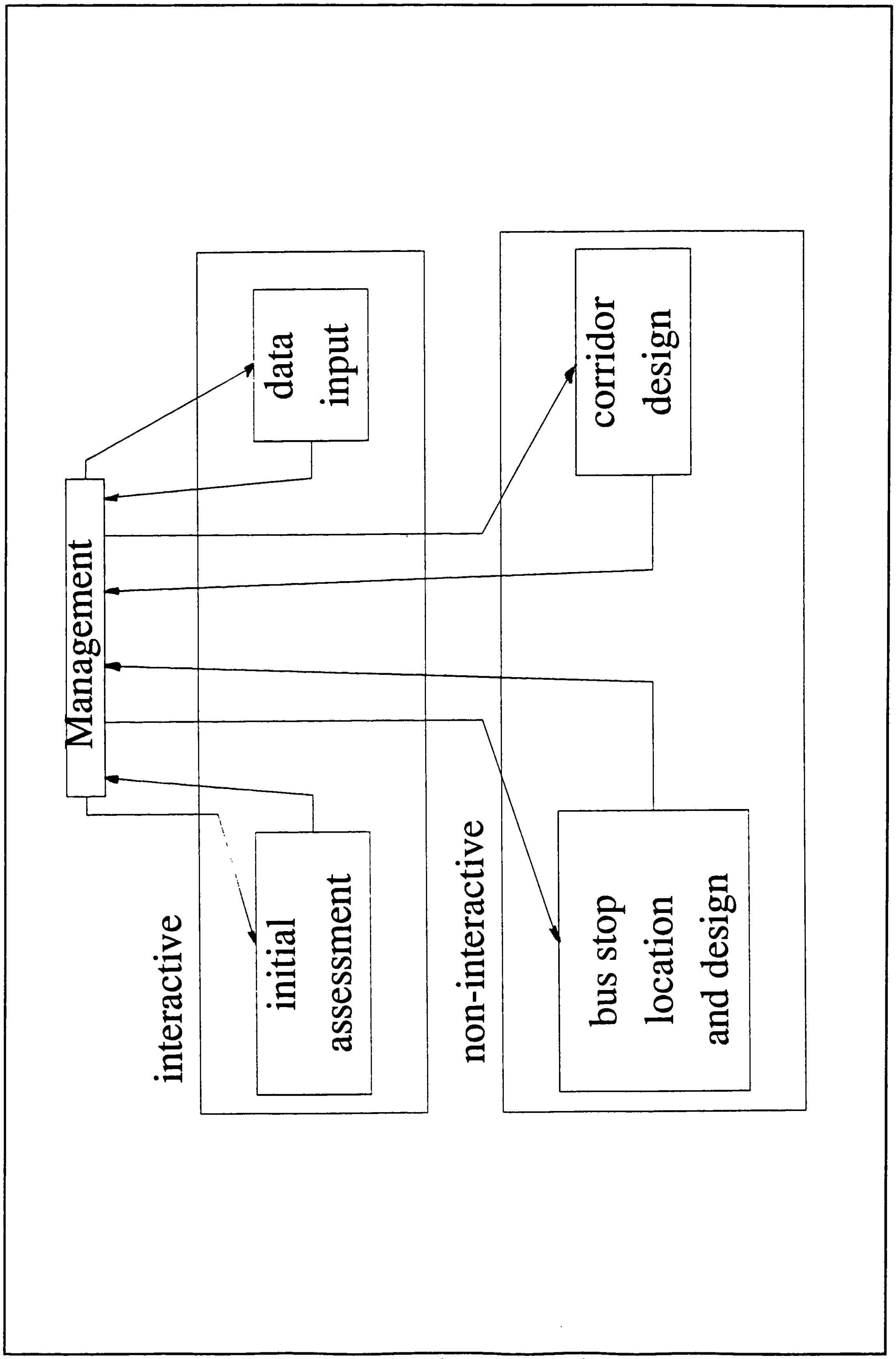


Figure 7.3 The overall structure of the modules within the design model.

7.7.2.1, in order to set the management role in context, the structure of the overall program is described, and the role of the management module within this structure is examined. Section 7.7.2.2 discusses the technical aspects of the management module.

7.7.2.1 Program structure

The program structure is illustrated in Figure 7.3. It can be seen from this diagram of the structure that it is possible to alter the design of the bus system by calling different modules, and even different programs within a module. Thus at its simplest, a module is a single executable program, which can be called independently of any other as long as there exists within the appropriate data files the necessary data in order for a particular program to operate. The design module described in Section 7.6 is more complex and consists of a number of executable programs which are called as required.

7.7.2.1.1 Executable programs

By maintaining different processes within functionally separate executable programs, it is possible to revise designs at any stage in the process, without the need for complex routines within the programs themselves. Thus a revision of the design (for example to remove overtaking lanes from certain bus stops) is undertaken simply by calling the appropriate module. Since each module records its activities in the appropriate data files and the notepad file, it is possible to trace the course of the decision process from the initial assessment to the final design.

It could be remarked that the design modules (described broadly in Section 7.6) could be amalgamated to provide one overall executable program, rather than a number of executable programs. The reason why this is not practical is due to the memory limitations of personal computers. As each program records its progress in the data files, the size of these files (particularly if there is a lot of revisions or the iterative processes are extensive) may become very large. A single design program could also be expected to be large, and therefore it is possible that such a program would cease to function. This would be as a result of the combination of the memory requirements of the program itself, the data files and any internal needs (such as the iterative processes). In practice, therefore, it has been found to be simpler and more reliable in this respect to divide the design process into some constituent parts with a separate executable program in each case.

7.7.2.1.2 Data files

It is also clear from Figure 7.3 that the data files are of considerable importance within the functional structure of the program. Not only do they contain all the appropriate data, but they provide the means of communication between executable programs. It is therefore not necessary to have a low-level control of memory addresses, for example, in order to pass information through the system. The use of data files for this purpose is, of course, slower than the use of random access memory. Where the processes are not interactive, however, this is not a particular problem.

7.7.2.1.3 Management role

The role of the management module in the program is to control the overall process. This is not, however, just a question of calling the correct programs in the correct sequence - such a role would be trivial. In this case, the management module is also required to choose the correct programs on the basis of the results of previous processes. Thus the management module is at a higher level than the other programs and oversees the whole process, with access to the data files and (very importantly) the notepad and expertise files.

The management program makes choices on the basis of a combination of the output from a program within the design module and the expertise. This combination reveals the appropriate choice of subsequent activity, which is then called by the management module by means of a system call.

7.7.2.2 Management methods

The management module is required to be small as it is the only module of the entire program which is required to be resident in memory throughout the design process. In order to achieve the type of control described in Section 7.7.2.1, therefore, it is necessary to have a level of contribution from the other modules. As described in Section 7.7.2.1, many of the management decisions are taken on the basis of the expert's opinion of the results of a particular process. Clearly, if the management module performed this function, it would become similar in size to the other modules, and the size constraint may be infringed. For this reason, the determination of the expert's opinion of such a result is made by the program concerned and recorded in the notepad file as a decision set (as described in Chapter 6). The management module is then required only to read this decision set and to take the appropriate action.

This process is made easier by the use of indices to identify the source of each decision set so that the management module reads the appropriate set for each case. The management module is therefore a very simple heuristic-based program which contains predicates which are combined by the use of AND and OR operators in order to establish the correct procedure through the design process. Thus it is this module which represents the basic design process: the details needed for the validation of the conditionals are obtained from the other modules as necessary. It is interesting to note, therefore, that the management module has no need to read any data file other than the notepad file.

Of course, not all decisions are taken on the basis of expert opinion. Other decisions are simply a question of the correct sequence of programs (for example, the bus stop design module is called in every case, and the type of bus stop to be designed is established within the module).

7.7.3 Implications of modular program construction

The use of a number of executable programs for the various design processes, together with a small management program has some implications for the representation of such a complex process in the form of a computer model. These are now discussed. Section 7.7.3.1 considers the disadvantages of this form of structure, and Section 7.7.3.2 discusses the advantages.

7.7.3.1 Disadvantages

There is a number of disadvantages which arise as a direct result of the use of this strategy for the representation of a complex process.

7.7.3.1.1 Speed

First it is cumbersome. There is a large number of separate programs which must be available to the management module when required. Each program must find and read the appropriate data files in order to function, and must write any alterations to these files before it terminates. The reading and writing processes can be very slow especially when the data files are large.

This is offset by the use of different data files for different purposes (as described in Section 7.3 above). Because the data are split into their functional form and recorded in separate files, it may not be necessary to read from - or write to - all of the files in every case. This division of the data also reduces the size of the files which eases the drain on the memory resources and also the time required to read and write.

7.7.3.1.2 Size

The total size of the program is very large. This is due, to a large extent, to the necessity to repeat functions in many modules. For example, the opinion model is included in each module (except the management module). The design modules have several functions which are nearly identical, but which must be coded in their entirety in each case because the programs are independent.

The comments made in Section 7.7.2 at various points, to the effect that the separation of these functions allows for easier control and management could be reiterated here to offset the disadvantage of the program size. Comments made in Section 7.7.3.2 below on this subject are also to be considered in this context.

7.7.3.1.3 Appearance

The repeated use of system calls to call the various programs has an effect on the appearance of the program to the user. Because programs each have to read and write data, there appears a number of hiatuses in the running of the program which may appear strange to the user.

On the other hand, there is the possibility of terminating the program in mid stream with few unforeseen problems of continuity. This is particularly important when the program is being used in countries with an uncertain electricity supply. If the power fails during the operation of a particular module, the data will not be saved. Once power starts again, the management module simply finds where it is in the process and starts the module again. Where the procedure is interactive (for example the data input modules), the user can continue the input from the last record. Where the procedure is not interactive, the program, once restarted, will start again from the beginning and read the appropriate data files before continuing. This of course also applies to a user-initiated termination of the process.

7.7.3.2 Advantages

Four advantages are considered in this section. Some comments further to those made in Section 7.7.3.1 are made in Sections 7.7.3.2.1 and 7.7.3.2.2 on the subjects of size and speed. In addition the flexibility of the program is considered in Section 7.7.3.2.3. Finally, issues concerning any subsequent modification of the program are discussed in Section 7.7.3.2.4.

7.7.3.2.1 Size

Although the size of the program may be a problem, as considered in Section 7.7.3.1.2, the fact that each executable program is self-contained means that it can be contained on and run from a floppy disk. The whole program can therefore be contained on a number (albeit large) of floppy disks. Although the management module does not currently have the appropriate procedures to instruct the user in the use of the appropriate disks, this is not difficult to add.

7.7.3.2.2 Speed

The assertion is made in Section 7.7.3.1.1 above that the program is slow as a result of the reading and writing procedures, and the loading of the programs themselves. While this is undoubtedly true, as suggested in Section 4.5 and above in this chapter most of the design process does not require attention from the user during its progress. Thus if the program is used on a very slow computer, the design process could be undertaken overnight, or during some other occasion when the user may be concerned with other activities.

The modular approach also allows the user to run just one or two modules to investigate an idea independently of the management module, or of the design process itself. This is of course conditional on the availability of the appropriate data in the appropriate form for the operation of the modules concerned. In this way, the user can obtain very quickly an indication of the results of particular constraints or design choices.

7.7.3.2.3 Flexibility

The modular approach to the design of the computer program allows (as has been stated above) a degree of flexibility for the design process which would be difficult to achieve by other means. Although it is of course possible to construct a large program which could be as flexible, it would be large, and problems may occur when some of the data files become large.

Subject to the necessary data being available at the time in the correct form, the programs may be run independently of the management module and therefore individually or in any order. Thus the model can be used to design particular bus stops, or to see what the design would be if the choice of cross-sectional position were restricted.

7.7.3.2.4 Modifications

The fourth advantage of the modular approach as represented in this program to be discussed in this section is the ease with which modifications may be made to the program. Modifications can take two forms: the adaptation or alteration of an existing module, and the addition of particular design models.

Adaptations to existing modules can be accommodated easily since this can be achieved independently of the other modules. As long as the input and output data conform to the standard formats, it should be possible to alter any module without the difficulties of interactions between particular procedures. This avoids any difficulties which may arise as a result of the desire to alter a procedure for one design type only, where this procedure is used by other design types.

In order to expand the range of the program in terms of the design types available to the user, it is necessary only to add the additional module(s) in the form of executable programs. These can be produced independently of the existing programs. The only necessity is that the requisite data is available to the new program, and that its outputs are included in any subsequent decisions which may be affected.

7.7.3.2.5 Language

Finally, because each executable program is called uniquely and is functionally independent from all the others, the language used in the construction of the programs is not restricted. In some cases, for example, it may be preferable to write some procedures in C and others in PROLOG. Although this is technically possible, there is always a risk of incompatibility where this is done within one program. As a result it is possible to write programs to undertake design processes in the most appropriate computer language, irrespective of its ability to link directly with any other.

One effect of this which has potential implications is that it should be possible, by adding the appropriate lines to the management module, to include other software packages in the design process. For this, it is necessary to have a program to convert the relevant data into a suitable form for the particular package.

7.8 COMMENTS

Some important factors have not been included in the program, for example the effects of general traffic or passenger behaviour on the design. In the former case, the effect of general traffic on the design is a result of the changes in capacity (especially at junctions) which arise as a result of the change in allocation of the road space. In the latter case, passenger behaviour (particularly with respect to changes in accessibility) is likely to change as a result of the new design. This effect is not represented in the model and should be investigated further. Of particular importance in this respect are the elasticities which describe the relationships between frequency and demand, and accessibility and demand. The design model presented here is intended to produce a preliminary design upon which such effects can be tested. The analysis of these effects is clearly a matter for further research.

CHAPTER 8
VALIDATION

8.1 INTRODUCTION

Chapter 7 has considered the construction of a computer model which can produce an initial design for a high capacity bus system in a corridor. This is derived on the basis of the techniques available for the design of high capacity bus systems based on the knowledge acquired using the techniques described in Chapter 6 and discussed in Chapter 3. The computer model of the design process has been developed using the methods available for the representation of this knowledge described in Chapter 4, and the opinion model derived in Chapter 5. It is now necessary to check that the computer model is a reasonable representation of the design process.

This chapter therefore sets out to describe the validation of the computer model. The validation is performed by using data available concerning a project for a high capacity bus system which has been designed and implemented by one of the teams of experts consulted during this research. The project concerned is the design of the high capacity bus system in the Santo Amaro corridor in Sao Paulo, Brazil.

This chapter is divided into three sections. First the methodology used in this validation exercise is explained. Section 8.3 considers the use of the model to generate a design for a bus system in this corridor. Finally, Section 8.4 draws some conclusions about the validation exercise.

8.2 METHODOLOGY

This section considers the methods to be used in order to validate the model. The Section is in two parts: the methodology for analyzing the actual corridor is considered in Section 8.2.1, and the methodology for the use of the model is described in Section 8.2.2.

8.2.1 Analysis of the corridor

This section is divided into two parts. The first considers the data relevant to the situation before the project implementation, and the second discusses the data relevant to the results of the implementation.

There is a particular problem with this exercise which should be considered before the validation process itself is considered. This concerns the fact that the experts were consulted initially just after the project had been implemented. Since other acquisition sessions were conducted later, there is a possibility that the knowledge acquired from the experts may have been affected by the results of this implementation. For this reason, it is possible that the design produced by the model may differ from that actually implemented. The extent and any implications of such differences are discussed in Section 8.4 below.

8.2.1.1 'Before' data

The first exercise to be completed in order to validate the model is to obtain the relevant data concerning the corridor to be used for the validation process. In this case, this is achieved fairly simply because there exists the report containing the original proposal (CET 1984), which contains a large part of the details necessary for the design of a high capacity bus system. Reference to this report has already been made in Chapter 3 above, and its significance for this exercise is that it was compiled by one of the design teams consulted during this research.

The report reveals a large amount of the data regarding the physical layout and constraints for the corridor, as well as details of the bus, traffic and passenger flows observed during the feasibility analysis conducted for the project. These data are used as inputs to the computer model in order to ascertain whether the model, using the expertise acquired with the techniques described in Chapter 6, will produce a design similar to that produced by the experts concerned.

Data available from the CET report include the physical details of the various parts of the corridor: the width of the roadway, the number of lanes and carriageways and the type of land use. Data concerning commercial speeds and the flow of passengers, buses and other traffic are also contained in the report and are used as inputs to the model. These data are illustrated in Appendix A.

These data are not, however, necessarily presented in the form which is appropriate for the model and some changes in form must be made in order to use the model. These are discussed in Section 8.2.2. The data which are used in the validation exercise are given in Appendix B.

8.2.1.2 'After' data

The CET report used for the pre-implementation data also contains the proposed physical changes to the corridor: the location of bus stops, their design and the design of the operating system. This report is therefore used to check these elements of the design process. In some cases, sufficient data are not contained in the report and these have had to be confirmed by visual inspection and measurement which have been performed specifically for this research.

Data referring to the performance of this bus system have not yet been recorded by the authorities in Brazil.

8.2.2 The use of the model

This section is divided into three sections. The way in which the model is used in order to produce a design for validation is discussed in Section 8.2.2.1. The way in which the data (obtained as described in Section 8.2.1) are prepared for use in the model is discussed in Section 8.2.2.2. Finally, the methods for the comparison with the actual project are considered in Section 8.2.2.3.

8.2.2.1 Methodology

The model is used in the validation process in order to produce a design for the Santo Amaro corridor using the data which apply to the corridor before the implementation of the high capacity bus system which is now in operation. Thus the methodology is to enter relevant data which describe the pre-implementation conditions and to use the expertise obtained using the methods described in Chapter 6 from the experts who designed the Santo Amaro project in order to produce some designs. The resulting designs are then compared with the actual project in order to see whether they correspond with the actual implementation.

Two aspects of this methodology deserve close scrutiny: the preparation of the data, and the comparison with the actual implementation. These are discussed in the following two sections.

8.2.2.2 Data preparation

In order to use the model with the available data concerning the Santo Amaro corridor, it is first necessary to consider the form in which the data are presented. The form of the data in the CET report is not necessarily suitable for use with the model. The suitability of data is initially dependent upon the type of data in question. The model expects single values for physical data such as the road widths and link lengths, and the preparation of these data is considered in Section 8.2.2.2.1. Section 8.2.2.2.2 then considers the more complex issues surrounding the preparation of imprecise data such as those describing flows.

8.2.2.2.1 Physical data

The physical data contained in the report consists of the width available for the proposed design (including footways and central reservations (where these exist)), and the lengths of particular links. It neither includes details of the junction designs along the corridor nor the permitted turning movements at these junctions. The data as recorded in the CET report is not exactly ideal for direct input to the computer model, however, and some preparation is necessary.

The first type of physical data considered here is the width of the corridor. The difficulty with using the data from the CET report in this case is that the model assumes a less detailed knowledge of the physical constraints of the corridor than is available in the report. The details of the available widths are contained in the CET report for each stretch of the corridor between each junction. The computer model requires only some estimate of this width for each section of corridor between consecutive important junctions. Thus the complex data in the report must be simplified. One strategy for this is to take the narrowest width which occurs as this is the worst possible case. This may not present a fair picture of the corridor, however, because the narrowest width may exist for only a small part of the stretch in question. To reduce the possibility of presenting a false representation of the constraints to the computer model, a number of widths are tried for each stretch: the narrowest, the widest and the width which could describe the majority of the stretch. Where the width of particular sections of a stretch are likely to cause a problem, this can then be identified by the results of these attempts.

The second issue about the physical data concerns the design of the junctions in the corridor. Junction design is not detailed in the report, although there are indications about the designs of the main junctions. These have been used in order to establish the junction design for the program. Junction design is not a major feature of this model, and is restricted to the changes that arise as a result of the implementation of a bus system along the corridor. Thus no consideration is taken of any changes which may occur to arms of the junction which do not lie along the axis of the corridor.

The third area to be discussed is the question of link lengths. These are defined to be the distance between adjacent junctions. The problem here is that the data in the report contain information about every junction in the corridor, however minor. Consideration of the state of the system as it is implemented shows that minor junctions are closed - at least to some turning movements - as a result of the implementation of the bus system. Nevertheless, it is not clear from the CET report which junctions are considered to be minor in this sense.

The design team, in considering whether a junction may be closed, take into consideration the potential effects on the surrounding network of the closure of a junction. The problem with modelling this process is that it requires a large amount of detail about the surrounding network in order to make the decisions about which junctions are actually minor, and which turning movements may be sacrificed. For these reasons these decisions are not represented in the model.

The preference is to allow the user to try out different strategies for junction closure. This is achieved by redefining the links to provide different lengths between adjacent junctions. For the validation exercise, the link lengths may be defined in two ways. The first is to choose link lengths which are defined by a reasonably constant width along the link. This eases the problem described above about the differences in link widths. The second definition is to keep the junction structure as it has been implemented. This has the disadvantage that the design team's decisions in this respect are not tested, but since the model does not attempt to represent these decisions, there is no real need to validate this part of the process. Accordingly, in this validation exercise the link lengths are chosen according to the current design.

Bus stop location is loosely described in the report, largely because the pre-implementation bus stops were loosely defined in practice. Information is given that a particular bus stop is located between two specified junctions. The junctions used for this purpose in the CET report include the minor junctions. Since, for the reasons discussed in the preceding paragraph, the model only works with specified major junctions, bus stop location is converted to a location between two major junctions. Thus more than one bus stop may be located between two adjacent major junctions. Bus stops which are located near to particular attractors (such as a metro interchange) can be fixed to the current location by the user, and this practice is used in the validation exercise where this is desirable. The computer model then attempts to calculate an even spacing between these stops as a base approximation.

8.2.2.2.2 Variable data

The term 'variable data' is taken to refer to data which are not known precisely, such as commercial speeds and flows. The data describing these speeds and flows, as presented in the CET report, are not suitable for direct entry to the model. This is because the data are presented as single values. These numbers are the averages of a number of observations and constitute the values used in the calculations performed by the design team during the design process.

Where the data are describing attributes which are not known precisely, the model uses distributions of opinion of the data values in preference to a single value. The reasons for this are discussed in Chapter 5 above. For this reason it is necessary to have some estimate of the precision of the values in order to derive a suitable opinion distribution. Such estimates are not available from the CET report and therefore it is necessary to consider the level of precision of the data.

The opinion model derived in Chapter 5 does, however, determine the thresholds at which decisions would change if imprecisely-known data were actually to take particular values. Thus it is possible, for example, to assume that the opinion distribution has a wide variance, to leave the model to identify the important thresholds and to identify the effects of traversing them. In order to generate the opinion distribution, however, it is necessary to know the minimum and maximum values of the variable concerned and the computer model described in Chapter 7 accordingly expects inputs of this type. Since this validation exercise uses data which were not collected specifically for this model, it is necessary to approximate the imprecision. The method used for this is now discussed briefly using as an example the number of boarding passengers per bus at a bus stop.

Clearly the number of boarding passengers is never less than 0, so this could provide a reasonable minimum value. At the other extreme, the number of boarding passengers cannot exceed the absolute capacity of the vehicle. This could be used as the maximum value if no further information is to hand. It is extremely unlikely, however, that the number of boarding passengers would normally attain this value at ordinary on-street stops and therefore it could be argued that a lower value could be used.

The choice of any particular value carries an element of assumed knowledge, and therefore this number should be chosen so that the level of assumed knowledge is minimized. In this validation process, the maxima have been assumed to be derived from a probability distribution where the mean is the average value and the coefficient of variation is one third of the mean value. Thus the maximum value is taken to be twice the mean:-

$$\begin{aligned} \sigma &= \frac{\mu}{3} \\ \text{max}\mathit{value} &= \mu + 3.\sigma \\ \therefore &= \mu + \mu \\ \therefore &= 2.\mu \end{aligned} \tag{8.1}$$

The next value to determine is the mode of the distribution. With no more information than the data in the report, the mode is assumed to be equal to the published data value.

Finally, as discussed in Chapter 5, it is also necessary to have an opinion value for another variable value. As a default, the model presents the opinion distribution using the minimum value for this variable value, and a value of zero for the opinion value.

In this way, the data inputs for the number of boarding passengers per bus at a bus stop comprise:-

(i) a minimum variable value of zero;

(ii) a maximum variable value of twice the mean;

(iii) a modal variable value equal to the mean;

(iv) a variable value equal to the minimum value in (i) above; and

(v) an opinion value of the variable value defined in (iv) of zero.

8.2.2.3 Comparison with the Santo Amaro project

Some aspects of the design implemented in the Santo Amaro project are available in the CET report. These include the design of bus stops and the operating system. Bus stop location is indicated on maps, but is not otherwise defined in the report. The actual bus stop locations have been checked for this research. The design produced by the computer model can therefore be compared with these data in order to establish any similarities.

Performance aspects of the implemented design, however, are not included in the report as these would require observation of the implemented scheme and the CET report was produced before the scheme was constructed.

Since the computer model is only concerned with the generation of an initial design for the corridor, it is not concerned with the performance of this design. The question for the validation exercise is to ascertain whether the computer model, as described in Chapter 7, produces a design which is similar to the one produced in reality from the same data. The test is therefore whether this similarity is sufficient for the conclusion to be drawn that the model is a reasonable representation of the design decision process.

