LABORATORY FOR
COMPUTER SCIENCE

MASSACHUSETTS
INSTITUTE OF
TECHNOLOGY

MIT/LCS/TR-388

AN EXPERT SYSTEM FOR DIAGNOSING GAIT FOR CEREBRAL PALSY PATIENTS

David Edward Hirsch

May 1987

545 TECHNOLOGY SQUARE, CAMBRIDGE, MASSACHUSETTS 02139

This blank page was inserted to preserve pagination.

An Expert System for Diagnosing Gait for Cerebral Palsy Patients

by

David Edward Hirsch

Copyright © 1987 Massachusetts Institute of Technology

Support for this research has been provided in part by the National Institute for Handicap Research of the Department of Education under contract G008300074 and also in part by the National Institutes of Health Grant No. R24 RR01320 from the Division of Research Resources.

Laboratory for Computer Science
Massachusetts Institute of Technology
Cambridge, MA 02139

This empty page was substituted for a blank page in the original document.

Table of Contents

1. Introduction 4

1.1 History 5

1.1.1 Artificial Intelligence in Medicine 5

1.1.2 Examples of Past Medical Data Interpretation Programs 6

1.1.2.1 PUFF 7

1.1.2.2 VM 7

1.1.3 Examples of Past Gait Analysis Programs 8

1.1.3.1 Stanford's Gait Diagnosis Program 8

1.1.3.2 GAITSPERT 10

1.2 Approach Taken in Designing Dr. Gait 11

1.3 A Roadmap through the Thesis 12

2. A Primer on Human Gait and Gait Analysis 14

2.1 Human Gait 14

2.1.1 Gait Cycle 14

2.1.2 Components of Gait 15

2.1.3 Gait Disorders 17

2.1.4 Treatments of Gait Disorders 20

2.2 Gait Analysis 20

2.2.1 History 20

2.2.2 Current Gait Analysis Techniques 21

3. Dr. Gait's Task 24

3.1 How the Dr. Gait System Will be Used 24

3.2 Inputs 25

3.3 Output 26

3.4 Goals and Success Criteria 28

3.5 Summary 28

4. Dr. Gait-1 29

4.1 Dr. Gait-1's Structure 29

4.1.1 GENIE 31

4.1.2 Medical Knowledge Base 31

4.1.2.1 Motion Class Object 32

4.1.2.2 Time-Measure Class Object 32

4.1.2.3 Anatomical-Entity Class Objects 34

4.1.3 Patient Frame 34

4.1.4 Rule Bases 35

4.2 Dr. Gait-1's Analysis Process 37

4.3 Example 38

4.3.1 The Case 38

4.3.2 The Analysis 38

4.3.3 Problems with Step 5: Incorrect Paragraphs 46

4.4 Conclusions About Dr. Gait-1 47

5. Structure of Dr. Gait-2 50

5.1 Motivation 50

5.2 Dr. Gait-2's Structure 51

5.2.1 GENIE 51

5.2.2 Medical Knowledge Base 51

5.2.3 Patient Frame 51

5.2.4 Preprocessor Module 53
5.2.5 Problem Identifier Module 53
5.2.6 Problem Enhancer Module 54
5.2.7 Causality Module 54
5.2.8 Therapy Module 59
5.3 Dr. Gait-2's Operation 61
5.4 Summary 61
6. Dr. Gait-2: Examples **62**
6.1 Example 1 62
6.1.1 Step 1: Gather all the Patient Data 