Since the computer model uses explicitly imprecise data, it is possible that the design produced by the model may differ from the one finally implemented. It is therefore important to establish some measure of similarity so that an estimate may be made of the degree to which the designs are similar.

8.2.2.3.1 Similarity

This section describes the Similarity Measure used in the validation exercise.

In order to calculate the Similarity Measure, it is necessary to obtain a suitable measure for the design for each relevant parameter obtained from each source. These values, denoted Q_{Dp}, are obtained as described below in this Section for each parameter. The subscript D indicates the source of the design under consideration and takes the value M when this is the model, and the value E when the design source is the expert. The parameter index, p, is given for each parameter during the subsequent paragraphs.

A measure of comparison is denoted in this chapter by C_p, where p is the parameter index which identifies the parameter under consideration. C_p is calculated according to Equation 8.2:-

$$C_p = \left| Q_{M_p} - Q_{E_p} \right| \tag{8.2}$$

The value of C_p indicates the degree to which the design for the particular parameter produced by the computer model approximates to that produced by the design team. A Relative Comparison Measure, R_p, is used to indicate the level of comparison of each parameter design as a function of the expert's design. R_p is calculated from C_p by dividing the value of C_p by the value of Q_{Ep}:-

$$R_p = \frac{C_p}{Q_{E_p}} \tag{8.3}$$

Thus a value between zero and one is obtained which indicates the degree to which the design produced by the model is different from that produced by the expert.

In order to have an overall measure of comparison, the values of R_p are added. As the absolute value is used in the calculation of C_p, the cancellation of differences in different categories in the calculation of the overall measure of comparison is avoided. The overall Similarity Measure, S, is then calculated by subtracting the sum of all the values obtained for R_p divided by the number of parameters being assessed from one, in order to obtain an overall average measure for similarity.

$$S = 1 - \left(\frac{\sum_{i=1}^{p} R_i}{N} \right) \tag{8.4}$$

where N is the number of parameters to be assessed (in the case of this validation exercise N = 10).

The design model produces a design on three levels: the system characteristics; bus stop location and design and the general street infrastructure. It seems appropriate therefore that the Similarity Measure defined in this section should evaluate the design at these levels, and in addition at the overall level. Each parameter to be tested for similarity is identified by a subscript index, the value of which is indicated for each parameter during the following discussion.

8.2.2.3.2 System characteristics

The system characteristic to be compared is the choice of position of the bus system. As discussed in Chapter 3 and also in Chapter 6 above, this is a binary choice and can take only one of two values: median or kerbside. This characteristic also applies to the system as a whole: it is not possible for the bus system to have a different position at different points along the corridor. For this reason, it is sufficient to have a binary representation of the chosen operating system. In this case, the parameter value, Q_{D1}, takes a value of zero when allocated to the choice of a median position, and a value of one when referring to the kerbside.

This value is applied to each design: that produced by the model and that described in the CET report. The Comparison Measure is calculated by Equation 8.2 by subtracting the value allocated to the expert's design (Q_{E1}) from that of the model design (Q_{M1}) and taking the absolute value. If the two designs are the same in this respect, the resulting value for C_1 is zero. If the designs are different in this respect, C_1 is 1.

8.2.2.3.3 Bus stop location and design

The location of bus stops is assessed by the mean stop distance over the corridor. This allows perturbations from the optimal distance to be reduced in their effect on the comparison. The parameter index used for the comparison of stop distances is 2, and therefore the Parameter values are Q_{M2} and Q_{E2}, and the Similarity Measure is C_2.

In this case the value of Q_{D2} is the mean stop distance over all the corridor. For this reason, the values of Q_{D2} and therefore C_2 are not restricted to zero or one. Nevertheless, as with C_1, the smaller the value of C_2, the closer is the approximation of the model design to that of the experts.

Bus stop design, on the other hand has four characteristics which may differ between the model design and the expert design. Bus stop design is denoted by the parameter index 3_j, where j is given by the parameter subindex which identifies each characteristic. These are:-

1 the number of overtaking lanes at the bus stop;

2 the number of bus groups using the bus stop;

3 the length of the bus stop platforms; and

4 the number of berths.

Since the comparison process measures differences between the two designs, the corresponding values for each of these categories can be compared. Thus Q_{D31} is given by the number of overtaking lanes at the bus stop, Q_{D32} is given by the number of groups, Q_{D33} is the length of the stop platform, and Q_{D34} is the number of berths at the stop.

In order to calculate the value of C_{3j}, it is necessary first to compute all the values of Q_{D3j}, and then to evaluate C_3:-

$$C_3 = \sum_{j=1}^{4} \left| Q_{M_{ij}} - Q_{E_{ij}} \right| \tag{8.5}$$

Once again, the closer the approximation, the smaller the value of C_3.

8.2.2.3.4 General infrastructure

Four characteristics of the general infrastructure are of interest when comparing the design of a bus stop produced by the model with that produced by an expert. These are:-

1 the number of traffic lanes at the bus stop;

2 the width of these traffic lanes;

3 the width of the bus lanes; and

4 the width of the footways in the bus stop area.

As with the design characteristics discussed in Section 8.2.2.3.3, these characteristics are numbered using two parameter indices. Thus Q_{D41} is the number of traffic lanes at the stop, Q_{D42} denotes the width of the traffic lanes, Q_{D43} the width of the bus lanes and Q_{D44} the width of the footways in the stop area. The calculation of C_4 is equivalent to that for C_3 described in Section 8.2.2.3.3.

8.2.2.3.5 Overall comparison

The overall comparison between the model and the experts is calculated as the sum of all the values of R_p (Equation 8.2). While the Comparison Measures (both absolute and relative) indicate similarity by the smallness of their value, the overall Similarity Measure, S, increases in magnitude as the overall similarity between the model and expert designs increases. Since absolute values are obtained for C_p, R_p and S are both by definition within the range zero to one.

As an indicator for the similarity between the designs from the two sources, the hypothesis that the computer design is similar to the expert's design is accepted if the overall measure S is greater than or equal to 0.9.

8.2.2.3.6 Decision factors

As described in Chapter 5, the opinion model records the decision set for each decision which is based on a qualitative assessment. The chosen preferred action is, of course, the one with the highest opinion value. It is, however, informative to assess the degree to which this value is clearly higher than the next opinion value. This is even more important when there is a choice of actions depending on the actual sub-range of the data which applies.

To perform this assessment, the opinion value of the preferred action is divided by the next highest value. The result is called the Decision Factor, and indicates the relationship between the opinion values of the two actions. If the data are such that there is a choice of available actions, the opinion value of the preferred action is divided by that of the alternative choices.

The Decision Factor (*DF*) indicates the degree of confidence which may be held about the chosen action. If the value of DF is near to one, this suggests that there is a feeling of ambivalence about the selection. In this situation, the two opinion values must be almost equal, implying that the decision is being taken on the basis of data values which are very near to a decision threshold. Therefore, there is a certain ambivalence about the actual chosen action: a slight change in the data values could alter the choice. On the other hand, if the value of *DF* is very high, it is extremely unlikely that a threshold is near, and therefore only a considerable change in the data would cause a change in the decision.

For a decision to be unequivocal, therefore, the value of *DF* should be clearly in excess of one. A value of 2, for example, suggests an opinion value for the second choice of only half that for the first choice, and could be considered to be unequivocal. Values for *DF* of less than two suggest some equivocation, and this increases as the value tends towards one.

8.3 THE APPLICATION OF THE MODEL

This section considers the results of the application of the model to the data for the Santo Amaro corridor. These results are considered in the three groups identified in Section 8.2.2.3 above. This section is therefore divided into four further sections. The first three of these consider in turn the comparison of the system characteristics, the location and design of the bus stops and the design of the general infrastructure. The fourth considers the overall comparison.

8.3.1 System characteristics

As discussed above in Section 8.2.2.3.1, the system characteristic to be compared in this validation exercise is the choice of cross-sectional position of the proposed bus system. This section discusses the results of the application of data derived as described above in Section 8.2.2.2.2 to the computer model.

As discussed in Chapter 3 and represented in the computer model as described in Chapter 7, the data required for this are concerned with the levels of parking and enforcement. As the corridor passes through some areas of shopping, it is felt that the parking level is likely to be quite high. When asked about this, the design team suggested that both parking and enforcement were difficult before the implementation of the bus system, and therefore this was a major consideration in their choice of cross-sectional position for the system.

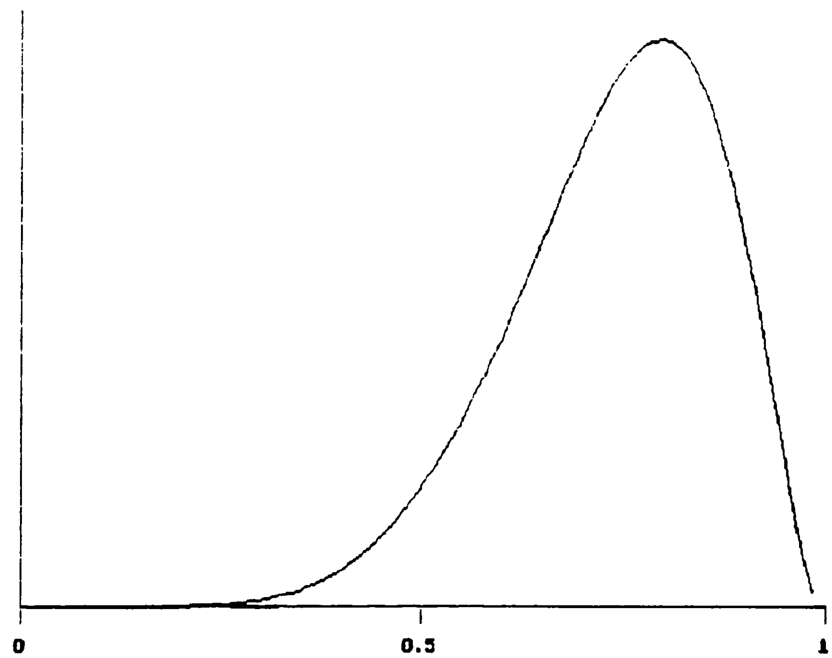


Figure 8.1 Opinion distribution of the level of illegal parking in the Santo Amaro corridor prior to the implementation of the high capacity bus system.

Table 8.1 Decision set for the system characteristics

PARKING					
data range	*very low*	*fairly low*	*fairly high*	*very high*	**decision factor (DF)**
1	0.12	0.57			1.70
2			0.97		1
3				0.62	1.56
overall	0.12	0.57	0.97	0.62	1.56
ENFORCEMENT					
1	0.97				1
2		0.59	0.14	0.08	1.64
overall	0.97	0.59	0.14	0.08	1.64

Of course, this is not quantified in the CET report, but the experts produced opinion distributions which represented their opinion of the parking and enforcement levels before implementation. These are shown in Figures 8.1 and 8.2. These distributions were then applied to the relevant expertise by the model and the resulting decision set is given in Table 8.1.

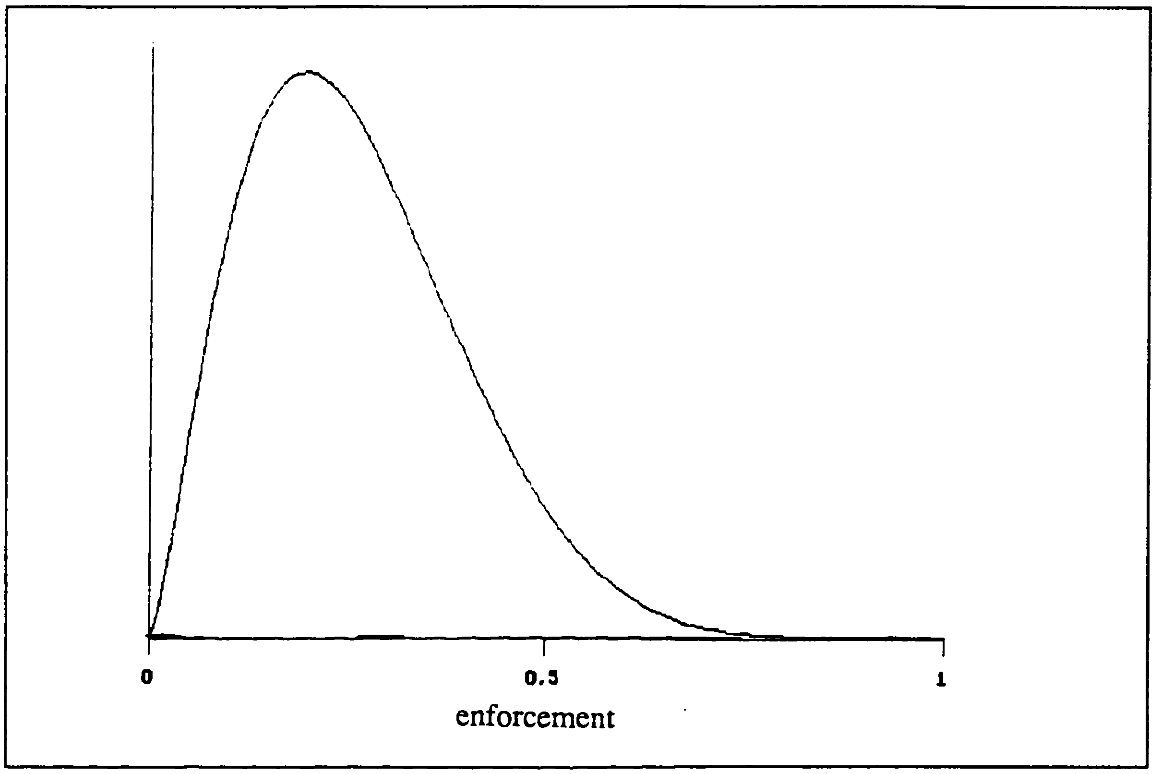


Figure 8.2 Opinion model of the enforcement level for parking in the Santo Amaro corridor prior to the implementation of the high capacity bus system.

From this it can be seen that the option with the highest opinion value for the land use parameter is for the descriptor *'fairly high'*. That of the enforcement parameter is *'low'*. Using the logical formula discussed in Chapter 6, the decision taken by the computer model is therefore to install a bus system in the median. Therefore the value of the Parameter value Q_{M1} is 0.

This is also the decision of the experts in this case. The value of Q_{E1} is therefore also zero and the Comparison Measure (the difference between the two parameter values) is equal to zero.

8.3.2 Bus stop location and design

The parameter used to compare the location of the bus stops is the average stop distance over all the corridor. This is given by the model as 287.35 metres. The average stop distance over the corridor in the implemented design is 500 metres.

The Comparison Measure, C_2, is therefore 212.65. This appears to be rather large, and it is useful to consider the reasons why this difference is so great. The stop location model used by the computer model is derived from a model in current use by one of the design experts. Since the time of the design of the Santo Amaro corridor (and possibly as a result of the implementation) the view of this expert has changed about the optimum stop distance. He now considers that it is better to place the stops closer together than in the Santo Amaro implementation in order to give better access for the passengers.

The bus stop design is compared over the stops as a whole. Since the number of bus stops is different in the two designs, it is not possible to compare the designs for particular stops. It is, however, possible to compare the designs which occur within the same physical constraints. These constraints are defined by the available width in the links and junctions.

The first set of design characteristics to consider are identified in Section 8.2.2.3.3. These are the number of overtaking lanes at the stop, the number of bus groups, the length of the platforms and the number of berths for each group. The design selected by the computer model is that of a double group, double berth bus stop with one overtaking lane in each direction. The platform length is 74 metres.

The bus stop design actually implemented in this corridor has one overtaking lane per direction, two bus groups, a bus stop platform of 75 metres and two berths allocated to each bus group. Two lanes are allocated to general traffic. Thus the values of Q_{E3j} are equal to those for Q_{M3j}, with the exception of Q_{M33}, which differs by 1. The Comparison Measures are therefore equal to zero except for C_{33} which takes the value of one.

The effectiveness of this bus stop design in the given circumstances is indicated by the degree of saturation, and the important aspect is the associated opinion of the expertise. This is given in Table 8.2, together with the decision factors, which are calculated as described in Section 8.1.2.3.6 above. The descriptor with the highest opinion value is '*fairly low*'. The decision factor is high ($DF = 2.33$) which indicates that this opinion value is obtained from data values which are not in the proximity of a decision threshold. The value in the third column is, however, derived from a different sub-range of the data values, and therefore if the degree of saturation were to be within this sub-range this would be considered to be '*fairly high*' and the recommendation would be to increase the capacity of the bus stop by changing the design.

Table 8.2 Decision set for the degree of saturation at a bus stop in Link 1.

SATURATION					
data range	*very low*	*fairly low*	*fairly high*	*very high*	**decision factor (DF)**
1	0.16	0.93			5.81
2			0.40	0.26	2.33
overall	0.16	0.93	0.40	0.26	2.33

Table 8.3 Decision sets for the widths of traffic lanes, bus lanes and footways in and around bus stop areas.

ELEMENT WIDTHS					
parameters	*too narrow*	*narrow but possible*	*satisfactory*	*too wide*	**decision factor (DF)**
traffic lanes	0.55	0.99	0.61	0.42	1.62
footways	0.20	0.56	0.79	0.19	1.41
bus lanes	0.25	0.02	1.00	0.48	2.08

The last characteristics to consider are those identified in Section 8.2.2.3.4, which are concerned with the general infrastructure around the bus stop areas. These characteristics are the number of traffic lanes at the stop, Q_{D41}, the width of the traffic lanes, Q_{D42}, the width of the bus lanes, Q_{D43}, and the width of the footways in the bus stop area, Q_{D44}.

8.3.3 General infrastructure

Although the design can be accommodated in this link without losing the overtaking lanes (and therefore without a capacity loss to the system), the traffic lanes and footways, as can be seen from Table 8.4, are not necessarily at their preferred widths. It is first necessary to calculate Q_{M4j} for each characteristic so that the Comparison Measure can be calculated. Taking, for example, the bus stops constrained by the available width in Link 2, the model produces a design which has two traffic lanes per direction. These lanes have a width of 3.11 metres. The bus lanes are 3.5 metres wide and the footways are allocated 1.055 metres. These values are compared with the measured values for stops built within the same width constraints in Table 8.4.

It is therefore necessary to establish the extent to which the expertise is satisfied with the resulting widths. This requires an analysis of the opinion values in the final decision sets. These are given in Table 8.3, along with the decision set for the bus lane width.

The decision factors in the last column indicate that the expert is unambiguous about the choice of bus lane width. The opinion value in the decision set for '*satisfactory*' is more than double that of the descriptor with the next highest opinion value (DF = 2.08). The traffic lane width is less satisfactory to the expert. In fact the descriptor with the highest opinion value is '*not satisfactory, but possible*'. The decision factor in the last column suggests that the expert is concerned that the traffic lanes are becoming a little narrow: the next highest opinion value is for the descriptor '*satisfactory*', and the DF is 1.62. This indicates that the expert may consider lanes which are even narrower before rejecting the design of the bus stop by reducing the number of traffic lanes. The width of the footways is also a matter of concern to the expert, but not to the extent that the bus stop design should be changed. Here, the descriptor with the highest opinion value is '*satisfactory*'. Some ambivalence is indicated, however, by the fact that the descriptor with the next highest opinion value is '*not satisfactory, but possible*', with a DF of only 1.41.

The bus stop design implemented by the experts in this part of the Santo Amaro corridor is also of this form. The bus stops are staggered, with overtaking lanes and two traffic lanes. Often, the traffic lanes and footways are narrower than the ideal (sometimes even narrower than in this computer design). The bus lanes are not less than the ideal width in the corridor (although for one stretch in which the corridor passes through a tunnel, the bus lanes are suspended).

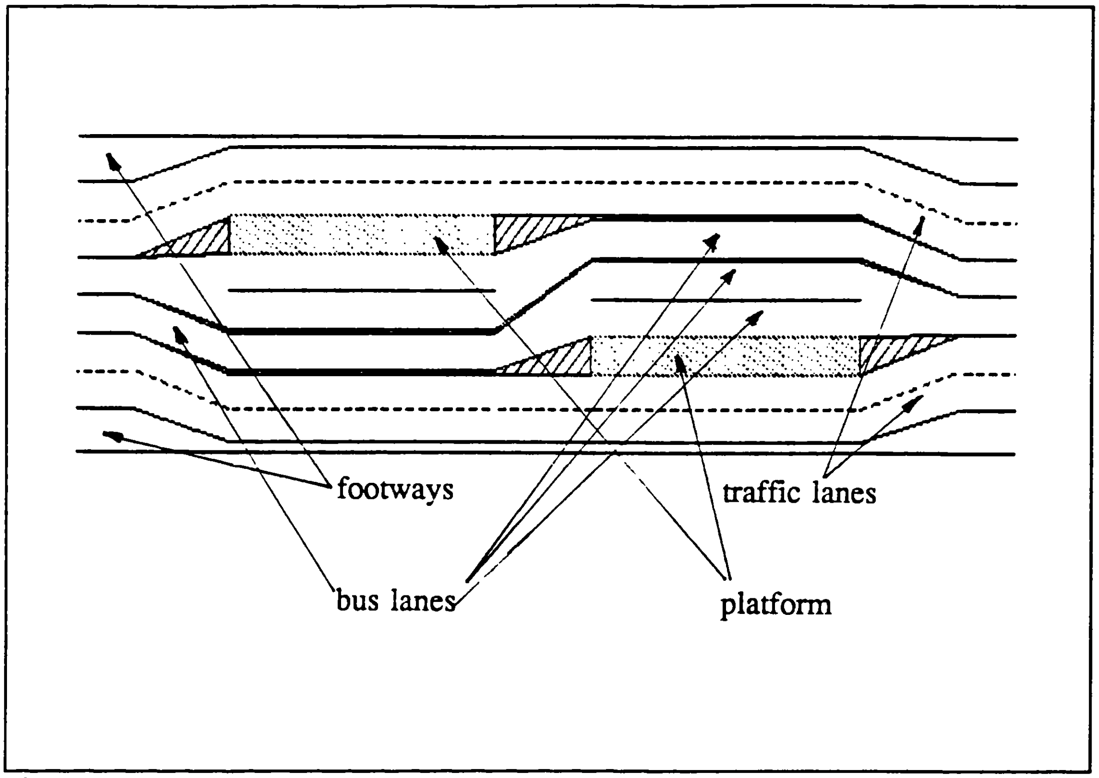


Figure 8.3 A staggered bus stop in the median in Link 1.

Figure 8.3 shows the computer design for a bus stop designed within these constraints. This should be compared with Figure 8.4, which shows two photographs of a bus stop in the Santo Amaro corridor, showing clearly the staggered design, with the overtaking lane for each direction and the number of traffic lanes reduced to two for the duration of the bus stop. The two bus groups, each with two allocated berths may also be seen.

8.3.4 Overall comparison

The values of the parameters, the Comparison Measures and their relative values are given in Table 8.4. This table also gives the value for the Similarity Measure obtained from this validation exercise. In this case, the Similarity Measure is 0.908. Since the rule for acceptance that the model has produced a design which is similar to that of the experts is that the Similarity measure is greater than or equal to 0.9, the hypothesis may be accepted.

The most dissimilar parameter is the stop distance, which accounts for nearly half (46.3%) of the balance (referred to as the dissimilarity) between the Similarity Measure and a perfect comparison (in which the value of S would be 1). The difference in the traffic lane widths

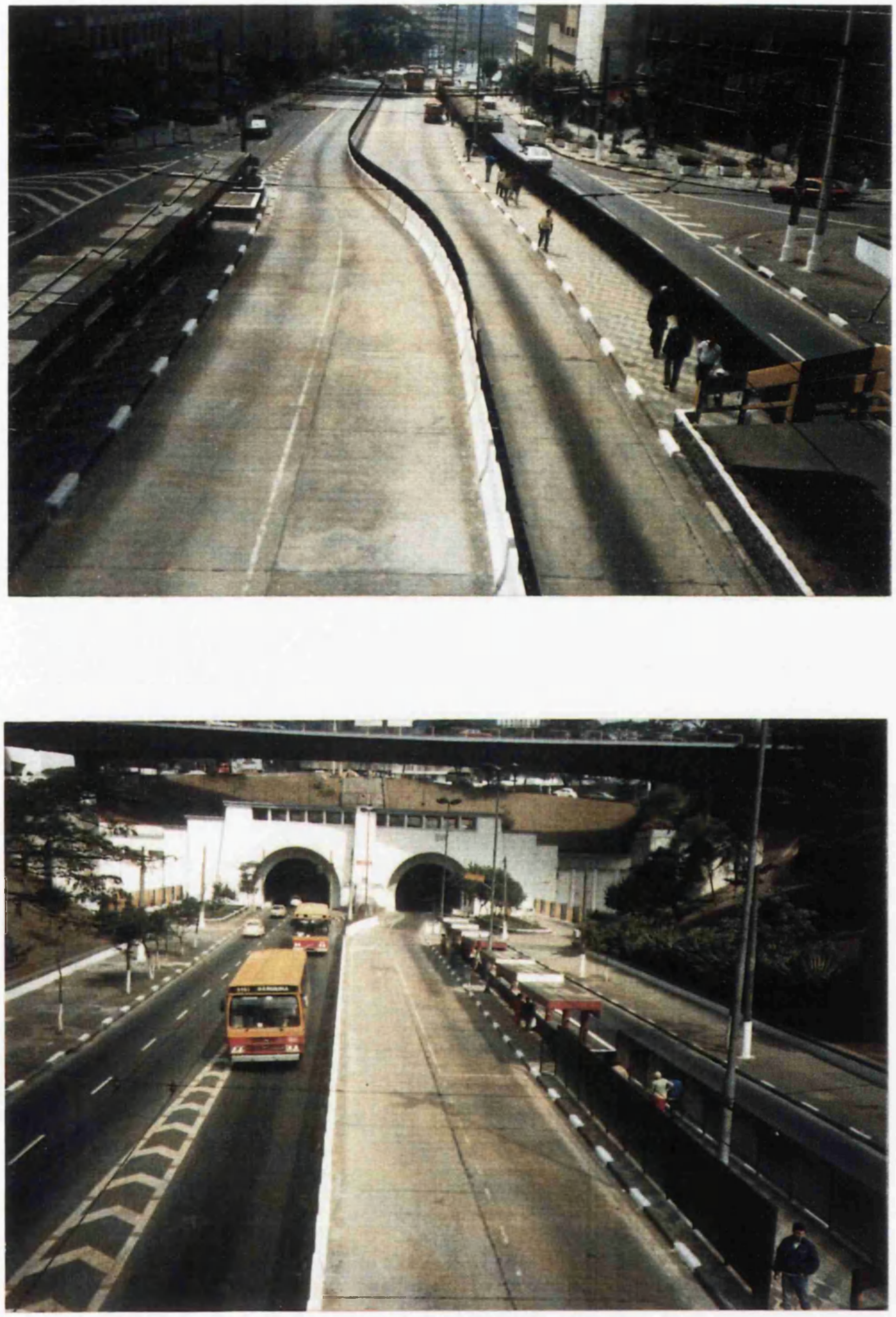


Figure 8.4 Two views of a bus stop in the Santo Amaro corridor. These may be compared with the computer design shown in Figure 8.3. The views are taken from the same point, but in opposite directions, and show the two sides of the bus stop.

Table 8.4 Comparison measures and the Similarity measure for this validation exercise.

Parameter index	Value (model)	Value (expert)	Comparison measure	Relative comparison
1	0	0	0	0
2	287.35	500	212.65	0.426
31	1	1	0	0
32	2	2	0	0
33	74	75	1	0.013
34	2	2	0	0
41	2	2	0	0
42	3.11	2.8	0.31	0.111
43	3.5	3.5	0	0
44	1.055	1.675	0.62	0.370
Total comparison				0.92
Similarity measure				**0.908**

accounts for 12.1% of the dissimilarity, and the variation in the design width of the footways accounts for 40.2%. The minor difference between the platform lengths accounts for the remainder.

Since the Similarity Measure is intended to record and evaluate differences, no account is taken of any logical connection between the parameters concerned. A result of this is that the Similarity Measure may be lower than would otherwise be the case. This applies, for example, to the consideration of the traffic lane and footway widths. A decision to accept a particular width for one of these parameters necessarily implies the acceptance of another particular width of the other parameter. Thus it could be argued that the fact that the traffic lanes are wider in the model design than in the expert design implies that the footway width is necessarily narrower. Thus the use of absolute values counts against the model. Nevertheless, the use of absolute values is supported for two reasons.

First, the Similarity Measure should record all differences, whether these result from a positive or negative difference between the two designs. The reason for this is that it is important to know the degree to which the model is capable of reproducing the expert's design practice. In this case, an underestimate in one parameter is as important as an overestimate in another.

Second, the comparison is made between the chosen design widths. There is always a pair of footways in each design, but there could be a larger number of traffic lanes over which the available width should be allocated. Thus a small difference between traffic lane widths could force a larger difference between the footway widths. This is the case in this validation exercise, and accounts for the difference, observed above, between the proportion of the dissimilarity taken up by the two parameters concerned.

8.4 CONCLUSIONS

This validation exercise is intended to test whether the computer model is capable of representing the design process of the design team whose expertise has been used in its derivation.

In all the binary tests, the model produces the same decision as the design team.

In the measurement tests, the performance of the model is more variable. In the case of the bus stop platform length the difference between the model and the expert design is negligible. The bus lane widths are exactly the same in the two designs.

On the other hand, there is a considerable difference between the two designs over the issue of the distance between the bus stops. As discussed in Section 8.3.2 above, this is probably due to the revision by the expert concerned of his views about the optimal location of bus stops since the implementation of the design of the Santo Amaro corridor. The model contains the current expertise in this respect: the implemented design, of course, is an example of the previous thinking. This has implications for the construction of similar computer models. The experts can revise their opinions about such matters, and consequently their expertise evolves over time. A computer model, however, is frozen in this respect. Even if the computer model could learn from its own experience, this would be subject to suitable input from the expert. Such a learning process, while theoretically possible, would be time-consuming and would require a large amount of appropriate data. The revision in the expertise in this case, for example, is the result of observations that buses are spending too much time at the stops, and that a reduced stop spacing would reduce this problem. The potential of this reduction is, however, only realized after observations of the success of the overtaking facilities at the bus stops which, as discussed in Chapter 3 above, reduce the interactions between buses and subsequently their interdependence.

The difference in the traffic lane widths and footway widths is noticeable and reflects an interesting difference between the preferences of the model and the choices actually made by the experts. It appears that the experts, in practice, prefer to opt for a wider footway at the expense of narrower traffic lanes. This is not reproduced in the model, because the choice of narrow traffic lanes is not generally recommended - even by the experts whose expertise has been represented in the model. The Decision Factors recorded for these parameters indicate that the model accepts the narrow footways, but with a considerable degree of hesitation, and that there is a certain amount of room for manoeuvre for reducing the width of the traffic lanes to compensate. Such a decision is, however, subject to consideration of the particular circumstances at a particular site. The representation of such consideration is beyond the scope of such a simple model. Further work in the area of knowledge representation could produce a more sensitive appreciation of such matters.

The results of the validation test indicate that it is possible to accept the hypothesis that the model can represent the design expertise as represented using the logic, the opinion model and the various functional models as encoded in the computer model.

It is now of interest to discuss the type of designs produced by the model when given slightly different circumstances. Two situations are considered: where the available width is more than adequate, but a constraint is added with respect to the amount of the central reservation which may be used by the bus system; and where the available width is not sufficient for the design of such a system. These are discussed in the next Chapter.

CHAPTER 9
EXPERIMENTS

9.1 INTRODUCTION

The conclusions drawn from the validation exercise described in Chapter 8 show that when using the same data, the computer model is capable of producing a design which is similar to that produced by the expert whose expertise is represented within the model. The corridor used for the validation exercise is obviously suitable for a high capacity bus system: this is known because a high capacity bus system has been designed and implemented in the original corridor. It is important to establish, however, that the model is capable of reacting appropriately to other situations. This chapter discusses the situation where data are used which describe a corridor which may not be so suitable for such treatment.

Two experiments are discussed in this chapter. The first uses data made available by the Royal Borough of Kingston-upon-Thames in London, and included in Royal Borough of Kingston-upon-Thames (1990). This corridor leads into the town centre of Kingston and is distinguished by a restricted width over all of its length (by Brazilian standards). This experiment is described in Section 9.3. The second experiment uses data from Lima, Perú. The corridor under consideration in this experiment is wider than the London example, but there is a constraint imposed with respect to how the available width may be allocated. This experiment is described in Section 9.4.

There are, however, some difficulties with these experiments which affect the experimental design and assessment. Because of this, they are discussed before the experiments are described. These difficulties are the subject of Section 9.2. Some conclusions are drawn from the experimental results, and these are to be found in Section 9.5.

9.2 EXPERIMENTAL DESIGN AND ASSESSMENT

This Section discusses the conceptual difficulties which must be addressed before the experiments can be designed. These difficulties are discussed in Section 9.2.1. The design of the experiments is described in Section 9.2.2 and the assessment process is considered in Section 9.2.3.

9.2.1 Conceptual problems

The validation exercise described in Chapter 8 indicates that the representation of the expertise is sufficient to reproduce the expert's decisions and designs when given the same data and conditions. It is not possible, however, to show that when faced with different conditions, the representation of the expertise is still sufficient. This is because this type of situation is hypothetical: the expert has not been required to design a system, and therefore there is no concrete example to reveal the decisions that were made during its design.

9.2.1.1 Testing methodology

Several possibilities exist for the evaluation of such an experiment. First, the experts could be asked to provide a design when given the relevant data. Such a design - if it could be obtained - would be hypothetical, and would not necessarily be the design that would be implemented in reality.

A second possibility is to identify a project that is in the design stage, and to compare this design with that produced by the computer. Unfortunately there is no such project at a suitable stage in development for such comparison within the timescale of this research.

The results produced by the computer model consist of the design of infrastructure for a high capacity bus system which is defined by the model. There is no suggestion that the design model is capable of assessing the quality of the design through the performance of the bus system. There are two reasons for this limitation. First, the actual construction of a high capacity bus system is outside the bounds of this research, and therefore it is not possible to test the adequacy of the design against its performance in practice.

Second, the computer model developed in this research is designed to represent a decision process. The test is, therefore, whether this decision process is represented adequately. In the case of the validation exercise discussed in Chapter 8, it is shown that the design produced by the computer model is broadly similar to the design produced by the experts. This design, however, does not test the limits of the decision process: for example, where the corridor is too narrow for such a system, or where some other constraint is imposed. It is therefore essential to test that the model is capable of making reasonable decisions under these circumstances.

In the absence of an implemented system, it might be considered satisfactory to test the design with simulation. A number of problems arising from the technicalities of this approach show that this is not satisfactory. First, there is a conceptual difficulty over the use of one model to test another. A model represents a subset of the real world. The two models concerned must therefore include - and work with - a common subset of the real world. It is therefore necessary first to check that this is indeed the case. Second, it is necessary to ensure that these subsets are sufficient to represent enough of the real world to test the design without ignoring any important factors. Third, it is necessary to be certain that there are no inherent biases within the model used for the testing procedure which may affect its performance in terms of the design model. Fourth, the model used to perform the tests should include not only the same subset of the real world as the design model, but also a sufficient number of additional elements which can be used to establish a series of pseudo-objective tests. This suggests that in order to use a model to test another model it may be necessary to construct the testing model not only for this specific purpose but also for the particular design model.

Of the simulation models considered in this research, SIBULA (Lindau 1983) appears at first to be sufficient because it considers in exhaustive detail an exclusive bus lane and convoy operation over a corridor. Unfortunately, however, overtaking at bus stops is not included in its capabilities, and since this is a fundamental element of the type of design suggested during this research this would need to be included. SIBULA also has a rather simplistic view of the demand profile along the corridor, and this would detract from any results obtained from the use of this model to test the subtleties of the design process which have been addressed in this research and which attempt to design a bus system which is responsive to the particular demand profile of the corridor in question. Whilst BLAST (Lindau 1989), a derivative of SIBULA produced for microcomputers, has the facility to include some measure of different operating styles at the bus stops, the demand profiles are still based on the SIBULA system. Like SIBULA, BLAST is really a simulator for bus operation using convoys.

IRENE (Gibson *et al* 1989) treats the bus stops in much more detail, but considers them in isolation: there is no consideration about the effects of the design of one stop on the operation of another. Where a whole corridor is being redesigned with a different operational strategy and design, it is necessary to consider such effects.

The model of Oldfield Bly and Webster (1977) is less detailed than either of these models and is designed for a more macroscopic analysis of the workings of a British bus lane. Thus this only takes account of the type of priority offered by a point-measure bus lane (the details of this type of design are discussed in Chapter 3). In order to investigate the detailed and subtle design changes considered by the computer model described in Chapter 7, more detailed work is required on the effects of these changes on a whole corridor.

Clearly, therefore, if simulation were to be used to test the designs, a model would be required which could represent the details of bus operation. This model should include the ability to model results arising from subtle changes to bus stop design and operating style. In addition, it should be capable of modelling the effects of bus stops on the bus operation as it progresses along the corridor. The development of such a model is a major undertaking, and is a subject for urgent further research.

It is therefore concluded that the evaluation should refer to the representation of the decision process rather than to the performance of the (hypothetical) bus system.

9.2.1.2 Experimental objectives

Since the model developed in this research represents part of the decision process underlying the design process, it is sufficient to test its abilities to discern between possible system designs, and the alternative - to recommend that a high capacity bus system is not possible within the constraints imposed. Accordingly, three objectives are considered for these experiments.

The first objective is to design an experiment to test the performance of the decision model where the constraints which determine the feasibility of a bus system are critical. In this situation, it is important that the decision model produces the appropriate response.

The second objective is to test the ability of the decision model to recognize the extent to which it is using suboptimal values for its design elements.

The third objective is to determine the extent to which a design, even when using suboptimal values, may be considered to be similar to a design produced by the expert(s) from whom the expertise is obtained.

These objectives differ from those used in the validation exercise described in Chapter 8. The validation exercise is designed to test the representation of the expertise used by the decision model, and the decision model itself, against the performance of the human experts from whom this expertise is obtained. The experiments discussed in this chapter aim to test the performance of the decision model against the expertise. To consider how this may be done, the situation is discussed where the data are such that no constraints apply to the decision process.

In this situation, the expertise suggests that the modal values of the 'best' descriptor of each expertise distribution should be used. The 'best' descriptor is whichever one conveys the idea of '*satisfactory*' (whatever the actual wording of the descriptor), and for brevity in this discussion is referred to as the optimum.

When constraints affect the ultimate decision, by forcing the use of suboptima, the results obtained from the decision model can be compared with the optima within the expertise by assessing the degree of recognition of the effects of using such suboptima. The design of an experiment to fulfil these objectives is now considered.

9.2.2 Experimental design

9.2.2.1 Preliminary considerations

The two experiments performed in this research conform to the same experimental design. This is intended to provide the opportunity for the model to produce a design for a high capacity bus system (where possible) for a given corridor. Data describing real corridors are used and supplemented where necessary with data obtained from observation. Two experiments are therefore conducted in order to establish the way in which the model approaches two types of constraint. The first is that the available width is very restricted, and the second is that the total width has a restriction on its use.

The results of these experiments are of interest in terms of the objectives stated in Section 9.2.1.2. Thus three aspects of the results are considered. The first is the question of whether the model can detect that it is not possible to construct a high capacity bus system. The second issue is the evaluation of a design (if one is produced) in terms of the 'ideal' as expressed by the optima in the expertise. The third aspect is whether the decision process represented by the decision model has resulted in a design which, despite being suboptimal in one or more respects, is acceptable to the expertise from which it has been derived.

In the first case, the feasibility of the design must be considered. As discussed in Section 9.2.1.1, it is not desirable to discuss with an expert the results of such a hypothetical case. The decision of the model not to produce a design for a high capacity bus system (if this is the result) should also be critically assessed.

The second condition, regarding the evaluation of the resulting design, considers these points. The model, if it cannot produce a design for a high capacity bus system, should be able to ensure that whatever is produced as a design is acceptable to the expertise. The relevant criteria are essentially simple: the bus and traffic lanes should be sufficiently wide for the appropriate vehicles. Footways should also be of a reasonable width. The ways in which sufficiency and reasonableness may be evaluated are discussed in Section 9.2.3.

The first experiment is designed to test the model's ability to recognise that the physical constraints are too severe for the implementation of a high capacity bus system. In this case, the overall widths are restricted and the existing flow data are low.

The second experiment aims to test a more subtle interpretation of the data. In this experiment, the overall widths are sufficient for the implementation of a bus system, but some constraints are imposed on the way in which the overall width may be allocated. Thus the imposition by the user of constraints is tested. The design of the experiments is discussed in the following section, and the assessment procedure is discussed in Section 9.2.4. The conduct of the experiments is described in Sections 9.3 and 9.4.

9.2.3 Assessment

It is important to ensure that the assessment measures used for these experiments do not assume that more is achieved in the validation exercise than is the case. The validation exercise described in Chapter 8 shows that when handling data which are known to have resulted in the implementation of a high capacity bus system, the computer model performs satisfactorily in the sense that it reproduces, within defined limits, the infrastructure design actually implemented. In this sense it can be assumed that the expertise as represented within the system, when combined with the set of decision models within the program, is adequate for assessing the limits of the various elements of the design. Therefore it is possible to assess the performance of the model in the more restricted cases used for the experiments without further recourse to the experts.

These assessments are therefore different from those conducted during the validation exercise. Rather than assessing the design for its similarity with the implemented design, it is necessary to evaluate the design for the degree to which it is deemed satisfactory by the expertise. This requires some form of quantitative measure - such as the Similarity Measure defined in Chapter 8. As it is not possible to measure the similarity between the computer design and an implemented design, a proxy for the implementation is necessary. The proxy used for these experiments is to assume that the expert would attempt to design each element of the bus system using the preferred values. The measure used to evaluate the design is therefore the Tolerance Measure. This is the degree to which an expert is prepared to tolerate a suboptimal solution. The tolerance is computed in terms of the parameter values obtained from the computer design and the relevant optima found in the expertise base. This is likely to produce a severe test: the experts are often unable to use their preferred widths in a design for a variety of reasons.

The parameters are assessed against the theoretical optima rather than an actual design (which may also have used suboptima). It is therefore necessary to adapt the assessment measures used in the validation exercise so that they can indicate the degree to which the values used in the design differ from the optima.

In order to achieve this, the Relative Comparison Measure, *R*, used in the validation exercise is replaced by a Relative Assessment Measure, *A*. This measure is designed to reflect the fact the suboptima can be evaluated in three ways:-

1. the value of opinion indicates the degree to which a descriptor is held to be valid;

2. the magnitude of the difference between the optimum and the design value indicates the extent to which a value is different from the optimum; and

3. the relationship between the opinion value for the preferred descriptor and that for the next option indicates a degree of confidence that the preferred descriptor is correct for the given value.

The Relative Assessment Measure, A, is therefore calculated as a combination of these estimators:-

$$A = \left(\frac{\left(\frac{|D - E|}{E} \right)}{(O \, . \, DF)} \right) . \, M \tag{9.1}$$

where

D	=	the value of the design variable
E	=	the value of the optimum defined by the expertise
O	=	the opinion value associated with the value of the design variable
DF	=	the Decision Factor
M	=	a multiplier dependent on the descriptor (these examples refer to the concept of width, but the equivalent applies to other types of descriptor where relevant):-

M = 2^0 if the descriptor is '*satisfactory*';

= 2^1 if the descriptor is '*narrow but possible*' or '*too wide*'; and

= 2^2 if the descriptor is '*too narrow*';

Thus A tends to increase if the difference between the design value and the optimum increases, or the descriptor changes. Alternatively, A tends to decrease if the opinion value increases or the Decision Factor, DF, increases. The magnitude of the difference between the design value and the optimum is therefore scaled according to the strength of the opinion that this value can be described by the relevant descriptor. The opinion value is also scaled by the Decision Factor in order to reduce its effect on A if there is a degree of ambivalence about the choice of descriptor. If necessary, the fact that the descriptor applied by the decision model to the design value is different from the optimum is included to increase the significance of the measure. A perfect match between the design value and the optimum gives a value for A of zero.

The assessment for the whole design is called the Total Assessment Measure, TA, and is calculated:-

$$TA = \sum_{i=1}^{N} A_i \tag{9.2}$$

where

N = the number of parameters under consideration (in the case of these experiments, $N = 10$)

i = the parameter index

In order to assess the complete design, the Similarity Measure of the validation exercise is replaced by a Tolerance Measure, *T*. *T* represents the degree to which the average of the parameter assessments is such that the design could be accepted, or tolerated, by the expertise:-

$$T = 1 - \left(\frac{TA}{N} \right) \tag{9.3}$$

The range for T is therefore $[-\infty, 1]$. If all the design values were equal to their associated optima, A would take a value of 0, and T would take a value of 1. Therefore, the smaller the value of T, the less tolerance there is and the less acceptable the design is to the expertise. For the purposes of these experiments, the condition for an acceptable value for T is taken to be $T \geq 0.900$.

9.3 EXPERIMENT 1: LONDON

9.3.1 Introduction

The corridor from which data are used in order to test the computer model is situated between Tolworth and the town centre of Kingston-upon-Thames. the corridor is approximately 4 kilometres in length and is fairly narrow over most of its length. There is a particularly rigid restriction where the road crosses the railway line serving Surbiton station, a major attractor and generator of trips from and to Kingston Town Centre. This station is not on the corridor itself, but the corridor is one possibility for the route from the station to the town centre for a high frequency bus service.

Table 9.1 Link and Junction data describing the corridor between Kingston Town Centre and Tolworth. *Source: Royal Borough of Kingston-upon Thames (1990)*

Link	From	To	Length	Width
1	Brook Road	Maple Road	970	15
2	Maple Road	St. Mark's Hill	570	13
3	St.Mark's Hill	Kingston By-pass	2250	19

Junction	Arm	stop line		entry line		central reserve	footways
		lanes	width	lanes	width		width
1	1	3	10	3	10	5	2
1	2	3	10	3	10	3	2
1	3	3	10	3	10	3	2
2	1	2	5	2	5	0	2
2	2	2	6.5	2	6.5	0	2
2	3	1	4	1	4	0	2
2	4	2	7	2	7	0	2
3	1	2	6	2	6	0	2.25
3	2	2	5.5	2	5.5	0	2
3	3	1	4.5	1	4.5	0	2
3	4	2	5.5	2	5.5	0	2
4	1	3	7.5	3	7.5	10	6
4	2	3	7.5	3	7.5	10	6
4	3	3	9	3	9	0	2
4	4	3	9	3	9	0	2

9.3.2 Data

For the purposes of this study, the corridor is divided into three links with four associated junctions. The appropriate data are shown in Table 9.1. These data are in effect a summary of the data made available. As described in Chapter 7, the model takes a particular width for each link, and this is used to evaluate the possibilities for the link.

9.3.2.1 Physical data

As explained in Chapter 7, the links used in the experiment are chosen on the basis of the importance of the junctions, and it is highly improbable that the junctions identified in this experiment could be closed without some major inconvenience to road users. The choice of junctions also permits links to be defined in which the width is reasonably constant. Given the knowledge of different link widths within each link, it is necessary to choose a particular width for the experiment. For the purposes of this experiment, therefore, a width is chosen for each link which is appropriate for a reasonable proportion of its length, but which is not duplicated in one of the other links. This prevents any repetition of calculations, and therefore allows the effects of different link widths to be explored.

The siting of the seven bus stops in the corridor is given quite accurately in the data. As explained in Chapter 7, however, the computer model is constructed to be able to proceed with informal stops and therefore cannot locate bus stops with this degree of accuracy. Using the methods described in Chapter 7, the bus stops are located equidistantly within each link. Stop 2 is located near to the Kingston Polytechnic, and therefore it is not desirable to allow this stop to be moved. Its location is therefore fixed. No stops are located in Link 2. The stop locations used in this experiment are shown in Table 9.2. The location of each stop is estimated by the computer model in terms of the distance of the stop from the corridor start, and this information is also given in Table 9.2.

Table 9.2 Bus stop locations used by the computer model for experiment 1.

stop number	link number	distance from link start (metres)	distance from previous stop (metres)
1	1	356.67	
2	1	713.33	356.67
3	3	2173	1459.67
4	3	2563	390
5	3	2953	390
6	3	3343	390
7	3	3733	390

9.3.2.2 Variable data

Three types of variable data are used in the model: vehicle data, passenger data and speed data. These are considered in the following three sections.

9.3.2.2.1 Vehicle data

Bus flow data are determined from the various bus timetables. There is a total of 12 bus routes using all or part of this corridor. In some cases, the published timetable gives only an estimate of the frequency, in terms of a range of headways. These ranges have been used to form the ranges of the bus flow for data input. The mid point of the range has been used as the mode.

The type of buses used to operate the various routes is recorded as observed in the corridor. The most important aspects of vehicle type are the size, capacity and the number of doors. It is also important to know how many passengers may be on the vehicle as it enters the corridor. The bus flow and bus characteristics data used in this experiment are shown in Table 9.3.

Table 9.3 Bus flow data for the Kingston to Tolworth corridor.

Bus line	Links operated	Minimum flow	Maximum	Preferred	Length / capacity	Doors
72	1	1	3	2	13 / 90	2
K1	3	1	3	2	10 / 25	1
281	1,2,3	5	7	6	13 / 90	2
514/714	1	0.5	1.5	1	13 / 50	2
727A	1	0.5	1.5	1	13 / 50	2
727B	3	0.5	1.5	1	13 / 50	2
406	1,2,3	3	5	4	13 / 90	2
X5	1,2,3	0.5	1.5	1	10 / 25	1
X71	1	1	3	2	13 / 90	2
K4	1	1	3	2	10 /25	1
479A	1	1	3	2	13 / 90	2
479B	3	1	3	2	13 / 90	2

9.3.2.2.2 Passenger data

Passenger flow data are inferred from various data made available by only some of the bus companies. These data are not available from all the companies because some considered the information to be commercially sensitive. Since the corridor of interest only constitutes a part of each bus route, it is reasonable to assume that the demand characteristics are similar for all buses using the corridor.

One data source in particular gives the number of boarding passengers for each stage during a particular time period as recorded by the electronic ticket machine on each bus. Two problems exist with these data. First, most of the data is an aggregate of the boarding data for all the buses on each line in the direction concerned during the appropriate time period, and it is not possible to obtain the data for individual vehicles. Second, some of the data is only available aggregated between the two directions, which makes it impossible to analyze one particular direction. Third, the data are classified by stages (points at which the machine is updated with a new fare stage). These stages may not necessarily correspond exactly with the bus stops.

The part of the corridor under investigation has seven stages and seven bus stops. The second problem identified above does not therefore appear in this case.

The data aggregated from individual vehicles (but identifiable by direction) are divided by the range of bus flows in order to obtain a range of boarding passengers per bus per stage. This makes the assumption that the demand during the relevant time period is constant. The time period concerned covers the morning peak period, and therefore it may be understating the passenger demand for the busiest part of the peak. This could be offset a little by moving the mode of the opinion distribution so that it is a little higher than the mid point. The data aggregated from the two directions for each line are not helpful, and therefore these are generated to conform to a similar pattern to that which is revealed for the other lines. The weakness of the data describing the passenger flows per bus is therefore significant. The passenger flows for each bus and each bus stop are represented by the same opinion distribution. The data are not available to provide a more accurate estimate of the boarding pattern for each vehicle on each line at each stop. It is useful that it is possible to allow for a wide range of possibilities as a result of the flexibility provided by the opinion distribution to describe such approximate data.

Data for alighting passengers were not available from these sources, and therefore it is assumed that the number of alighting passengers can be described by an opinion distribution where the minimum, maximum and preferred values are one half of those used in the distributions describing passenger boardings. The opinion distribution used to describe passenger flow data per bus per stop is shown in Figure 9.1.

9.3.2.2.3 Speed data

As described in Chapter 7, the speed data are used for the initial assessment of the potential for a high capacity bus system design. The data used in this experiment are obtained from timings of various buses travelling along the corridor during the appropriate time period. The

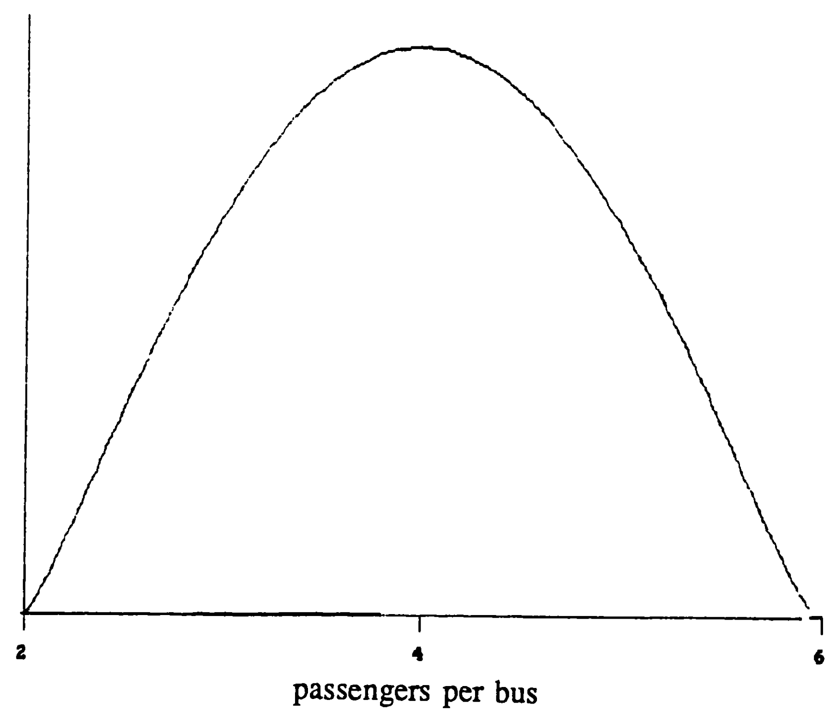


Figure 9.1 Passenger flow data for the Kingston to Tolworth corridor

free-flow speeds were obtained on a Sunday morning in order to provide an idea of the speed constraints imposed by the corridor itself. The opinion distributions used for the peak and free-flow speed data are shown in Figures 9.2 and 9.3 respectively.

9.3.2.2.4 Parking and enforcement

The parking opinion distributions are derived from the local knowledge of the traffic engineers in the Borough of the potential level of parking at various points in the corridor. This varies along the corridor, so the information relating to the busiest part is used. An interesting feature of this corridor is that a kerbside point-measure bus lane at the Tolworth end of the corridor has been removed as a result of complaints about loss of trade by the local traders.

Enforcement is considered to be fairly good, and the opinion distribution is intended to reflect this. These distributions are shown in Figures 9.4 and 9.5 respectively.

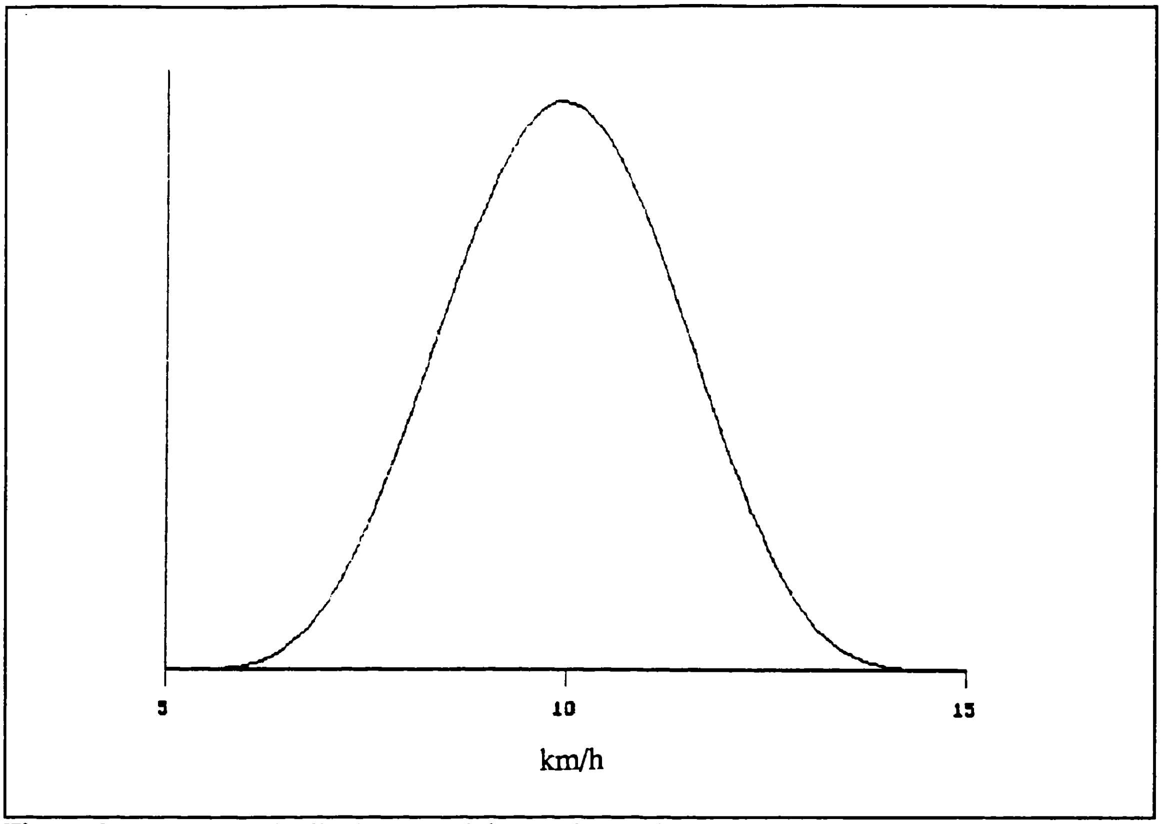


Figure 9.2 Opinion distribution describing peak speeds in the Kingston corridor.

9.3.3 Results

9.3.3.1 Initial assessment

The initial assessment suggests that a bus system could be attempted. This decision is based on the ratio of peak to off-peak speeds.

9.3.3.2 System characteristics

The decision matrices for the interpretation of parking and enforcement are shown in Table 9.4. The interpretation of the opinion distribution which describes parking indicates that the preferred descriptor is that the parking is described as *'fairly low'*. The Decision Factor is 1.47 which implies that the preferred choice is held with a reasonable degree of confidence. The second option is for the descriptor *'very low'*, suggesting that there is a general feeling of a low level of parking in the corridor.

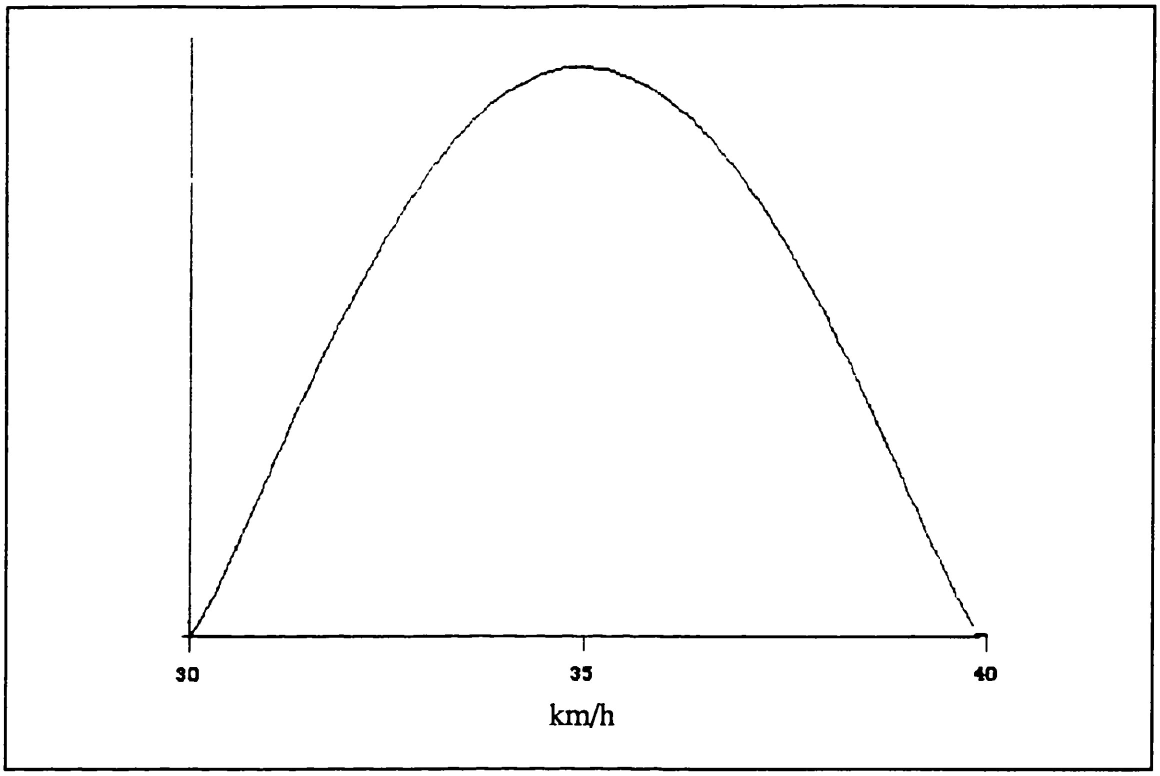


Figure 9.3 Opinion distribution describing the free-flow speed in the Kingston corridor.

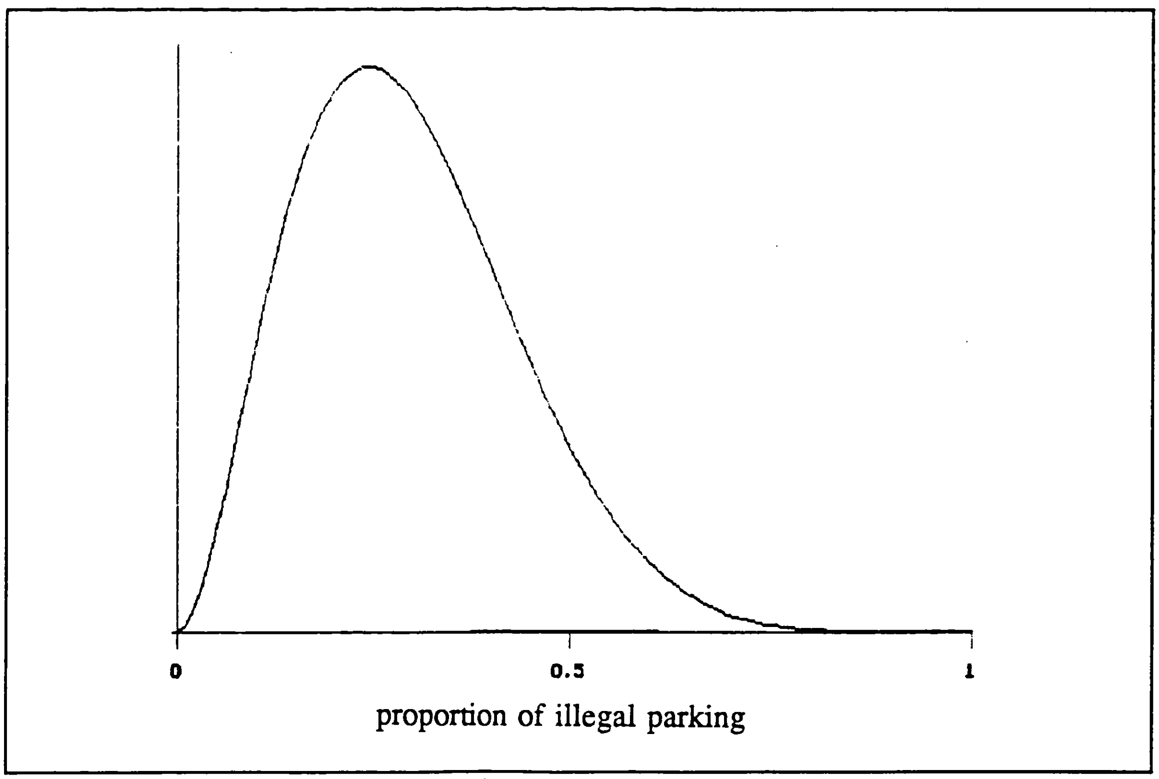


Figure 9.4 Opinion distribution describing land use effects on parking in the Kingston corridor.

The consideration of enforcement, however, is ambivalent. The preferred descriptor is '*fairly low*', and the second choice is for the descriptor '*fairly high*' This applies to a different sub-

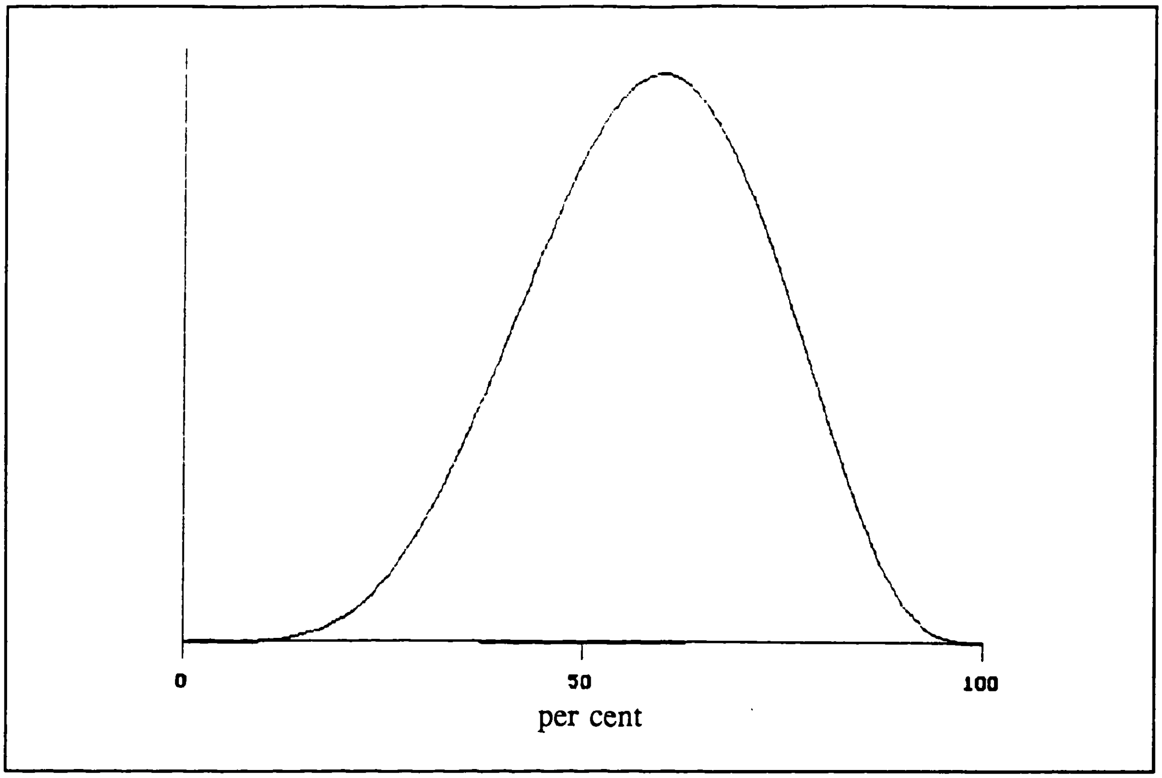


Figure 9.5 Opinion distribution describing the potential parking enforcement in the Kingston corridor.

range of the data. Within its own sub-range of the data, the preferred choice is held confidently (the Decision Factor is 2.11). When compared with the second choice, however, the degree of ambivalence is greater. In this case, the Decision Factor is only 1.11. This indicates that with the data as expressed, the ambivalence between the first and second choice descriptors suggests that the general feeling is that the level of enforcement could almost be described as *'fairly high'*. The actual choice of this as the descriptor for the data is, however, dependent on the sub-range of the data considered.

The cross-sectional position of the bus system in the roadway is therefore determined to be at the kerbside. The value of the parameter Q_{M1} is therefore 1.

Using the modes of the expertise, as described in Section 9.1.3, and since the land use is described as *'very low'*, the expertise also chooses the kerbside position. The value of Q_{E1} is therefore 1, and the Comparison Measure, C_1, is equal to zero. The Relative Assessment Measure, A_1, is thus also equal to zero.

9.3.3.3 Bus stop location and design

The average stop distance over the corridor is 228.41 metres. This is therefore the value of Q_{M2}. The optimum stop distance calculated according to the expertise is 226.72 metres. Thus the value of the Comparison Measure, C_2, is 1.69. It is not possible to assess a stop distance using an expertise distribution because the view of the expert is dependent upon the particular values of bus and passenger flows in the corridor. The Relative Assessment Measure is therefore taken to be the ratio of the Comparison Measure and the optimum stop distance calculated by the model. A therefore takes the value 0.007.

Table 9.4 Decision matrices for land use effects and potential parking enforcement in the Kingston corridor.

PARKING					
data range	*very low*	*fairly low*	*fairly high*	*very high*	**decision factor (DF)**
1	1.00	0.68	0.018	0.095	1.47
overall	1.00	0.68	0.018	0.095	1.47

ENFORCEMENT					
data range	*very low*	*fairly low*	*fairly high*	*very high*	**decision factor (DF)**
1	0.44		0.93		2.11
2		0.84		0.36	1.11
overall	0.44	0.84	0.93	0.36	1.11

For the remainder of the assessment, it is necessary to consider the links individually. The bus stop design for each link is discussed in Sections 9.3.3.3.1, 9.3.3.3.2, and 9.3.3.3.3. The design for the general infrastructure in each link is discussed in Sections 9.3.3.4.1, 9.3.3.4.2 and 9.3.3.4.3.

9.3.3.3.1 Link 1

Link 1 is too narrow to accommodate a bus stop. The consideration of whether the bus stop has overtaking lanes or not therefore does not apply. In order to register this in the assessment procedure, a value of 2 is allocated to Q_{M31}, which is used to indicate that one lane per direction is allocated for a bus lane. The values of Q_{M32}, Q_{M33}, and Q_{M34} are given the value of zero since the impossibility of a bus stop in this link precludes their consideration, and has already been registered in the value of Q_{M31}.

The value of Q_{E31} is one. This is because the expertise prefers to have a bus stop with an overtaking lane. By maintaining this value, the fact that the model has produced a design which is not compatible with the optimum of the expertise is registered. This should not be double-counted, however, and therefore Q_{E32}, Q_{E33}, and Q_{E34} are all given the value of zero. This produces a value of zero for each of C_{32}, C_{33} and C_{34}, which in turn produces a value of zero for A_2, A_3 and A_4. C_{31} is 1, and since no opinion distribution is available to assess this decision, A_1 is also 1.

9.3.3.3.2 Link 2

Link 2 is narrower than Link 1 and therefore is also unable to accommodate a bus stop. The consideration of whether the bus stop has overtaking lanes or not therefore does not apply. As with Link 1, therefore the values of Q_{D3j} are the same for this link as for Link 1. These are discussed further in Section 9.3.3.3.1.

9.3.3.3.3 Link 3

Two options were tested for Link 3: an available width of 16 metres and an available width of 19 metres.

9.3.3.3.3.1 Link 3: 16 metre width

In this case, a bus stop with no overtaking is deemed to be possible. Accordingly, a value of 2 is ascribed to Q_{M31}. Q_{M32} is given the value 1, since only one bus group is anticipated at this stop, and Q_{M33} is given the value 13, as the capacity of the stop is provided by one berth which is sufficient for the longer buses using the stop. As there is only one berth at the stop, Q_{M34} therefore takes the value 1.

The value of Q_{E31} is 1, since the expertise calls for a kerbside stop with overtaking. The other values of Q_{Ej} all take the same values as their counterparts from the model.

9.3.3.3.3.2 Link 3: 19 metre width

In this case, the model considers that it is possible to fit a kerbside stop with an overtaking lane. The stop design index, Q_{M31} is therefore 1. Since this agrees with the optimum design of the expertise, the value of C_{31} is zero. The other stop design parameters, as discussed in Section 9.3.3.3.3.1, agree with the expertise. Therefore these also attract values of zero for each value of C_{3j}. The corresponding values for A_{3j} are therefore zero.

9.3.3.4 General infrastructure

The widths for the elements of general infrastructure are now discussed. These are subject to assessment by the opinion model, and therefore each element has a decision vector describing the opinion of the expertise about its width. These decision vectors are shown in Table 9.5.

9.3.3.4.1 Link 1

In the absence of the possibility of building a high capacity bus stop in this link, the assessment of the general infrastructure is made using the elements designed for the link where there is no bus stop, but where a bus lane might be attempted. The value of Q_{M41}, indicating the number of traffic lanes per direction, is therefore 1. Q_{M42}, which is the width of a bus lane, takes the value of 3.5, and Q_{M43} (the width of the traffic lane) is 2.71. The footway width under these conditions is represented by Q_{M44} and is 0.915. The values of the optima are used for Q_{Ej}. Table 9.6 shows the values used in the assessment of the experiment for this link.

The value for Q_{E41} is 1, and the value of Q_{E42} is 3.5. The value of both C_{41} and C_{42} is zero. The width of the traffic lanes in the expertise describing '*satisfactory*', represented by Q_{E43}, is 3.51, and this produces a value for C_{43} of 0.8. The width of the footways in the expertise describing '*satisfactory*', given by Q_{E44} is 8.2 metres. The value of the Comparison Measure C_{44}

Table 9.5 Decision vectors describing the width of a bus lane, traffic lane and footway (Kingston).

element	**element width**	***too narrow***	***narrow, but possible***	***satisfa-ctory***	***too wide***	**decision factor**
LINK 1						
bus lane	3.5	0.25	0.024	1	0.48	2.08
traffic lane	2.71	0.62	0.77	0.18	0.35	1.24
footway	0.915	0.25	0.70	0.76	0.18	1.09
LINK 2						
bus lane	3.5	0.25	0.024	1	0.48	2.08
traffic lane	1.71	0.76	0.00	0.013	0.19	4
footway	0.915	0.25	0.70	0.76	0.18	1.09
LINK 3 (16m)						
bus lane	3.5	0.25	0.024	1	0.48	2.08
traffic lane	3.11	0.55	1	0.612	0.423	1.63
footway	1.015	0.21	0.60	0.78	0.19	1.30
LINK 3 (19m)						
bus lane	3.5	0.25	0.024	1	0.48	2.08
traffic lane	2.91	0.59	0.95	0.33	0.39	1.63
footway	0.965	0.23	0.65	0.77	0.18	1.19

Table 9.6 The assessment and tolerance measures for experiment 1, Link 1 (Kingston).

Parameter index	Value (model)	Value (expert)	Comparison measure	Relative Assessment
1	1	1	0	0
2	228.41	226.72	1.69	0.007
31	2	1	1	1
32	0	0	0	0
33	0	0	0	0
34	0	0	0	0
41	1	1	0	0
42	2.71	3.51	0.8	0.477
43	3.5	3.5	0	0
44	0.915	8.2	7.285	1.072
Total Assessment				2.556
Tolerance				**0.744**

is therefore 7.285. Using the appropriate elements of the decision vectors from Table 9.5 and Equation (9.1), the value for A_{43} is determined to be 0.477, and that for A_{44} is 1.072.

9.3.3.4.2 Link 2

As with Link 1, the construction of a bus stop is not feasible in this link, and therefore the assessment of general infrastructure is taken in terms of the allocation of the available space to the design elements where no bus stop exists. The design suggests that a bus lane may be possible and therefore the number of traffic lanes allocated is one, with one bus lane per direction. Table 9.7 shows the assessment data for Link 2. The value of Q_{M41} is 1. The bus lane width is 3.5 metres, but the traffic lane width is only 1.71 metres. As may be seen in Table 9.5, the decision vector for this element is very strongly indicative of an unwillingness to accept this as a feasible width (the Decision Factor is 4, and the preferred descriptor is '*too narrow*'). The footway width is only just deemed to be acceptable at 0.915 metres (the Decision Factor is 1.09). The resulting values for A_{43} and A_{44} are calculated using Equation (9.1), and are respectively 0.675 and 1.072.

Table 9.7 The assessment and tolerance measures for experiment 1, Link 2 (Kingston).

Parameter index	Value (model)	Value (expert)	Comparison measure	Relative Assessment
1	1	1	0	0
2	228.41	226.72	1.69	0.007
31	2	1	1	1
32	0	0	0	0
33	0	0	0	0
34	0	0	0	0
41	1	1	0	0
42	1.71	3.51	1.80	0.675
43	3.5	3.5	0	0
44	0.915	8.2	7.285	1.072
Total Assessment				2.754
Tolerance				**0.7246**

9.3.3.4.3 Link 3

As stated in Section 9.3.3.3.3, Link 3 is considered under two conditions. These are now considered independently.

Link 3: 16 metre width

Table 9.8 shows the assessment measures for the 16-metre option in Link 3. This design assumes that a bus stop may be fitted with no overtaking lane, which makes it possible to allocate the resulting space to the other elements. In this case, the traffic lane width is a little wider than is possible in Link 1 or Link 2, and the assessment measure is correspondingly lower at 0.14. The footway width, however, is only marginally wider than that in Links 1 and 2, and therefore the assessment measure for this element is only marginally lower than in the other cases at 1.014.

Table 9.8 The assessment and tolerance measures for the 16-metre option, Link 3 (Kingston).

Parameter index	Value (model)	Value (expert)	Comparison measure	Relative Assessment
1	1	1	0	0
2	228.41	226.72	1.69	0.007
31	2	1	1	1
32	0	0	0	0
33	0	0	0	0
34	0	0	0	0
41	1	1	0	0
42	3.11	3.51	0.4	0.140
43	3.5	3.5	0	0
44	1.015	8.2	7.185	1.014
Total Assessment				2.161
Tolerance				**0.784**

Link 3: 19 metre width

In this case, it is possible to consider a bus stop with an overtaking lane. Table 9.9 shows the assessment measures for this option. The value of Q_{M41}, indicating the number of traffic lanes per direction, is 1. Q_{M42}, which is the width of each bus lane, takes the value of 3.5, and Q_{M43} (the width of the traffic lane) is 2.91. The footway width under these conditions is represented by Q_{M44} and is 0.965. The decision vectors describing these widths are also shown in Table 9.5.

The assessment measure for the traffic lane is less than that for the 16-metre option because the lane is narrower (since an overtaking lane is now incorporated at the bus stop). This is determined to be 0.210. The footway is also narrower and therefore the assessment measure is also lower at 0.963. This is an interesting result because the model has shown that its allocation of space to the design elements is less acceptable to the expertise than the allocations in the 16-metre option. Whether the overall design, including the benefits to be obtained from the implementation of the overtaking lane is sufficient to offset these disadvantages can be determined

Table 9.9 The assessment and tolerance measures for the 19-metre option, Link 3 (Kingston).

Parameter index	Value (model)	Value (expert)	Comparison measure	Relative Assessment
1	1	1	0	0
2	228.41	226.72	1.69	0.007
31	1	1	0	0
32	1	1	0	0
33	13	13	0	0
34	1	1	0	0
41	1	1	0	0
42	2.91	3.51	0.6	0.210
43	3.5	3.5	0	0
44	0.965	8.2	7.235	0.963
Total Assessment				1.180
Tolerance				**0.882**

in the Tolerance Measure.

9.3.3.5 Tolerance

As with the discussions in Sections 9.3.3.3 and 9.3.3.4, Tolerance is discussed for each link separately.

9.3.3.5.1 Link 1

The Tolerance Measure calculated using Equations (9.2) and (9.3) for Link 1 is shown in Table 9.6. In this case, the value is 0.744. This is considerably below the acceptance level of 0.9. This implies that the use of suboptima is such that in general it is not acceptable to the expertise. On this basis, the feasibility of the design should be rejected.

9.3.3.5.2 Link 2

The Tolerance Measure for Link 2 is shown in Table 9.7 and is 0.725. This is even lower than the value for Link 1, and this design should also therefore be considered infeasible.

9.3.3.5.3 Link 3

Because the bus stop design and general infrastructure for Link 3 has been considered for two options, the Tolerance Measure is also considered separately for each option.

Link 3: 16 metre width

Table 9.8 shows that the value of the Tolerance Measure for this option is 0.784. While this is a little better than the values for Links 1 and 2, the design is still unacceptable to the expertise.

Link 3: 19 metre width

The Tolerance Measure is shown in Table 9.9 and is calculated to be 0.882. This is marginally below the acceptance value of 0.9. Although this indicates that the expertise would not accept the design as ideal, this result is expected since the overtaking lane is feasible only at the expense of a rather narrow footway.

It is interesting to note that the use of an overtaking lane is deemed to offset the disadvantages perceived in having suboptimal lane widths for the general traffic and a narrow footway. Even so, this compensation is not sufficient to permit acceptance of the design.

9.3.3.6 Conclusions: Experiment 1

The overriding conclusion obtained from this experiment is that the design of a complete segregated high capacity bus system, as represented by the expertise used in the experiment, is not possible in this corridor. It is important, however, to recognize some points that have arisen during this experiment.

First, the model has worked with imprecise data and has produced a design for the corridor, relocating the bus stops to give better accessibility, suggesting designs for the bus stops where they can be constructed, and indicating where the available space is not sufficient to allow high capacity bus stops to be built.

Second, the comparison is between an attempt to design a high capacity bus system and an assumption that the only values considered by an expert are the modes of the appropriate opinion distributions. As mentioned in Section 9.1.3, this is a particularly severe test of the attempts of the model to use suboptimal widths.

Third, the use of modal values for the expertise in this experiment does not constrain the expertise to work within the requirement that the total width should be used in the design. As stated in Section 9.3.3.4, if these modal values were used, the total space required would be 15.75 metres, which is a little short of the total width available. Since the objective of the model is to work within this constraint, the dissimilarity between the model's results and those represented by the expertise is disproportionate: in this example, over 83% of the overall comparison measure is accounted for by one parameter - the width of the footway when compared with the expertise.

Fourth, the objective of the experiment is to test the ability of the model to distinguish where it is not possible to design a high capacity bus system as a result of insufficient space for the infrastructure required. As far as this objective is concerned, the experiment has been a success. Such difficulties have been identified, and the appropriate advice has been given: that a high capacity bus system is not feasible in this corridor because of the impossibility of constructing the infrastructure necessary to protect buses from traffic and *vice versa*. The possibility of constructing a bus lane and bus stops in Link 3 is acknowledged, and a design suggested, although the resulting bus stop design is also considered to be infeasible.

The model has therefore succeeded in the sense that it has identified the problems and has suggested in which part of the corridor a solution may be practicable.

9.4 EXPERIMENT 2: LIMA

9.4.1 Introduction

The London experiment is intended to show that the model can function appropriately in situations where the physical constraints are such that a high capacity bus system could not be implemented. The Lima experiment, on the other hand, is intended to show that it is a fairly simple matter to test the possibilities which may become available when constraints are changed.

The corridor on which this experiment is based joins the centre of Lima to one of the suburban areas (Miraflores). The physical constraints of this corridor, which are almost the same along its entire length, are generally restrictive on the basis of environmental and historical grounds. This corridor is characterized by a wide formation, but with a garden in the centre of the roadway by which the traffic lanes in each direction are separated. Some corridors are much wider than the one used for this experiment, but this corridor is fairly typical of those in the older suburbs of Lima.

Brazilian engineers visited Lima in 1986 to design and implement a series of high capacity bus systems in some of the principal corridors of the city. The engineers used the garden areas in these corridors to provide space for the bus system, thus leaving the capacity for other traffic intact.

There is now, however, a degree of resistance in Lima to the implementation of similar bus corridors for a number of reasons. First, in these corridors the buses operate in the median, while the passengers are used to a stop-on-demand system operating at the kerbside (this is particularly true of the extensive network of microbuses). Thus passengers are required to cross to the central reservation in order to gain access to the bus system in the median. This inevitably reduces the level of stop-on-demand operation as it is neither possible to walk along the median, nor to wait for a bus, except at the formal bus stops. This effect is exacerbated by the second reason for the unpopularity of these bus systems. The bus stops are located at a minimum of 500 metre intervals. Even the state-run buses, which operate in a slightly more formal way than the microbuses, typically have stop spacings of about 200 metres in other corridors.

The third reason is that the bus lanes are segregated with fences to prevent passengers from crossing the road. It is now only possible to cross one of the Brazilian-designed corridors legally at the bus stops, unless other infrastructure exists which allows this manoeuvre. Thus the bus system is seen as a divisive element which separates communities on either side of the corridor. The distances involved in crossing one of these avenidas can be quite substantial: some have four lanes per direction for general traffic, and two lanes per direction for buses. Fourth, the removal of the central gardens is now seen to be environmentally unsound.

As a result, the conditions likely to be imposed upon any proposal for a high capacity bus system include the constraint that any central garden area should be retained.

This experiment therefore aims to show that the corridor may be analyzed under different conditions. Using the computer model, these conditions may be tested without recourse to expensive data collection surveys. Two scenarios are tested in the experiment: the first scenario retains the central garden area, and the second explores the possibilities where this space is made available for the transport system.

9.4.2 Data

This section considers the data used for this experiment. The physical data are discussed in Section 9.4.2.1, and the variable data are considered in Section 9.4.2.2. The physical data used in this experiment are based on information made available by INVERMET (the transport authority in Lima) and personal observations. The bus flow data are based on information from INVERMET (Invermet 1990) and the passenger flow data are based on observations at a selection of the bus stops along the corridor. Clearly, with a high degree of stop-on-demand operation, estimates of the number of boarding and alighting passengers are likely to be approximate.

9.4.2.1 Physical data

The corridor used as a data set for this experiment is Avenida Arequipa, from the junction with the Avenida 28 de julio to the junction with Avenida Ricardo Palma. There are six junctions (including the two mentioned) and five links. As mentioned above, the overall width of this corridor is consistent along this length, and the width of the central garden only varies by a small amount. Two lanes are available for all traffic in each direction, and there is currently no specific provision for the bus system. The data used in this experiment are shown in Tables 9.10 and 9.11. The overall physical data are the same for both experiments.

Table 9.10 Link data describing Avenida Arequipa, Lima (Perú)

Link	From	To	Length	Width
1	28 de julio	Tirado	820	27
2	Tirado	José Pardo	780	27
3	José Pardo	Javier Prado	810	27
4	Javier Prado	Angamos	1870	31
5	Angamos	Ricardo Palma	500	30

Access to the bus system at formal bus stops is possible at intervals of approximately two hundred metres, and stops have been defined accordingly. As mentioned in Section 9.4.1, however, there is an element of stop-on-demand operation and this does affect the stop frequency, especially where the microbuses are concerned. The way in which this affects the passenger flow data is described in Section 9.4.2.2.2.

9.4.2.2 Variable data

9.4.2.2.1 Vehicle data

The vehicle data are estimated from observations of journeys along the corridor. Bus flows are determined from the surveys carried out by INVERMET and which provide an estimate of the bus flows over the peak fifteen minutes. This value is therefore used as the mode of the opinion distribution. In order to obtain a maximum amount of information, the minimum and maximum values are set differently for each link, while the modal value is retained.

Table 9.11 Junction data for Avenida Arequipa, Lima (Perú).

Junction	Arm	stop line		entry line		central reserve	footways
		lanes	width	lanes	width		width
1	1	2	6	2	6	15	1.5
1	2	2	6	2	6	15	1.5
1	3	2	6	2	6	0	1.5
1	4	2	6	2	6	0	1.5
2	1	2	6	2	6	15	1.5
2	2	2	6	2	6	15	1.5
2	3	2	6	2	6	0	1.5
2	4	2	6	2	6	0	1.5
3	1	2	6	2	6	17	1.5
3	2	2	6	2	6	15	1.5
3	3	2	6	2	6	0	1.5
3	4	2	6	2	6	0	1.5
4	1	2	6	2	6	18	1.5
4	2	2	6	2	6	17	1.5
4	3	3	10	3	10	0	2
4	4	3	10	3	10	0	2
5	1	2	6	2	6	17	1.5
5	2	2	6	2	6	18	1.5
5	3	3	10	3	10	0	1.5
5	4	3	10	3	10	0	1.5
6	1	3	10	3	10	20	1.5
6	2	2	6	2	6	18	1.5
6	3	2	6	2	6	4	1.5
6	4	2	6	2	6	4	1.5

Two bus types operate along this corridor. These can be generally represented as large and small buses. The large buses are taken to have a capacity of about 90 passengers, to have two doors and to be approximately 13 metres in length. The small buses, on the other hand, seat 15 passengers as a maximum and are allocated a length of 8 metres (which may be a little generous in some cases). The small buses account for approximately 80-90 per cent of the bus flow. The data relating to buses is shown in Table 9.12.

Table 9.12 Bus data for Avenida Arequipa, Lima, Perú

Bus line	Links operated	Minimum flow	Maximum	Preferred	Length / capacity	Doors
2	1,2,3,4,5	15	25	20	13 / 90	2
53	1,2,3,4,5	15	25	18	13 / 90	2
24	1,2,3,4,5	120	180	140	8 / 15	1

9.4.2.2.2 Passenger data

As mentioned in Section 9.4.2.1, the passenger flow is a little difficult to estimate as a result of the informal nature of the bus stop operation. Nevertheless, as stated in Chapter 3, this situation is very common in developing countries. In order to include this demand, the corridor is divided into sections of approximately two hundred metres and the passenger flows described refer to the boarding of buses within each section, whether or not this takes place at a formal bus stop. This accords with the principle, noted in Chapter 3 and represented in the computer model as described in Chapter 7, that bus stops are defined as areas rather than points in space in order to allow situations such as these in Lima to be described. Because bus stops are defined, in terms of their location, to be between two junctions, this presents no problem for the data entry process.

To give an indication of these demand levels, the average passenger demand for each stop for all bus lines is shown in Table 9.13. The model then calculates the demand along the corridor as if each of these sections contained one bus stop with the appropriate demand level.

9.4.2.2.3 Speed data

Speeds are estimated from journey times along the corridor. Since these data are very weak, the opinion distribution is accordingly characterized by a large variance. The opinion distribution for peak speed is shown in Figure 9.6.

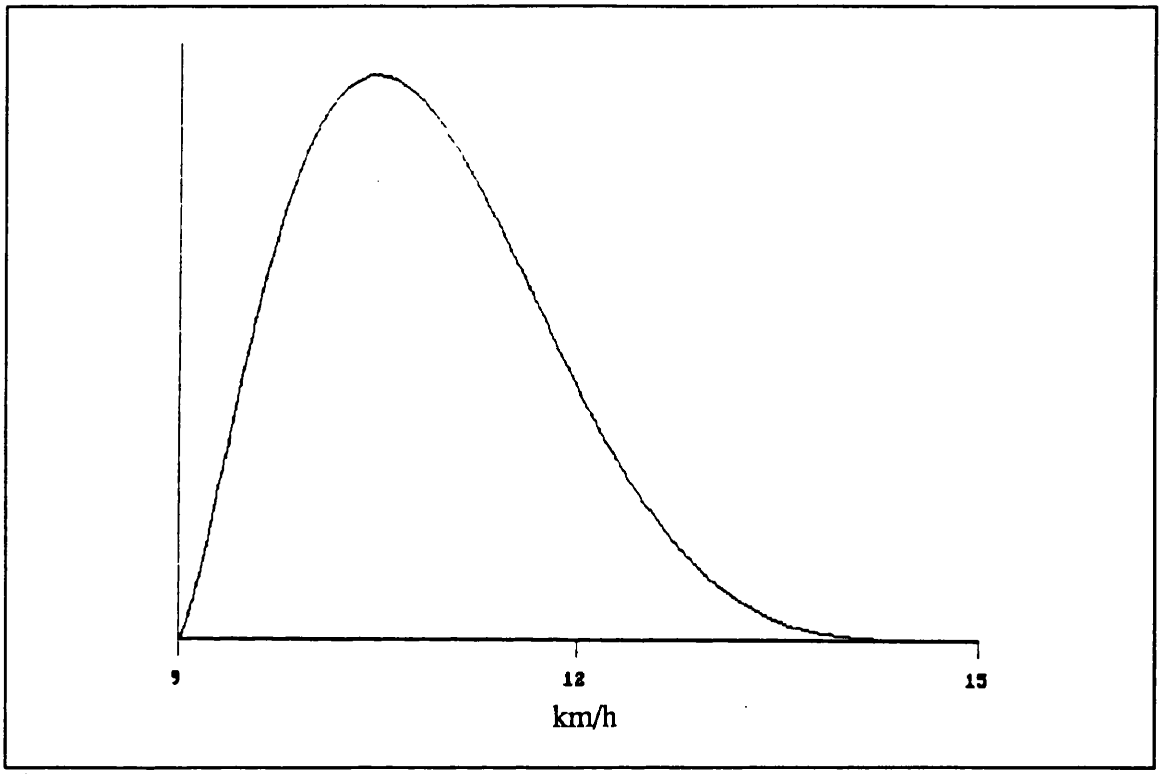


Figure 9.6 Opinion distribution describing peak speeds in Avenida Arequipa, Lima, Perú

9.4.2.3 Land use and enforcement

Because the space allocated to traffic is limited in the current context, the land use tends not to affect the parking levels along the corridor itself. The level of parking is therefore very low. Enforcement is, as is generally the case in Lima, poor. These data are therefore suitably described by the distributions illustrated in Figures 9.7 and 9.8.

Table 9.13 Demand data for Avenida Arequipa, Lima, Perú

Stop number	minimum	maximum	preferred
1	1	6	5
2	1	6	4
3	1	6	5
4	1	6	5
5	1	6	4
6	1	5	3
7	1	6	5
8	1	5	4
9	1	6	5
10	1	6	5
11	1	6	4
12	1	4	3
13	1	6	5
14	1	6	4
15	1	7	4
16	1	7	6
17	1	6	5
18	1	7	4
19	1	3	2
20	1	7	6
21	1	5	3
22	1	7	5
23	1	8	7

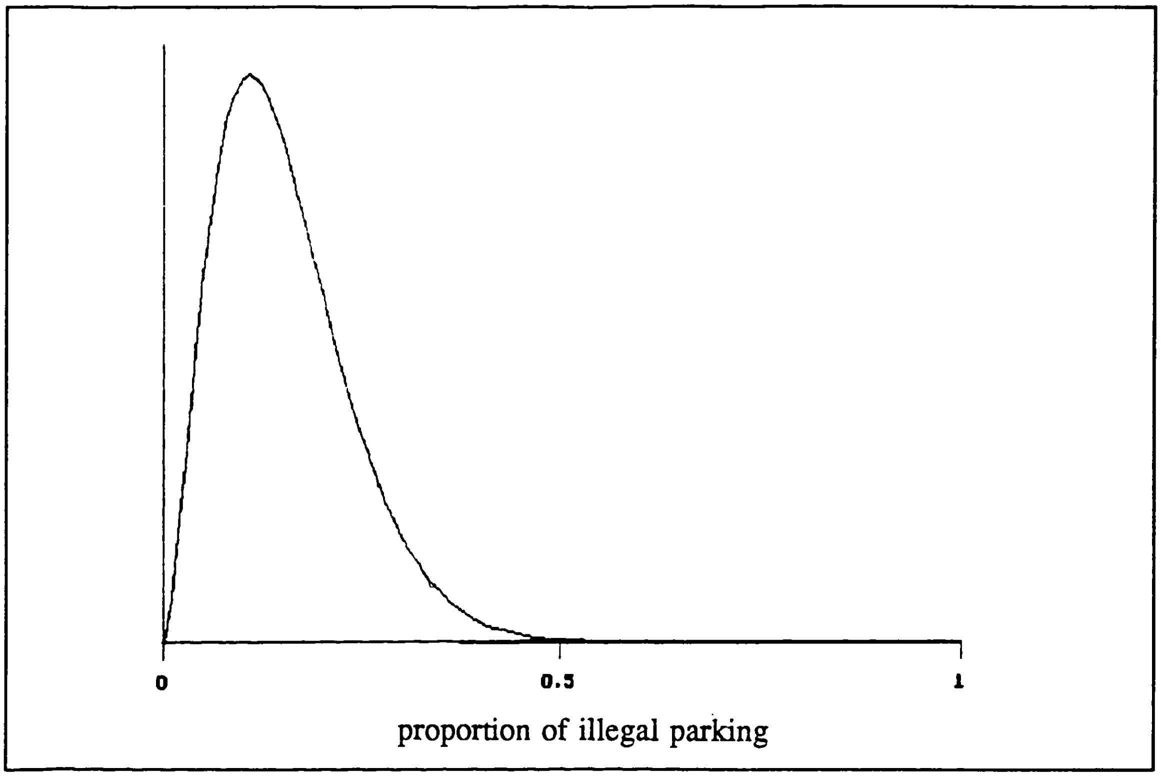


Figure 9.7 Opinion distribution for the land use effect in Avenida Arequipa.

9.4.3 Results

Since this experiment considers two sets of constraints concerning the availability of space for the bus system, the results are reported in two parts where necessary. The change in these constraints only affects the allocation of space to bus stops and general infrastructure, and therefore the sections dealing with these design characteristics (Sections 9.4.3.3 and 9.4.3.4) describe each set of results in turn. Correspondingly the discussion about the Tolerance Measure in Section 9.4.3.5 also considers the two situations separately.

9.4.3.1 Initial assessment

The initial assessment shows that an analysis should proceed on the basis of the comparison of peak and off-peak speeds.

9.4.3.2 System characteristics

The decision matrices resulting from the consideration of parking and enforcement are shown in Table 9.14. The interpretation of the opinion distribution which describes parking

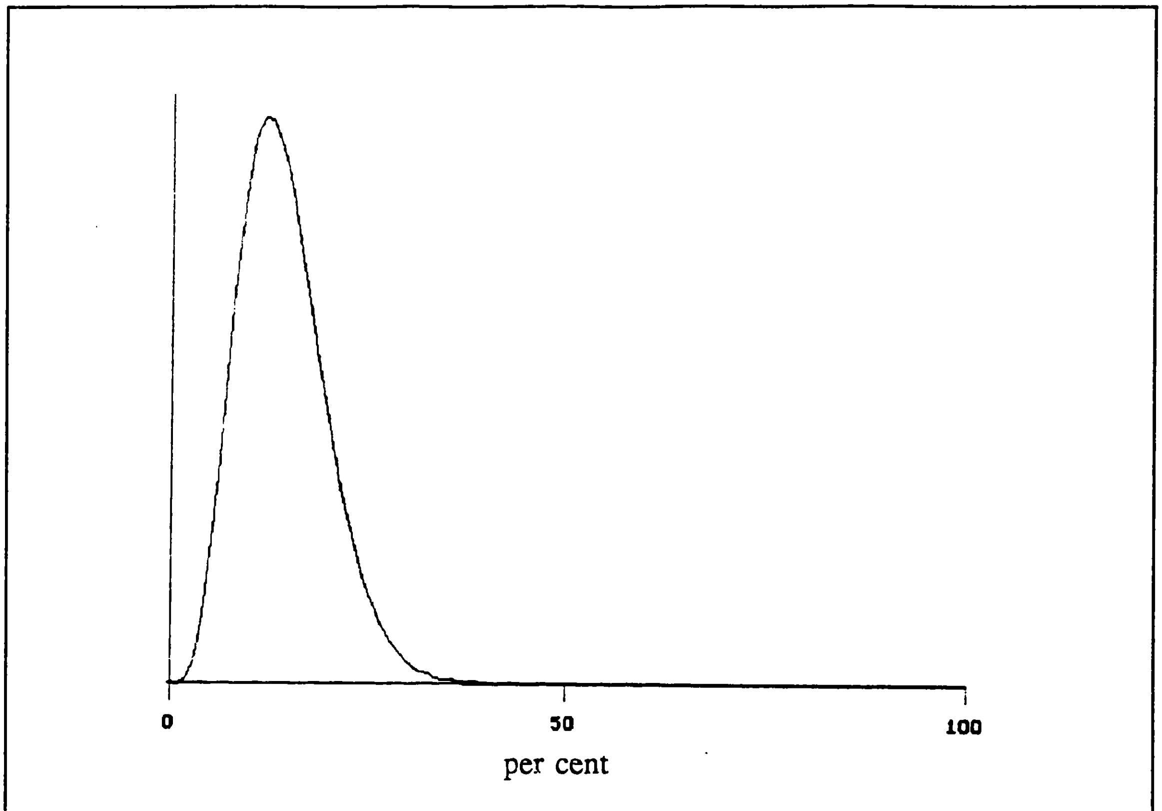


Figure 9.8 Opinion distribution describing enforcement in Avenida Arequipa.

provides an alternative depending on the value of the input data. The preference is that the parking is described as '*very low*', and the second option is for the descriptor '*fairly low*'. Consideration of these descriptors gives a Decision Factor of 3.23, indicating a fairly confident feeling that the true descriptor is '*very low*'. The smallness of the opinion values for the other sub-range of the data indicates that it is not considered likely that this data sub-range would occur. The opinion value for the preferred descriptor in this sub-range is extremely low (resulting in a Decision Factor of 41.76), suggesting that this could be reasonably ignored as a serious possibility for consideration.

Similarly, enforcement is also considered to consist of two alternatives depending on the data value. As with parking, the consideration of '*very high*' could be ignored on the basis of the very low opinion value which results in a Decision Factor of 330.00. Although the preferred descriptor, '*very low*', has a fairly low associated opinion value (0.66), the overall Decision Factor is 5.08. This suggests that the preference for this descriptor is very strong.

The cross-sectional position of the bus system in the roadway is therefore determined to be at the kerbside. The value of the parameter Q_{M1} is therefore 1.

Table 9.14 Decision matrices for parking and enforcement in Avenida Arequipa, Lima (Perú).

PARKING					
data range	*very low*	*fairly low*	*fairly high*	*very high*	**decision factor (DF)**
1	0.71	0.22			3.23
2			0.014	0.017	41.76
overall	0.71	0.22	0.014	0.017	3.23

ENFORCEMENT					
data range	*very low*	*fairly low*	*fairly high*	*very high*	**decision factor (DF)**
1	0.66	0.13		0.006	5.08
2			0.002		330.00
overall	0.66	0.13	0.002	0.006	5.08

Using the modes of the expertise as described in Section 9.1.3 and since the land use is described as '*very low*', the expertise also chooses the kerbside position. The value of Q_{E1} is therefore 1, and the Comparison Measure, C_1, is therefore equal to zero.

9.4.3.3 Bus stop location and design

The optimum stop distance for the corridor obtained by the computer model is 202.06 metres. The average stop distance under the new design is 203.17 (this compares with the current average stop distance over the corridor of 203.88 metres). The value of the Comparison Measure C_2 is therefore 1.11. The consideration of stop design depends on the available space. There is therefore a choice of responses, according to the criterion adopted with respect to the central garden.

9.4.3.3.1 Central garden retained

When the central garden area is retained, the model finds that it is not possible to accommodate a high capacity bus system using segregated bus lanes in any of the links. In this circumstance, bus stop design is therefore also impossible. A value of 2 is therefore entered for the bus stop type to indicate that it is not possible. As discussed in Section 9.3.3.3.1, to avoid counting this result twice, the other values in this assessment are considered to be zero for both the model design and the expertise.

9.4.3.3.2 Central garden removed

If the area currently occupied by the central garden is made available for the bus system, however, it is possible to design a kerbside bus stop with an overtaking lane in each link. The length of the bus stop platform required for this level of bus operation is 74 metres, and accommodates a double group double berth stop. The resulting values for Q_{M3j} are given in Table 9.18. Consideration of the values of Q_{E3j}, also shown in Table 9.18, shows that the value of each of the Comparison Measures C_{3j} is also zero.

9.4.3.4 General infrastructure

The general infrastructure is now discussed. In Section 9.4.3.4.1, the discussion considers the design which arises when the constraint is imposed that the central garden should be retained. Section 9.4.3.4.2 discusses the alternative, where the central garden may be removed.

9.4.3.4.1 Central garden retained

Using the mode of the expertise distributions, the width required to accommodate a bus stop with no overtaking lane with the central garden retained, it would be necessary to have an available space of 46.17 metres in Links 1, 2 and 3. A width of 49.17 metres would be required in links 4 and 5. Since the width for the first three links is only 27 metres (and that for the last two is 30 metres), this would not appear to be feasible in any link. The results discussed in this section (and those discussed in Section 9.4.3.5.1) are therefore of interest because they may explain the rationale behind the decision of the model not to implement a bus system under these constraints.

Table 9.15 Decision vectors - central garden retained.

ELEMENT	width	*too narrow*	*narrow-ish*	*satisfactory*	*too wide*	decision factor
LINKS 1 - 3						
bus	3.5	0.25	0.48	1.00	0.02	2.08
traffic	1.11	0.85	0.00	0.00	0.11	7.73
footway	1.015	0.21	0.60	0.78	0.19	1.30
LINK 4						
bus	3.5	0.25	0.48	1.00	0.02	2.08
traffic	1.71	0.76	0.01	0.00	0.19	4.00
footway	0.915	0.25	0.70	0.77	0.18	1.10
LINK 5						
bus	3.5	0.25	0.48	1.00	0.02	2.08
traffic	1.11	0.85	0.00	0.00	0.11	7.73
footway	1.015	0.21	0.60	0.78	0.19	1.30

Table 9.15 shows the decision vectors for the element widths under these constraints. The narrowness of the traffic lane is registered in two ways: the descriptor '*too narrow*' has a reasonably high opinion value of 0.85, but this is reinforced by the value of the Decision Factor (7.73). This is because the lane width is clearly too narrow, and the expertise is almost unequivocal in its decision about this descriptor. The value of 0.11 for the descriptor '*too wide*' is obtained as a result of the shape of the expertise distribution describing this vector. In order to obtain the decision threshold between the descriptors '*satisfactory*' and '*too wide*' with the correct opinion value and at the correct lane width, the experts defined an opinion distribution for the latter which increases slowly over the whole range of widths considered. In this case, this has produced the strange result that such a narrow lane width has a measurable opinion value for the descriptor '*too wide*', but not for the descriptors '*narrow, but possible*' and '*satisfactory*'. In practice, this is not too troublesome: the Decision Factor is very high. This effectively reduces the relative importance of the second alternative.

9.4.3.4.2 Central garden removed

With the central garden area available for use in the design, the width required if all the elements were to be set at their optimum widths would be 47.96 metres in all links. This allows for three traffic lanes in each direction. The solution which is constrained by the available width in each link is shown in Table 9.18 which therefore gives the values of Q_{M4j}. The associated opinion values are given in Table 9.16. In the first three links, the width of the traffic lanes has been reduced to 3.41 metres, with two lanes. The width of the footway is 1.055 metres. Link 4 is able to accommodate two lanes of 3.51 metres, and a footway width of 2.855 metres. Link 5 also has traffic lanes of 3.51 metres, but the footway is reduced to 2.355 metres in width. In each case, the decision matrix shown in Table 9.16 indicates that the descriptor '*satisfactory*' may be applied to the width of each element with some confidence: the lowest Decision Factor is 1.25, and the highest is 2.97.

Table 9.16 Decision vectors - central garden removed.

ELEMENT	width	*too narrow*	*narrow-ish*	*satisfactory*	*too wide*	decision factor
LINKS 1 - 3						
bus	3.5	0.25	0.48	1.00	0.02	2.08
traffic	3.41	0.51	0.77	0.97	0.48	3.41
footway	1.055	0.20	0.56	0.79	0.19	1.41
LINK 4						
bus	3.5	0.25	0.48	1.00	0.02	2.08
traffic	3.51	0.49	0.64	1.00	0.50	1.56
footway	2.855	0.01	0.01	0.91	0.33	2.76
LINK 5						
bus	3.5	0.25	0.48	1.00	0.02	2.08
traffic	3.51	0.49	0.64	1.00	0.50	1.56
footway	2.355	0.02	0.03	0.89	0.30	2.97

9.4.3.5 Tolerance

As mentioned in Section 9.4.3.5 with reference to the corridor in Kingston, the Total Assessment Measure is an indication of the degree to which the model has chosen suboptimal characteristics, and the level of happiness felt about these suboptima when these are compared with the appropriate preferred value. The Tolerance Measure is an indication of the degree to which the expert would tolerate the differences in design which may be considered when compared with a design where only optimal values are used for each element.

9.4.3.5.1 Central garden retained

In the case of Avenida Arequipa, Table 9.17 shows the Total Assessment and Tolerance Measures for the corridor when the central garden is retained. Three assessments are made on the basis of the available width in different parts of the corridor: for the first three links, link 4 and link 5.

In the case of links 1 to 3, the Total Assessment Measure is 2.285 and the Tolerance Measure is 0.772. This is considerably less than 0.9, which is the threshold for acceptance of suboptimal designs, and indicates that the design is not thought to be sufficiently acceptable to the expertise.

Link 4 has a Total Assessment Measure of 2.729, which gives rise to a Tolerance Measure of 0.727. This is broadly comparable with the Tolerance Measure obtained for Links 1 to 3. An interesting feature in this case is that the Tolerance Measure for this link is lower than that for Links 1 to 3. Since the traffic lanes are wider in Link 4 than in the others, it might be expected that the Tolerance Measure would be higher. That it is not is due to the effect of the extremely high Decision Factor which characterizes the assessment of Links 1 to 3.

A high Decision Factor suggests that the preferred descriptor is considered to be correct with a high degree of confidence, and this is reflected in the calculation of the Relative Assessment Measure. In this case, the Decision Factor more than offsets the reduction in the Relative Assessment Measure which is brought about by the fact that the preferred descriptor is '*too narrow*'. Thus confidence in the chosen preference (or a lack of ambivalence about the

Table 9.17 The assessment and tolerance measures for experiment 2 (central garden retained).

Parameter index	Value (model)	Value (expert)	Comparison	Relative Assessment
ALL LINKS				
1	1	1	0	0
2	203.17	202.06	1.11	0.005
31	2	1	1	1
32	0	0	0	0
33	0	0	0	0
34	0	0	0	0
LINKS 1 - 3				
41	1	1	0	0
42	1.11	3.51	2.4	0.416
43	3.5	3.5	0	0
44	1.015	8.2	7.185	0.864
Total Assessment				2.285
Tolerance				**0.772**
LINK 4				
41	1	1	0	0
42	1.71	3.51	1.8	0.675
43	3.5	3.5	0	0
44	0.915	8.2	7.285	1.049
Total Assessment				2.729
Tolerance				**0.727**
LINK 5				
41	2	2	0	0
42	1.11	3.51	2.4	0.416
43	3.5	3.5	0	0
44	1.015	8.2	7.185	0.864
Total Assessment				2.285
Tolerance				**0.772**

choice) may counteract the fact that the preferred descriptor is not the optimum.

As the opinion value associated with the preferred descriptor of a particular element width reduces, it tends towards some point at which it is equal to the opinion value associated with the second choice. As discussed in Chapter 5, near to this threshold, the importance of the actual opinion value and its relationship with that of the second choice increases. The effect of this importance is reflected in the Relative Assessment Measure: a lower opinion value and a lower Decision Factor would reduce the magnitude of the denominator of Equation (9.1), thus tending to increase the value of the measure. The Relative Assessment Measure therefore tends to prefer a poor choice which is held confidently to a better choice held with a greater degree of ambivalence.

Although the total width of Link 5 is wider than that for Links 1 to 3, this is offset by an equal increase in the width of the central garden. The element widths are therefore the same as for Links 1 to 3. The Total Assessment Measure is therefore 2.285, and the Tolerance Measure is 0.772.

9.4.3.5.2 Central garden removed

Table 9.18 shows the Total Assessment and Tolerance Measures for the situation where the central garden area is made available.

For links 1, 2 and 3, the Total Assessment Measure is 0.816. This implies a considerable change from the optimal values contained in the expertise. Nearly all of this change (95.8%) is accounted for by the suboptimal footway width. The effect of the change in footway width is not only because of the magnitude of this change. As described in Section 9.2., three other factors combine with the magnitude to produce this evaluation measure: the opinion value, the Decision Factor and the descriptor. In this case, although the opinion value is reasonably high, at 0.79 it is not excessively so. The Decision Factor is 1.41 which is not overwhelmingly high and suggests that the decision to describe this width as '*satisfactory*' may be a little ambivalent. Since the opinion value is expressed for the descriptor '*satisfactory*', the descriptor is the same as that used for the optimum width in the expertise, and therefore there is no influence from this parameter in the evaluation of the value of the Total Assessment Measure. The resulting Tolerance Measure is 0.919, which is deemed acceptable by the rule, suggested in Section 9.2.3, that a reasonable

Table 9.18 The assessment and tolerance measures for experiment 2 (central garden removed).

Parameter index	Value (model)	Value (expert)	Comparison	Relative Assessment
ALL LINKS				
1	1	1	0	0
2	203.17	202.06	1.11	0.005
31	0	0	0	0
32	1	1	0	0
33	74	74	0	0
34	2	2	0	0
LINKS 1 - 3				
41	2	2	0	0
42	3.41	3.51	0.1	0.025
43	3.5	3.5	0	0
44	1.055	8.2	7.145	0.782
Total Assessment				0.812
Tolerance				**0.919**
LINK 4				
41	2	2	0	0
42	3.51	3.51	0	0
43	3.5	3.5	0	0
44	2.855	8.2	5.345	0.260
Total Assessment				0.265
Tolerance				**0.974**
LINK 5				
41	2	2	0	0
42	3.51	3.51	0	0
43	3.5	3.5	0	0
44	2.355	8.2	5.845	0.270
Total Assessment				0.275
Tolerance				**0.973**

acceptance of suboptimality is that the Tolerance Measure should be greater than or equal to 0.9.

Considering Link 4, the Total Assessment Measure is 0.265, which is much lower. Once again the main element of this measure is the change in footway width which accounts for 96.7%. When compared with the same element in Links 1 to 3, the value of the Relative Assessment Measure is considerably smaller. The magnitude of the difference between the actual width and the optimal width is a little smaller (5.345 compared with 7.145). The major reason for the difference between the value of the Total Assessment Measure for Link 4 and that for Links 1 to 3 is obtained from an analysis of the opinion values and Decision Factors concerned. As can be seen in Table 9.16, the opinion value for the footway width in Link 4 is 0.91 for the descriptor '*satisfactory*'. This is some 15% higher than that for Links 1 to 3. The Decision Factor for the footway width in Link 4 is 2.76. This is 95.7% higher than that for Links 1 to 3, implying a much higher degree of confidence that this width can be described as '*satisfactory*'. The resulting Tolerance Measure is 0.974 which suggests that the degree of tolerance felt between the element characteristics of the design produced by the model and the optimum widths of the expertise is, on average, quite high (and a considerable improvement over that for Links 1 to 3).

Link 5 is very slightly more constrained than link 4, having one metre less over the whole width. In this case, the Total Assessment Measure is 0.270, with 96.8% of this accounted for by the footway width. This results in a Tolerance Measure of 0.973, which is marginally less than that for Link 4, but still acceptable. The suggested design for a bus stop in this corridor is shown in Figure 9.9.

9.4.3.6 Conclusions: Experiment 2

The major conclusions drawn from Experiment 1 also apply in this case: the model can distinguish where it is not possible to locate a high capacity bus system and it can operate with imprecise data. Three major additional conclusions are, however, obtained from this experiment.

First, a high capacity bus system should be implemented in this corridor, but with the existing bus and passenger flows, this is not possible unless the central garden area is removed and the space made available for the implementation of a bus system. If this can be done, the bus system should not be located in the median, but should remain at the kerbside. There is sufficient space to allow a double width bus lane along this corridor under these circumstances, and a design

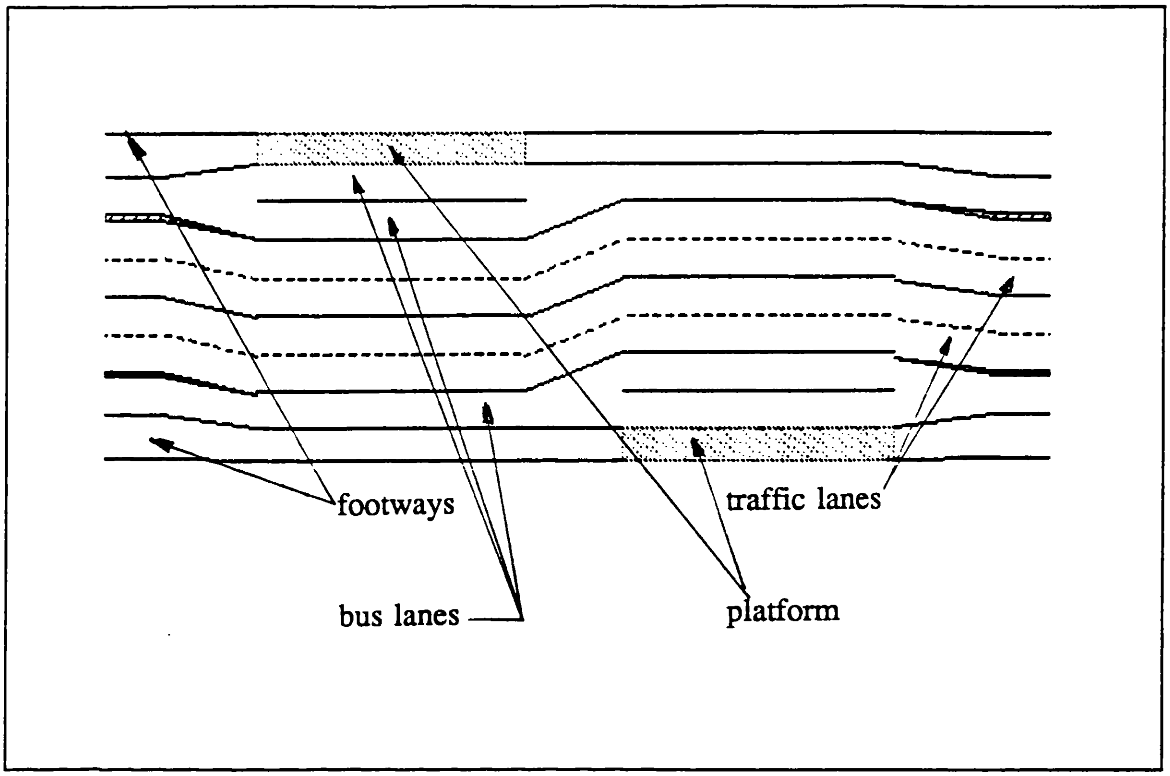


Figure 9.9 A sketch of the suggested design for a bus stop in Avenida Arequipa, Lima (Perú), with the central garden removed.

of this type would allow the stop-on-demand operation (with its attendant benefits and disadvantages) to continue.

Second, the model can produce alternative designs under different constraints, indicating the feasibility of the alternatives. The advantage of this is that the use of imprecise data in the production of these alternatives is that the comparison may be performed initially with only a modest investment in data collection.

Third, the test used is a severe measure (only the optima in the expertise are considered acceptable). In this experiment this has shown that the model can distinguish between a conclusion which is poor from the point of view of the design, but which is strongly held and a conclusion which is better in terms of design, but about which there is a higher degree of ambivalence.

9.5 GENERAL CONCLUSIONS

This chapter has described two sets of experiments carried out using the decision model. The aim of these experiments was to test whether the decision model is capable of making appropriate decisions about the design of a high capacity bus system in situations where the physical constraints of the corridor may not be sufficient for such a system.

In order to achieve this aim, the model was tested using data based on a corridor in London and a corridor in Lima (Perú). A range of constraints were tested, and the model succeeded in identifying where a high capacity bus system is not feasible, where suboptimal stop designs may be incorporated to ease problems, and how designs may be modified in order to obtain a satisfactory bus system. Satisfactory degrees of saturation have been achieved at bus stops, through a combination of stop location and stop design.

The experiments may therefore be considered successful as they have demonstrated the capability of the model to infer decisions from imprecise information, and to measure its ability to interpret these decisions in terms of the resulting degree of suboptimality. Several problems remain, however, and these are now considered.

First, the model treats the corridor very simply in terms of its physical constraints. It would be helpful if a link could be considered with a selection of widths without the necessity of running the program several times.

Second, the decision to close junctions is not included in this decision process. The reason for this is that it would be necessary to know more about the surrounding network, including the feasibility of introducing one-way systems. This would add considerably to the amount of data necessary for the program to operate. Nevertheless, some indication of the possibility of closing or retaining junctions could be gleaned from the knowledge of traffic flows using the non-corridor arms. These data could be used to advise about the potential benefits to be gained from closure, as well as the possible problems which may arise.

Third, although the recommendation to group bus lines is included as a possibility where this is appropriate, currently no advice is available as to how this should be done. Although the decision to recommend grouping and the number of groups recommended are both determined on the basis of the model discussed in Section 3.3.4.2.2, the exact choice of bus lines to have in each group is not suggested by the model. A more sophisticated approach to the grouping of bus lines would provide a more comprehensive solution, including, for example recommendations for groupings based on the combined frequencies of the bus lines. This would also need some indication of the proximity of the origins and destinations of all bus lines so that the geographical constraints can be maintained.

Fourth, where the model cannot provide a solution which makes use of a high capacity bus system, it would be helpful if it could recommend possible improvements to the existing system.

Fifth, as there is currently only a small number of experts in the field of designing high capacity bus systems, only one expertise set was tested in these experiments. The methodology of the decision model is designed to enable the user to choose amongst different expertise sets. The results of using different sets has therefore not been tested in this case.

Sixth, a vital part of the design process is the evaluation of the design in terms of its efficacy as a high capacity bus system. As the model being tested here is the decision model, this evaluation is not included within the design process. Clearly, however, this is an aspect of the design process which should be included in a design tool. As mentioned in Section 9.2.1.1, however, existing models are not sufficient to test the subtleties of the design changes considered within this process. This is a matter for urgent consideration, and further detailed research.

As described in Chapter 7, the allocation of space to the various elements of infrastructure is undertaken on the basis that the design should accommodate similar levels of capacity in both directions. This may not always be necessary, and therefore an improvement might be to assess individually the demand characteristics of each direction in the corridor.

CHAPTER 10
CONCLUSIONS AND RECOMMENDATIONS FOR FURTHER RESEARCH

10.1 INTRODUCTION

This chapter draws together the conclusions drawn from the various strands of this research in order to provide a set of general conclusions. These form the basis of an analysis of the possibilities for future research into the use of artificial intelligence techniques in the design of high capacity bus systems.

10.2 CONCLUSIONS

The discussion in Chapter 2 highlights a number of issues which are of importance in the consideration of this research. The travel patterns common in the cities of developing countries suggest that there is a requirement for some form of high capacity transport system. The economic situation which exists in these countries determines that the costs involved in developing and constructing heavy rail systems to meet this capacity may be prohibitive. The rate of population growth in many of these cities is such that the time taken for such developments is likely to be too great to solve the short and medium term capacity problem.

There is therefore a need for an alternative high capacity system which can be implemented more cheaply and more quickly than common alternatives such as a heavy-rail metro system. Brazilian experience shows that high capacity bus systems can provide a flexible, low-cost injection of capacity into the public transport system which can be implemented within a short timescale.

The conclusion is therefore drawn that it is desirable to examine the design principles used in the development of high capacity bus systems in order to facilitate the transfer of this technology to other environments where this may be appropriate.

Accordingly the engineers responsible for the Brazilian systems were approached in order to ascertain their design methods. Chapter 3 describes and discusses both the Brazilian expertise and that from other countries, including the United Kingdom, North America, Chile and several

European countries. Various models used by these engineers have been examined, and the way in which these are selected, and their data requirements have been analyzed. A number of conclusions may be drawn from the analysis of this expertise.

Efficient operation of a bus system is dependent upon the use of a combination of measures, designed to ease the progress of buses through critical points. In the United Kingdom, the critical points are typically seen to be junctions and the bus priority measures are therefore related to individual junctions and do not tend to be coordinated along a corridor as a whole. Capacity-oriented systems, however, tend to be implemented over a longer distance than the point measures used to allow buses to pass through difficult bottlenecks.

High capacity bus systems are constrained by the capacity of the bus stops. The capacity of the system may be enhanced by better design of bus stops and this is therefore the principal tool used by the designers of such systems in order to increase capacity.

The operation at bus stops is assisted if the buses are independent of other traffic and of each other. This suggests the need for some form of segregation which would preserve the integrity of the bus system.

In order to ensure that bus stop operation is maintained at a high level of efficiency, it is important that the headways between buses are consistent. This is affected by two factors: the operational efficiency of previous bus stops in the corridor and the ability of buses to travel between the bus stops uninterrupted by traffic-induced congestion. The former may be achieved by ensuring that bus stop design is sufficient so that the buses are not delayed unduly at previous bus stops. The latter requires the use of segregated bus lanes which serve the purpose of a "track" for buses between the stops.

Design of a bus stop revolves around the number of buses that it is possible to accommodate simultaneously. This can be increased with a larger number of berths, or by dividing the bus lines serving the stop into groups. The latter provides a more consistent increase in capacity, but it is necessary to ensure uninterrupted overtaking at the stop in order to maintain the functional independence of the groups. Generally, the more independent different vehicles can be of the activities of other buses, the more efficient is the operation of a bus stop.

The location of the stops can affect their capacity as a result of the effects of stop location on the number of passengers wishing to use each stop. Stop location is a function of the relationship between the passenger density and the bus flow and depends heavily upon the possibility of overtaking at the stops.

Using these techniques it is possible to provide a high capacity bus system which makes a significant contribution to the movement of large numbers of people.

In the design of high capacity bus systems, decision outputs are often discrete, but are based upon inputs which are continuous variables. The design process is therefore characterized by the large amount of time spent on the interpretation of poor quality data. Several of the design processes require the evaluation of models which are dependent for their inputs upon data which are not known precisely. The design process thus depends upon the way design engineers think about the problem. The design process is therefore not just a set of quantitative analyses: it is also subject to the interpretative abilities of the design engineer.

A computer model could assist in the transfer of this technology if it is capable of representing these interpretative processes. In order to facilitate this technology it is therefore necessary to consider how this interpretative process may be modelled. The possibility examined for this purpose in this research is the use of artificial intelligence techniques.

It appears that the design of high capacity bus systems is an area in which knowledge-based systems may be used. Before designing a knowledge-based system, however, it is essential to examine the nature of the expertise concerned.

In this case the expertise can be categorized into three types: factual knowledge (such as the knowledge that a particular model exists), modelling knowledge (for example, the expertise in using the model), and judgement (or opinion), which is used in the interpretation of the data and the output of these models.

The detailed expertise is often personal to particular experts who prefer different alternatives when given the same data. A model to represent such personal judgement is therefore required to address the question of the evaluation of opinion. Existing artificial intelligence techniques are not, however, capable of representing the flexibility required by these interpretative processes within the design processes.

A model is therefore derived which attempts to represent the use of judgement and opinion. This model identifies decision thresholds which indicate where decision outputs may alter for marginal changes in the inputs.

Opinion may be described in terms of shape, which can be represented conveniently by a β distribution. The opinions concerning a particular decision may be represented by a set of four such distributions, where a decision threshold is identified as the point at which two distributions have the maximum opinion value.

Fuzzy logic may be used to propagate opinion values through the decision process. Conventionally, however, the support values used in fuzzy set theory are assigned arbitrarily. This is a particular weakness if decision processes are to be derived on the basis of the propagation of these values. The conclusion drawn in this research is that the use of a mathematical function which allows the opinion distribution to be defined as a shape (rather than just a mathematical form) removes the requirement for arbitrary evaluation of numerical values of opinion. The numerical values required for propagation are derived from the shape by means of the opinion model. This also accords with the conclusion mentioned above in this section, that opinion is perceived in terms of a shape, rather than in a numerical form.

Fuzzy set theory attempts to use mathematical functions in order to derive support values for qualified variables. This is felt to be unsatisfactory, and a conclusion of this research is that the relationship between two descriptors which refer to different aspects of the same variable is semantic but is not necessarily mathematical. Accordingly the use is advocated of a set of opinion distributions which have the same mathematical form, but which have no functional relationship with each other.

Expertise, including the opinion distributions, is acquired from the design engineers using the medium of interviews. These may take place at a variety of locations including the sites where high capacity bus systems have been implemented.

A major conclusion of this research is that the view, commonly held in the artificial intelligence community, that knowledge acquisition is best undertaken by a knowledge engineer with no domain knowledge, should be questioned in this context. The knowledge acquisition process is much easier and more efficient if the acquirer of the expertise is familiar with the domain. This facilitates communication between the expert and the acquirer of the knowledge, and relieves the expert of the difficulties found in communicating complex ideas at the level of a novice. The knowledge acquirer therefore has the responsibility of ensuring that the complete expertise is encapsulated within the model at the appropriate level. In order to ensure that this is done successfully, the expertise is calibrated with the expert before use within the decision model.

The calibrated expertise is encapsulated within a computer program with the objective of modelling the decision processes of the design engineers in order to produce an initial design for a high capacity bus system in a corridor.

The performance of the model is tested using data from a corridor in Brazil in which the design engineers from whom the expertise was obtained were involved. This validation exercise shows that the decision model produces a similar design to that of the experts when given the same data.

A note of caution is, however, necessary. One result of the validation exercise is that the stop location model used in the design model is different from that used in the original design of the high capacity system concerned. This is because there is a time lag between the use of expertise by the expert during the course of a particular project, and the subsequent acquisition of the expertise for the decision model. Possibly as a result of the effects of the implementation of their design, the stop location model used in the design of the Santo Amaro corridor is not the model conveyed during the knowledge acquisition process. This implies that a knowledge acquirer is likely to be at a disadvantage when testing expert's expertise against previous work of the expert. Expertise includes experience gained throughout the working life of the expert and is subject to continual evaluation and reconsideration. It is therefore unlikely to be consistently the same before and after the implementation of a particular project.

The model was then tested in less favourable conditions, in London and Lima. In each case, the model identified the difficulties and produced suboptimal designs where this was possible. These designs were measured against the expertise in order to check that the conclusions were indeed compatible with the expertise. These tests were severe: the suboptimal designs were compared with the optimum design values. The model correctly identified the suboptimal designs which were not tolerable to the expertise.

The opinion model is able to operate with imprecise data, and indicates, where appropriate, the limits within which a decision which is bifurcated on the basis of imprecise inputs is viable.

It is therefore concluded that the use of the opinion model, as formulated during this research, is a reasonable representation of the interpretative aspects of decision making within the context of the design of high capacity bus systems.

10.3 EVALUATION OF THIS RESEARCH

This research has considered in some detail various aspects of bus operation, especially under high capacity operating conditions. The scope of the study has not been such that it has been possible to analyze all the characteristics of the urban transport system which may be affected (and be affected by) the operation of a high capacity bus system.

Such an analysis is an essential feature of the appraisal of infrastructure designs of the type discussed in this thesis. Such analysis could be incorporated within the modular structure of the computer program, through the addition of appropriate modules. Before this is possible, however, it is necessary to undertake the research required in order to analyze these effects and to develop appropriate models.

Similarly, this research has made no attempt to establish the costs and benefits of different designs. The decision to implement a high capacity bus system is generally likely to be taken on the basis of a cost - benefit analysis, and therefore the design process should include evaluations of the design for this purpose. As with the analyses discussed in the preceding paragraph, a module constructed to perform this function could be included once the essential research is undertaken.

The design model does, however, produce an initial design for a corridor (if this is considered desirable) and indicates the appearance of the design and the degree to which any suboptimality is considered to be a problem. This is the starting point from which all of these analyses must proceed. The research reported in this thesis is therefore an initial attempt to analyze the design of infrastructure for high capacity bus systems. This should be pursued through a number of areas for further research which have been identified during this research.

10.4 RECOMMENDATIONS FOR FURTHER RESEARCH

The reductive effects on capacity which arise as a result of an increase in the number of berths at a bus stop are recognized, but there is currently no robust model which can estimate their causes, magnitude or nature.

The models identified during this research which estimate the capacity of a bus stop are useful, but only represent the situation at a single bus stop. The effects on the efficiency of a bus stop which may be caused by the operation at a previous stop are not currently considered. There is scope for further investigation into these effects.

The headway between bus arrivals at bus stops is represented in the current models, either as a constant or as a stochastic process where the particular headway is chosen from a distribution. The deterministic effects resulting from interactions between buses (and passengers and buses) may be better represented directly within a modelling framework. There is therefore a need for further research into the possibility for the representation of bus operation using deterministic models.

Passenger arrivals at bus stops are also commonly represented as a Poisson process. At critical stops, however, the arrivals of passengers may be determined, for example, by the arrival of other buses or trains. The interaction between these non-random arrivals and the arrivals of buses is not currently represented in model form, and this is also an area for further research.

A simulation model which is capable of representing the subtle changes in design of bus stops and bus systems would facilitate evaluation of such designs prior to implementation. The usefulness of such a model is clear, but it requires further research in the above areas before it would be able to represent the effects of these changes within the bus system.

It is clear that the evaluation of bus system performance is still a matter for further research. This includes aspects of performance in terms of operation - such as a capacity measure which is more comprehensive than the line-haul capacities usually used to describe system capacity. The capacity of a bus system is also a function of its capacity to allow boarding and alighting, and this should be included in any such measure. A measure thus constructed would therefore include aspects of system design as well as operational performance, and give accessibility the prominence it deserves in the evaluation of transport systems.

The satisfactory combination of differing expertise is still elusive, and further research should be continued in this area in order to increase the usefulness of knowledge-based systems.

The opinion model, as formulated in this research, is required to perform a large number of calculations in order to reach a decision. It is correspondingly slow. While this is of little importance in a design process, it would be useful to be able to use an opinion model in the course of decision-making in real time. This requires further work in order to establish how the number of calculations might be reduced in order to reduce the processing time.

The evaluation of alternative hypotheses is simplistic in the model as described in this thesis. There is a potential benefit to be gained from a more comprehensive treatment of the relationship between the preferred solution and the nearest alternatives.

The ability of the opinion model to represent decision processes where data are scarce, or imprecise (or both) is registered in this research. There is scope for further work to evaluate its use in other areas of transport research.

Further work would also be interesting to investigate the level of data quality that it is desirable to require under various circumstances. This could have implications for the design and implementations of surveys.

The representation of aspects of judgement in decision processes has a potential importance within transport research. This is because many decisions are taken on such a basis in the real world. This research has shown that it is possible to approach directly this aspect of the decision process. The model discussed in this research is a first step in this direction, and is necessarily simple. An important area for further research is therefore to investigate improvements to this model. Questions for consideration in this respect include:-

is the β distribution the best form for the representation of opinion?

is it possible to replace the use of a set of distributions with a single (more complex) distribution which could represent the entire range of options?

could the evaluation of the opinion of data distributions be performed more effectively by the use of methods other than the use of composition (for example to consider the use of integrals rather than the minimax formulation used in composition)?

is the assumption viable that the computer-based representation of the user's opinion of data is numerically equivalent to the computer-based representation of the expert's opinions (for example, should the constraint that the mode should equal 1 be retained)?

10.5 FINAL REMARKS

The decision model described in this thesis is intended to demonstrate that it is possible to represent in model form decisions based on qualitative assessments of data, of the outputs of quantitative models, and of other qualitative decisions. This model therefore models decisions based upon the use of judgement or opinion. This approach was adopted because during the course of the research it became apparent that several of the decisions taken during the design process for a high capacity bus system are based upon such factors.

The model attempts to represent the decision process of the engineers, rather than trying to establish quantitative relationships between the variables which describe the bus system. This model operates with interpretations of data, and therefore addresses directly the inherent imprecision. It therefore concentrates on the interpretation of the outputs rather than the outputs themselves. Another area for further research is therefore to establish what opportunities this change of focus in the decision modelling process might offer in terms of the modelling of other processes.

REFERENCES

ALLEN M.K. (1986) The development of an artificial intelligence system for inventory management using multiple experts. Phd Dissertation. Unpublished. Ohio State University USA

ALLSOP R.E (1990) Private correspondence

ALONSO W. (1968) Predicting best with imperfect data. Journal of the American Institute of Planners, July.

ARMSTRONG-WRIGHT A. (1986) Urban Transit Systems. Technical Paper 52. The World Bank Washington D.C., USA

ARMSTRONG-WRIGHT A.,S. THIRIEZ (1987) Bus Services. Technical Paper 68. The World Bank Washington D.C., USA

BAEZA I., J. GIBSON (1989) Modelación de la capacidad y las demoras en paraderos de buses. Proc. IV Congreso chileno de ingenieria de transporte. Valparaiso, Chile, October

BARON P. (1987) Der Spurbus in Essen - ein zukunftsweisendes System für den öffentlichen Personennahverkehr. Internationales Verkehrswesen, 39(2) March/April 105-109

BAYLISS D. (1981) One billion new city dwellers - how will they travel? Transportation, 10, 311-343

BELLMAN R.E., L.A. ZADEH (1970) Decision-making in a fuzzy environment. Management Science, 17B. 141-164.

BEIER G., A. CHURCHILL, M. COHEN, B. RENAUD (1976) The task ahead for the cities of developing countries. World Development, 4(5), 363-409

BODEN M.A. (1987) Artificial Intelligence and Natural Man. MIT Press. London, UK

BRACHMAN R.J., S.AMAREL, C.ENGELMAN, R.S. ENGELMORE, E.A. FEIGENBAUM, D.E. WILKINS (1983) In Building Expert Systems (eds Hayes-Roth et al). Addison-Wesley. Reading, Mass, USA

BRUNDTLAND, G.H., M. KHALID, S. AGNELLI, A.A. AL-ATHEL, B. CHIDZERO, L.M. FADIKA *et al* (1987) Our Common Future. World Commission on Environment and Development, Oxford University Press, Oxford, UK

BUCHANAN B.G., E. SHORTLIFFE (1985) Rule-based expert systems. Addison-Wesley. Reading, Mass, USA

CADE-IDEPE (1988) Estudio de investigación de metodología de análisis y seguimiento de transporte público. Informe Final, Intendencia de la región Metropolitana, Santiago, Chile (unpublished)

CAGAN J., A.M. AGOGINO (1987) Innovative design of mechanical structures from first principles. Artificial Intelligence for engineering design, analysis and manufacturing. Academic Press. **1**(3). 169-189.

CHADWICK G. (1987) Models of urban and regional development in developing countries. Pergammon. Oxford, UK

CHAPMAN J.A., H.E. GAULT, I.A. JENKINS (1976) Factors affecting the operation of urban bus routes. Working Paper 23. Transport Operations Research Group, University of Newcastle upon Tyne, Newcastle, UK

COHEN S. (1984) Indicateurs d'allure et de consommation d'un autobus en exploitation. Recherche Transportes Sécurité, Janvier.

COHEN P.R., G. SHAFER, P.P SHENOY (1987) Modifiable combining functions. Artificial Intelligence for engineering design, analysis and manufacturing. Academic Press. 1(1). 47-57.

CUNDILL M.A., P.F.WATTS (1973) Bus boarding and alighting times. Laboratory Report 521. Transport and Road Research Laboratory. Crowthorne, UK

CMTC (1984) Relatorio: Projeto Trolebus Santo Amaro/Bandeira. Companhia Municipal de Transportes Coletivos. Sao Paulo, Brazil

CMTC (1985) Relatorio Técnico: Corredor Rio Branco/Emílio Carlos. Companhia Municipal de Transportes Coletivos. Sao Paulo, Brazil

DANAS A. (1983) Bus passengers' walking distances, arrivals at bus stop and waiting times. PhD Thesis, University of London (unpublished)

DREYFUS H.L., S.E. DREYFUS (1986) Mind over machine. Blackwell. Oxford, UK

DTp (1976) Guidelines for the implementation of bus lanes H6/76. Department of Transport, London UK

Dtp (1988) Priority for buses. Traffic Advisory Unit Leaflet 2/88. Department of Transport, London, UK

EFSTATHIOU J., V. RAJKOVIC (1979) Multiattribute decisionmaking using a fuzzy heuristic approach. IEEE Transactions on Systems, Man and Cybernetics, 9(6), 326-333.

EBTU (1982) Tratamento preferencial ao transporte coletivo por ônibus. Empresa Brasileira dos Transportes Urbanos, Brasília, Brazil

EBTU (1988) The standardisation of city buses (in English). Empresa Brasileira dos Transportes Urbanos. Brasilia, Brazil

FEIGENBAUM E.A. (1977) The art of artificial intelligence1: themes and case studies of knowledge engineering. Report: STAN-CS-77-621. Department of Computer Science, Stanford University, California, USA

FERNANDEZ R. (1990) Expert systems: application to the location of high capacity bus stops. MSc Thesis, University of London (unpublished)

FOURACRE P.R., A.C. MAUNDER, M.G. PATHAK, C.H. RAO (1981). Laboratory Report 1018. Transport and Road Research Laboratory. Crowthorne, UK

GAINES B.R., J.H. BOOSE (1990) Machine learning and uncertain reasoning. Academic Press. London, UK

GIBSON J. (1990) Private Correspondence

GIBSON J. (1991) Private Correspondence

GIBSON J., I. BAEZA, L.G. WILLUMSEN (1989a) Bus-stops, congestion and congested bus stops. Traffic Engineering and Control, 30(6), 291-296.

GIBSON J., I. BAEZA, R. BECKETT (1989b) IRENE: un programa de simulación de paraderos. IV Congreso latinamericano de transporte público y urbano. La Habana, Cuba, September

GREENSTEIN J., L. BERGER, A. STRULOV (1991) Public transportation problems and solutions in the historical center of Quito. Transportation Research Record 1266. Transportation Research Board, Washington DC, USA

GUENTHNER R.P., K.C. SINHA (1983) Modeling bus delays due to passenger boardings and alightings. Transportation Research Record 915. Transportation Research Board, Washington D.C. USA

HARDOY J.E., D. SATTERTHWAITE (1989) Squatter Citizen: life in the urban third world. Earthscan publications, London, UK

HAMBURG M. (1985) Basic Statistics: a modern approach. Harcourt Brace Jovanovich. Orlando, Florida, USA

HAYES-ROTH F., D.A. WATERMAN, D.B. LENAT (1983) An overview of Expert Systems. In Building Expert Systems (eds Hayes-Roth F., D.A. Waterman, D.B. Lenat). Addison-Wesley. Reading, Mass, USA

HENDRICKSON C. C. ZOZAYA-GOROSTIZA, S. MCNEIL (1987) A knowledge-based expert system architecture for computer-aided analysis and design of intersections. In Tranportation and traffic Theory (Eds N.H. Gautner, N.H.M. Wilson). Elsevier, New York, USA.

HOUNSELL N.B., M, MCDONALD (1988) Evaluation of bus lanes. Contractor Report 87. Transport and Road Research Laboratory. Crowthorne, UK

HUTTON B.J., J.M. CLARK (1989) Monitoring minibus services in Milton Keynes. Proceedings of 21st Conference of the Universities' Transport Studies Group. Edinburgh, UK

IUNES F.F., L.G. WILLUMSEN (1988) The computer-assisted design of with-flow bus lanes with BLISS. Traffic Engineering and Control, 29(7). 384-391.

INVERMET (1990) Informe de encuestas y conteos de transporte y trafico: Proyecto Avenida Arequipa, Avenida du Petit Thouars y Avenida Arenales. Municipalidad de Lima Metropolitana

JACKSON P. (1986) Introduction to Expert Systems. Addison-Wesley. Wokingham, UK

JACKSON R.L. (1976) An evaluation of bus control systems by simulation. Laboratory Report 712. Transport and Road Research Laboratory, Crowthorne, UK

JAMIESON MACKAY & PARTNERS (1988) Bangkok Traffic Research Project. Contractor Report 84. Transport and Road Research Laboratory. Crowthorne, UK

JESSOP A.T. (1990) Decision and forecasting models with transport applications. Ellis Horwood. London UK.

JOWITT P.W. (1984) Fuzzy sets and water quality. IFAC Triennial Congress, Budapest, Hungary.

JOWITT P.W., J.P. LUMBERS (1983) Water quality objectives, discharge standards and fuzzy logic. Optimal Allocation of Water Resources Symposium. Exeter, UK.

KIKUCHI S. (1985) Relationship between the number of stops and headway for a fixed-route transit system. Transportation Research, 19A(1), 65-71.

KIKUCHI S., V.R. VUCHIC (1982) Transit vehicle stopping regimes and spacings. Transportation Science, 16(3)

KIMBER R.M., P.N. DALY (1986) Time-dependent queuing at road junctions: observation and prediction. Transportation Research, 20B(3), 187-203.

KRAFT W.H., T.J. BOARDMAN (1972) Location of bus stops. Transportation Engineering Journal, 98(1), 103-116.

KULIKOWSKI C.A. (1980) Artificial Intelligence methods and systems for medical consultation. IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-2. 464-475.

LAMBERTI D.M., S.L. NEWSOME (1990) Presenting abstract versus concrete information in expert systems: what is the impact on user performance? In Machine Learning and Uncertain Reasoning (Gaines B.R., J.H. Boose (Eds). Academic Press. London, UK

LANE R., D.H. HODGKINSON (1973) Evaluating a network of bus segregation schemes. PTRC New Year Meeting, January.

LESLEY L.J.S. (1976) Optimum bus stop spacings: Part 1. Traffic Engineering and Control October

LEVINSON H.S., C.L. ADAMS, W.F. HOEY (1976) Bus use of highways: planning and design guidelines. National Cooperative Highways Research Program Report 155. Transportation Research Board. Washington DC, USA

LINDAU L.A. (1983) High-flow bus operation on urban arterial roads. PhD Thesis, University of Southampton (Unpublished)

LINDAU L.A. (1987) Bus priority systems in Brazil: from theory to practice. PTRC Summer Annual Meeting Seminar J. 89-101. September. Bath UK

LINDAU L.A. (1989) Private correspondence

LINDAU L.A., L.G. WILLUMSEN (1988) How far can bus priority be stretched? Proceedings CODATU IV. June. Jakarta, Indonesia

MAHER M.L. (1987) Engineering design synthesis: a domain independent representation. Artificial Intelligence for engineering design, analysis and manufacturing, 1(3). 207-213.

MAHER M.L. (1989) Expert systems for engineering design. Advanced Study Institute: Optimization and decision support systems in Civil Engineering. North Atlantic Treaty Organization. June/July. Edinburgh, UK

MARLER N.W. (1982) The performance of high-flow bus lanes in Bangkok. Supplementary Report 723. Transport and Road Research Laboratory. Crowthorne, UK

MCKEITHEN K., J.S. REITMAN, H. REUTER, S.C. HIRTLE (1981) Knowledge organization and skill differences in computer programmers. Cognitive Psychology, 13. 307-325.

OLDFIELD R.H., P.H. BLY (1985) Using smaller buses in Central London. Research Report 31. Transport and Road Research Laboratory, Crowthorne, UK

OLDFIELD R.H., P.H. BLY, F.V. WEBSTER (1977) With-flow bus lanes: economic justification using a theoretical model. Laboratory Report 809. Transport and Road Research Laboratory. Crowthorne, UK

PAWLAK Z., S.K.M WONG, W ZIARKO (1990) Rough sets: probabilistic versus deterministic approach. In Machine Learning and Uncertain Reasoning (Gaines B.R., J.H. Boose (Eds). Academic Press. London, UK

PRETTY R.L., D.J. RUSSELL (1988) Bus boarding rates. Australian Road Research, 18(3).

RAMSEY F.P. (1926) Truth and probability. In Foundations: Essays in Philosophy, Logic, Mathematics and Economics (ed D.H. Mellor 1978). Routledge & Kegan Paul, London, UK

ROARKE J.J. (1982) Enforcement of priority treatment for buses on urban streets. Synthesis of Transit Practice 2. National Cooperative Transit Research and Development Program. Washington DC, USA

ROYAL BOROUGH OF KINGSTON-UPON-THAMES (1990) A240 Route Study. Directorate of Engineering and Transportation.

SCHEEL J.W., J.E. FOOTE (1968) Bus operation in single lane platoons and their ventilation needs for operation in tunnels. Research Publication GMR-808. General Motors Corporation, Warren, Michigan, USA

SCHEEL J.W., J.E. FOOTE (1969) Comparison of experimental results with estimated single lane bus flow through a series of stations along a private busway. Research Publication GMR-888. General Motors Corporation, Warren, Michigan, USA

SHALIN V.L., E.J. WISNIEWSKI, K.R. LEVI, P.D. SCOTT (1990) A formal analysis of machine learning systems for knowledge acquisition. In Machine Learning and Uncertain Reasoning (Gaines B.R., J.H. Boose (Eds). Academic Press. London, UK

SHANNON C. (1948) The mathematical theory of communications. Bell Systems Technical Journal July/October

SHAPIRO A.D. (1987) Structured induction in expert systems. Addison-Wesley. Wokingham, UK

SHORTLIFFE E. (1976) Computer-based Medical Consultations: MYCIN. Elsevier, New York, USA

SPIEGELHALTER D.J (1986) A statistical view of uncertainty in expert systems. In Artificial Intelligence and Statistics (Ed W. Gale). Addison-Wesley, Reading, Mass. USA

STEFIK M., J. AIKINS, R BALZER, J. BENOIT, L. BIRNBAUM, F. HAYES-ROTH, E. SACERDOTI (1983) Basic Concepts for building expert systems. In Building Expert Systems (Eds Hayes-Roth F, D.A. Waterman, D.B. Lenat). Addison-Wesley, Reading, Mass. USA

SZASZ P.A. (1980) Private Correspondence

SZASZ P.A.(1990) Private Correspondence

SZASZ P.A. (1991) Private Correspondence

SZASZ P.A., L. de C. MONTANA, E.O. FERREIRA (1978) COMONOR: Ordinated bus convoy. Technical paper 9. Companhia de Engenharia de Tráfego, Sao Paulo, Brazil

SZASZ P.A., E. GERMANI (1985) Comboio de ônibus ordenado COMONOR-sp. Empresa Brasileira dos Transportes Urbanos. Brasilia, Brazil

SPM (1981) Monitoramento - Porto Alegre Relatório final - Avaliaçao "ex-post" pré-corredores Farrapos e Assis Brasil. Secretaria do Planajemento Municipal, Porto Alegre, Brazil

THAYSE A. (Ed) (1988) From standard logic to logic programming. John Wiley and Sons, London, UK

THOMAS V. (1980) Spatial differences in poverty. Journal of Development Economics, 7, 85-98

TRRL (1990) Study of bus priority systems. Interim Report (unpublished)

TRANSPORTATION RESEARCH BOARD (1985) Highway Capacity Manual. Special Report 209, Washington D.C., USA

Transporte Moderno (Reinaldo Andrade) (1990) Onibus ganham portas á esquerda. Transporte Moderno, 28 November.

TOMII N, S SHINCHARA, N. SATO (1989) Knowledge enhancement through word-processor-like interface in railway traffic control systems. Proceedings Control, Computers, Communications in Transportation, Paris, September

TYLER N.A. (1987) Decisions: complexity and precision, or simply magic numbers? MSc Thesis, Polytechnic of Central London, (Unpublished)

VUCHIC V.R. (1969) Rapid transit interstation spacings for maximum number of passengers. Transportation Science, 3(3)

VUCHIC V.R (1981) Urban public transportation systems and technology. Prentice- Hall. New Jersey, USA

VUCHIC V.R., G.F. NEWELL (1968) Rapid transit interstation spacings for minimum travel time. Transportation Science, 2(4)

WAYTE F.A., T.J. WILSON (1989) Guided bus rapid transit - design and operating experience, Adelaide. Private correspondence.

WEBSTER F.V., P.H. BLY (Eds) (1976) Bus priority systems. NATO Committe on the Challenges of Modern Society 45. Transport and Road Research Laboratory, Crowthorne, UK

WHITE P.R (1986) Public Transport: its planning and management. Pergammon. London, UK.

WINSTON P.H. (1984) Artificial Intelligence. Addison-Wesley. Reading, Mass, USA

WITTEN I.H., B.A. MACDONALD (1990) Using concept learning for knowledge acquisition. In Machine Learning and Uncertain Reasoning (Gaines B.R., J.H. Boose (Eds). Academic Press. London, UK

ZADEH L.A. (1965) Fuzzy Sets. Information and Control. 8. 338-353.

ZADEH L.A. (1975a) The concept of a linguistic variable and its applications to approximate reasoning. Part 1. Information Sciences, 8. 199-249.

ZADEH L.A. (1975b) The concept of a linguistic variable and its applications to approximate reasoning. Part 2. Information Sciences, 8. 301-357.

ZADEH L.A. (1975c) The concept of a linguistic variable and its applications to approximate reasoning. Part 3. Information Sciences, 9. 43-80.

ZOGRAFOS K.G., H.S. LEVINSON (1986) Passenger service times for a no-fare bus system. Transportation Research Record 1051. Transportation Research Board, Washington D.C. USA

ZWICK R., T.S. WALLSTEN (1990) Combining stochastic uncertainty and linguistic inexactness: theory and experimental evaluation of four fuzzy probability models. In Machine Learning and Uncertain Reasoning (Eds Gaines B.R., J.H. Boose).

APPENDIX A

DATA: SANTO AMARO / 9 DE JULHO CORRIDOR

A1 LINK WIDTHS

TRECHO	LARGURA (m)
entre R.Joaquim Guarani e R.Andrea Paulinetti	27,00
entre R.Andrea Paulinetti e Av.Roque Petroni Jr.	27.40
entre Av.Roque Petroni Jr. e R.Eleutério	27,10
	26,50
entre R.Eleutério e R.Joaquim Nabuco	27,00
entre R.Joaquim Nabuco e R.Bernardino de Campos	27,70
	26,50
entre R.Bernardino de Campos e R.Pedro Taques	26,50
	26,40
entre R.Pedro Taques e R.Bavas	27,00
entre R.Bavas e R.Frei Gaspar	27,10
entre R.Frei Gaspar e R.Edson	26,50
	27,00
entre R.Edson e R.Piracicaba	27,00
entre R.Piracicaba e R.Vieira de Morais	27,00
entre R.Vieira de Morais e R.Machado de Assis	26,50
entre R.Machado de Assis e R.Jesuino Maciel	27,00
entre R.Jesuino Maciel e R.Rui Barbosa	25.10
entre R.Rui Barbosa e R.Morais Barros	25,50
entre R.Morais Barros e R.Giorgia	27,40
entre R.Giorgia e R.Novo Mundo	26,50
entre R.Trauna e R.Jurema	24,50
entre R.Jurema e Av. Eucaliptos	23,00
entre Av.Eucaliptos e Rua Cotovia	24,20
entre R.Cotovia e Av Bem te Vi	24,50
entre R.Bem te Vi e R.Carajua	25,10
entre R.Carajua e R.Pavão	22,90
entre R.Pavão e R.Rouxinol	23,20
entre R.Rouxinol e R.Grauna	22,70
entre R.Grauna e R.Periquito	27,60
	28,20
entre R.Periquito e Av. Uberaba	28,70
entre Av. Uberaba e R.Silvania	30,00

TRECHO	LARGURA (m)
entre R.Natividade e R.Afonso Brás	28,80
entre R.Afonso Brás e R.Baltazar da Veiga	36,50
entre R.Baltazar da Veiga e R.Brás Cardoso	26,80
entre R.Brás Cardoso e R. Dom.Fernandes	24,50
entre R.Dom.Fernandes e R.Bueno Brandão	22,80
entre R.Bueno Brandão e R.João Lourenço	22,60
entre R.João Lourenço e R.Filadelfo Azevedo	23,10
	22,80
entre R.Bastos Pereira e Av.Antonio Joaquim de Moura Andrade	23,40
AV. SÃO GABRIEL	
entre Av.Brig.Luiz Antonio e R.Gal.Mena Barreto	34,80
entre R.Gal.Mena Barreto e R.Sarita Cyrillo	34,60
entre R.Sarita Cyrillo e R.Primavera	27,00
entre R.Primavera e R.Gironda	27,00
entre R. Gironda e R. Suzano	27,00
entre R. Suzano e R.Maestro Elias Lobo	26,80
AV.9 DE JULHO	
entre R.Groenlândia e R.Gal.Fonseca Teles	29,60
entre R.Fonseca Teles e R.Antônio Bento	29,60
entre R. Espéria e Av. Brasil	29,90
entre Av.Brasil e R.Estados Unidos	30,30
entre R.Estados Unidos e R.Caconde	29,60
entre R.Caconde e R.V. Calfat	27,60
entre R..V. Calfat e Al. Lorena	29,60
	29,20
entre Al.Lorena e R.José Maria Lisboa	29,70
	30,20
entre R.José Maria Lisboa e Al. Franca	30,00
entre Al.Franca e Al. Itú	30,00
entre Al. Itú e Al. Jaú	29,60
	35,50
entre R.São Carlos do Pinhal até Pça 14 Bis	30,60
	29,10
	42,50*

TRECHO	LARGURA (m)
continuação	
entre R.São Carlos do Pinhal até a Pça.14 Bis	60,20*
	68,00*
	68,60*
	69,40*
entre Pça 14 Bis e R. 14 de Julho	37,20*
	38,00*
	37,40*
entre R.14 de Julho e R.Major Quedinho	28,40
	30,00
entre R.Major Quedinho e Pça.da Bandeira	31,00
	30,00

*- Larguras da Pça.14 Bis

A2 TRAFFIC VOLUMES IN THE PEAK HOUR

	a	b	c	v.e.
Av. Adolfo Pinheiro x R.Isabel Schmidt (B-C)	755	366	15	1547
Av. Adolfo Pinheiro x R.9 de Julho (B-C)	2272	339	25	3050
Av. Adolfo Pinheiro x R.Liberdade (B-C)	2441	424	53	3501
Av. Adolfo Pinheiro x R.Fraternidade (B-C)	1478	321	14	2176
Av. Santo Amaro x R.Verbo Divino (C-B)	1572	366	39	2460
Av. Santo Amaro x R. Bela Vista (C-B)	475	7	19	565
	2676	340	56	3580
Av. Santo Amaro x R.São Sebastião (B-C)	1367	279	14	1981
Av. Santo Amaro x R.Joaquim Nabuco (C-B)	1885	233	26	2435
Av. Santo Amaro x R.Joaquim Nabuco (B-C)	1819	271	15	2421
Av. Santo Amaro x R.Vieira de Morais (C-B)	2038	294	40	2786
Av. Santo Amaro x R.Vieira de Morais (B-C)	2100	272	18	2716
Av. Santo Amaro x Av. Cotovia (C-B)	1984	265	34	2650
Av. Santo Amaro x Av. Cotovia (B-C)	2126	271	27	2776
Av. Santo Amaro x R.Casa do Ator (C-B)	1498	257	36	2156
Av. Santo Amaro x R.Casa do Ator (B-C)	2157	307	21	2855
Av. Santo Amaro x R.Fiandeiras (C-B)	1588	263	46	2298
Av. Santo Amaro x R.Fiandeiras (B-C)	2057	294	36	3189
Av. Santo Amaro x R.Baltazar da Veiga (C-B)	2505	298	65	3361
Av. Santo Amaro x R.Baltazar da Veiga (B-C)	1638	311	8	2292
Av. Santo Amaro x R.João Lourenço (C-B)	1332	283	31	2022
Av. Santo Amaro x R.João Lourenço (B-C)	1756	299	17	2422
Av. Santo Amaro x Av. Joaquim de Moura Andrade (B-C)	940	184	14	1364
Av. Santo Amaro x Av. Joaquim de Moura Andrade (C-B)	133	82	3	309
Av. Santo Amaro (sobre o túnel)(C-B)	635	198	2	1039
Av. Santo Amaro (sobre o túnel)(B-C)	1785	140	24	2161
Av. Santo Amaro (Pça D.Gastão L.Pinto)(B-C)	1308	105	29	1634
Av. São Gabriel x Av.9 de Julho (C-B)	1177	244	16	1729
Av. São Gabriel x Av.9 de Julho (B-C)	897	163	3	1235
Av. 9 de Julho x Av.Brasil (C-B)	1999	172	8	2375
Av. 9 de Julho x Av.Brasil (B-C)	1952	217	11	2430
Av. 9 de Julho x Pça das Guianas (C-B)	1883	267	24	2513
Av. 9 de Julho x Pça das Guianas (B-C)	1978	353	5	2704

	a	ô	c	v.e.
Av. 9 de Julho x R.José Maria Lisboa (C-B)	1950	340	25	2730
Av. 9 de Julho x R.José Maria Lisboa (B-C)	499	-	5	519

a = automóveis

ô = ônibus

c = caminhões

v.e. = veículos equivalentes (1 x a + 2 x ô + 4 x c)

B-C = Bairro/Centro

C-B = Centro/Bairro

A3 BUS STOP LOCATION AND PASSENGER FLOWS

NÚMERO DO PONTO	LOCALIZAÇÃO	VOLUME DE PASSAGEIROS/HORA EMBARQUE	DESEMBARQUE
37	Largo 13 de Maio Av. Adolfo Pinheiro nº 384	2341	1690
36	Av. Adolfo Pinheiro (Pça Santa Cruz)	549	424
35	Av. Adolfo Pinheiro nº 810	553	420
34	Av. Adolfo Pinheiro nº 1136	773	613
33	Av. Adolfo Pinheiro nº 1362	437	371
32	Av. Adolfo Pinheiro nº 1500	151	370
31	Av. Adolfo Pinheiro nº 1912	115	374
30	Av. Adolfo Pinheiro nº 2100	69	462
29	Av. Adolfo Pinheiro nº 2546	134	493
28	Av. Santo Amaro nº 5461	82	229
27	Av. Santo Amaro nº 5137	71	225
26	Av. Santo Amaro nº 4693	63	484
25	Av. Santo Amaro nº 4281	54	471
24	Av. Santo Amaro nº 4061	405	993
23	Av. Santo Amaro nº 3785	374	476
22	Av. Santo Amaro nº 3496	217	742
21	Av. Santo Amaro nº 3105	151	420
20	Av. Santo Amaro nº 2283	155	773
19	Av. Santo Amaro nº 2027	119	564
18	Av. Santo Amaro nº 1645	121	481
17	Av. Santo Amaro nº 1327	41	538
16	Av. Santo Amaro nº 1057	97	437
15	Av. Santo Amaro nº 785	381	804
14	Av. Santo Amaro nº 441	31	645
13	Av. Santo Amaro nº 281	200	622
12	Av. São Gabriel nº 344	57	623
11	Av. São Gabriel nº 626	37	289
10	Av. 9 de Julho nº 4241	72	192
09	Av. 9 de Julho nº 4015	13	373
08	Av. 9 de Julho nº 3789	128	779

NÚMERO DE PONTO	LOCALIZAÇÃO	VOLUME DE PASSAGEIROS/ HORA	
		EMBARQUE	DESEMBARQUE
07	Av. 9 de Julho nº 3553	186	1663
06	Av. 9 de Julho - antes do Túnel	265	1303
05	Av. 9 de Julho em frente à G.V.*	113	1195
04	Av. 9 de Julho nº 1737	181	635
03	Av. 9 de Julho nº 1217	262	281
02	Av. 9 de Julho (INPS)	103	432
01	Av. 9 de Julho nº 359	97	736
00	Praça da Bandeira	13	3709

* G.V. : Escola de Administração de Empresas Getúlio Vargas

APPENDIX B

DATA USED IN VALIDATION EXERCISE

BUSES

LENGTH 13 metres

DOORS 2

CAPACITY 60 passengers

FLOW

minimum: 200

maximum: 450

preferred: 300

BUS STOPS

LOCATION

1 to 8 located between junction 1 and junction 2

9 to 27 located between junction 2 and junction 3

BOARDING PASSENGERS

LINK 1

minimum: 1

maximum: 3

preferred: 2

LINK 2

minimum: 1

maximum: 5

preferred: 3

LINKS

LENGTH

link 1: 2618 metres

link 2: 3129 metres

WIDTHS

link 1: 27.4 metres

link 2: 30.3 metres

JUNCTIONS

Junction	Arm	stop line		entry line		central reserve	footway
		lanes	width	lanes	width		width
1	**1**	4	14	4	14	1	2.7
1	**2**	4	14	4	14	1	2.7
1	**3**	2	7	2	7	0	2
2	**1**	4	14	4	14	1	2.7
2	**2**	4	14	4	14	1	2.7
2	**3**	2	7	2	7	0	2
3	**1**	4	14	4	14	1	6.4
3	**2**	4	14	4	14	1	2.4
3	**3**	3	10.5	3	10.5	3	2

APPENDIX C
KNOWLEDGE ACQUISITION PROCEDURE

C1 INTRODUCTION

This Appendix contains details of the experts consulted during this research. The next section consists of a list of the experts consulted during this research with an indication of the area of expertise in which their contribution was sought. Section C3 describes the selection criteria for the experts and some comments for consideration in such exercises in the future. Section C4 describes the knowledge acquisition procedure. Section C5 is a basic list of the variables and parameters which were discussed with the experts during this research.

C2 EXPERTS CONSULTED AND THEIR AREAS OF INTEREST

1 Jaime Gibson Aldunate, Universidad de Chile, Santiago, Chile: high capacity bus systems, in particular the capacity of bus stops in such situations and the effects of different bus stop designs on the capacity;

2 Vicente Pardo Díaz: Coordinador Técnico, Comisión de Transporte Urbano, Santiago, Chile: Operating systems with unregulated bus operators;

3 Luis Antonio Lindau Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil: analysis of COMONOR operations;

4 Resiere Pavanelli Filho Sindicato das Empresas de Transportes de Passageiros do Município do Rio de Janeiro, Brazil: bus operations and operators' needs;

5/6 Luiz Carlos S. Neves Pereira, Olóvis Emílio Costa Nogueira: Empresa Brasileira de Transporte Urbano, Brazil: the design of buses for high capacity;

7 Isaac Popoutchi: Gerente de Planajamento e Projeto de Transporte Metropolitano, Companhia do Metropolitano de Sao Paulo, Sao Paulo, Brazil and members of the design team: bus system design in segregated median corridors;

8/9 Sandra Maria Ratton, Joao Carlos Cascaes (Diretor de Planajamento e Desenvolvimento): Urbanizaçao de Curitiba S.A, Brazil: Bus operations in Curitiba;

10 Antonio Luiz Mourao Santana, Diretor técnico, KGS Engenharia (Consulting Company), Sao Paulo Brazil and various members of the design team: Projeto Trolebus (Santo Amaro/9 de julho corridor (CET 1984); and

11 Pedro Alvaro Szász Independent Consultant, Sao Paulo, Brazil (formerly Technical Director of Companhia de Engenharia de Tráfego, Sao Paulo): Design of COMONOR system; design of high capacity bus systems; bus operations at bus stops.

C3 SELECTION OF EXPERTS

The selection of the experts is important to the whole knowledge acquisition process. There is no predetermined method for the selection of experts, so the choice must be made on the basis of the criteria determined by the problem at hand. For this research, it was necessary to select experts who had experience of designing and working with high capacity bus systems. Since there is not a large number of such systems, the number of experts available with this experience was inevitably small. It was also felt to be important that the experts should be analytically inclined as it was thought that this would facilitate the acquisition process. In the event the main contributors of the expertise were indeed analytically disposed, although two comments should be made. First, the fact that the most helpful experts were analytical in their approach was in fact coincidental - these experts would have been chosen for their design experience with high capacity systems even if they were not particularly analytical. Second, a non-analytical expert would still be useful, although the knowledge acquirer would then have to provide the analysis of the expertise in order to encapsulate it within the knowledge base. It is therefore important to recognize that the main criterion for selecting any expert is that he or she has the appropriate experience and knowledge. Any interpretation problems are for the knowledge acquirer to solve. For this research, the experts were selected on a number of bases.

First, the experts were chosen on the basis of their proven experience in the design and implementation of high capacity bus systems which are currently operating (or which have been in operation in the past). Experts 10 and 11 were selected under this category.

The second criterion for the choice of experts is proven experience in the analysis of bus operations, particularly where these involve high capacity systems. Experts selected with this characteristic were experts 1, 3 and 11.

The third type selected consisted of experts with experience of operating high capacity bus systems, either as operators (expert 4), or as administrators of the municipality (experts 2, 8, 9 and 11).

Fourth, experts were selected with particular experience of a particular detail of the operation of high capacity bus systems (for example the estimation of the capacity of bus stops and the degree of saturation). Experts selected within this category were experts 1, 5, 6 and 11.

It is interesting to note that the contributions provided by each of these experts were not equivalent in terms of content or usefulness with respect to the objectives of the research. A recommendation for knowledge acquirers is therefore to have a clear view of the objectives of the acquisition process when selecting experts. Two experts in particular (experts 1 and 11) provided the majority of the analytical information used in this research. Expert 10 provided a large amount of information about the design process and the implementation of the high capacity bus system in the Santo Amaro/9 de julho corridor, and expert 7 provided similar (although less detailed) information about their corridor project (although this was not intended to be a high capacity system). Expert 3 provided a thorough objective analysis of the convoy operating system in Porto Alegre. The other experts provided varying amounts of information on details, for example the design of buses for high capacity operation (experts 5 and 6). Others provided more background information (for example experts 2, 4, 8 and 9).

It might therefore be concluded that the knowledge acquisition process could have been conducted with a smaller number of potential experts providing the core information (for example experts 1 and 11), and other experts selected for their expertise about specific issues (for example, experts 3, 5 and 6). Such an approach should be viewed with some caution, however, because while it is possible to make such an appraisal at the end of the acquisition process, it is extremely difficult to make such an objective assessment before it starts.

C4 KNOWLEDGE ACQUISITION METHODS USED

Where available, published papers manuals and articles were analyzed prior to the initial contact with each expert. This was possible only in the cases of experts 1, 3, 5, 6, 10 and 11, and reference is made to relevant documents throughout this thesis.

The basic decision structure was established during a series of initial interviews in which this was discussed with experts 7, 10 and 11 without reference to particular models or the details of the activities undertaken at each point in the decision process. In the cases of experts 7 and 10, the interview took place in the experts' offices in Sao Paulo, Brazil. Expert 11 was invited to London in order to conduct some of the knowledge acquisition process away from his home base. The benefits of this approach are discussed in Section 6.3.1.1.1.

A model of the decision structure, generated on the basis of these interviews and similar in form (but not content) to Figure 6.1, was then presented to each expert for comment to allow the opportunity for adjustment. This process was also undertaken with the experts individually because it was not possible to gather them together for this activity. Agreement on the basic decision structure was therefore achieved iteratively by taking a version to one expert, making the adjustments suggested and taking the revised version to the next expert. This was repeated until a version (shown in Figure 6.1) was agreed by all the experts. The necessity for repeated visits to the experts resulted in a time delay in the process since it was necessary to ensure that the basic model was agreed by all the experts. It was felt that if it had been possible to arrange a meeting with all the experts together, this would have been helpful in the sense that it could have reduced the amount of time taken to establish the basic model. In the event, the basic structure of the decision process was agreed between the experts without too much difficulty.

The use of particular models within the design process was discussed during interviews which were structured according to the basic decision structure. This was to ensure that all aspects of the design process were considered. This process was carried out with experts 1, 3, 5, 6, 10 and 11. In each case, models which had been derived by the experts were discussed in order to establish the feasibility of their use and their effectiveness for the relevant problem. In some cases, more than one model was considered for a particular aspect of the problem (for example, the calculation of bus stop capacity discussed in Section 3.3.3). These discussions enabled a rational choice to be made between the various alternatives knowing the strengths and weaknesses of each approach. These discussions took place in the experts' own offices, except in the cases of experts 3 and 11. Discussions concerning the models of these experts took place both in London and in Brazil.

In the case of the bus stop location, the model proposed by expert 11 was the analytical model discussed in Section 3.3.5.3 (Equation (3.23)). This model was included in the model because it represented the more recent thinking by expert 11 on the subject. As discussed in Chapter 8, a marked difference appeared between the results of this model and the implementation and this was discussed with the experts. The reasons for this difference appeared to be that the model used to estimate the distance between stops in the implementation was based on the heuristic "*bus stops should be located about 500 metres apart*". Partly because of this implementation, it was realized by expert 11 that the bus stops could be placed more closely together without losing too much of the benefits. Expert 10 was presented with this model and subsequently agreed that the newer approach was an improvement on the heuristic.

This problem with the knowledge acquisition process arises because expertise is not static, and is constantly being revised in the light of experience. The improved analytical model is clearly more satisfactory than the heuristic, although the validation process highlighted the difference between the expertise available before the implementation of the system and that available now, even when these are obtained from the same experts.

Examples of the type of questions asked during these interviews, the form of the answers received and the subsequent translation into a suitable form for use in the decision model are discussed in Sections 6.3.1.2, 6.3.2.3 and 6.4.1.1.

When there appeared to be a difference (for example with respect to marginal boarding times) between the expertise as presented at the interviews and the reality of the implemented solutions, the procedure was to take the expert to the appropriate location and to discuss the problem on site. This has two advantages. First, the expert is forced to discuss the theory in the presence of the reality and this reduces the opportunity of ignoring (or reducing in scale and importance) the differences between them. Second, it affords the opportunity for the expert to explain why the reality does not work or why it was not possible to implement the theoretical solution. The degree to which the implementation is suboptimal can also be observed on such occasions.

In addition, as mentioned in Section 6.3.1.1.1, expert 11 was brought to London. One objective of this visit was to observe how this expert set about designing a high capacity bus system in an unfamiliar environment: the characteristics that appeared important and those which did not appear to be influential. This was useful because it revealed the decision structure used by this expert and also confirmed how any suboptimalities might be approached. For example, the implementation of a scheme in a corridor should be considered even though there were parts of the corridor in which it would not be feasible. Although this did not alter the basic structure of the problem as represented in the model, it did alter the computer implementation: after this exercise the facility was introduced to attempt designs on specific parts of the corridor quickly and interactively before the design was attempted on the whole corridor. This allows particularly difficult parts of the corridor (for example where the available space is particularly constrained) to be examined with the possibility of ascertaining where the decision thresholds lie in terms of the various suboptimal solutions. For this reason, the final part of the initial assessment module described in Section 7.3.4 was added to the computer program.

C5 VARIABLES AND PARAMETERS DISCUSSED IN THE KNOWLEDGE ACQUISITION PROCESS

This is a list of the basic variables and parameters discussed with the experts and used in the design model. The generation of an opinion model required the generation of the set of opinion distributions as discussed in Section 6.4. The use of a model generated by an expert required the appraisal of its effectiveness, strengths and weaknesses with respect to the design model. The generation of a decision model involved the modelling of the decisions taken with respect to the variable or parameter concerned. Almost invariably, this also involved the generation of a set of opinion distributions. The qualitative assessment of data inputs and model outputs required the generation of a set of opinion distributions to describe the data concerned and a decision model to determine the subsequent action to be taken on the basis of this interpretation.

VARIABLE/PARAMETER	APPEARANCE IN DESIGN MODEL
Bus flow:	opinion model generated
Passenger flow:	opinion model generated

Bus types and designs:	parameter estimation techniques used in calculations
Bus stop capacity:	models incorporated, decision and opinion models generated
Saturation:	model incorporated, decision and opinions model generated
Traffic lane widths:	opinion model generated
Bus lane widths:	opinion model generated
Footway widths:	opinion model generated
Allocation of available space to element widths:	design and decision models generated
Position of staggered bus stop platforms:	decision and opinion models generated
Operating systems:	decision and opinion models generated
Bus stop design:	variables and parameters incorporated in design model, decision and opinion models generated
Number of berths:	models generated
Platform design:	model generated
Segregation infrastructure:	incorporated into design model
Commercial speeds:	decision and opinion models generated

Peak/off-peak speed ratio:	decision and opinion models generated
Bus stop position:	decision and opinion models generated
Stop spacing:	model incorporated
Data inputs:	decision and opinion models generated
Model outputs:	decision and opinion models generated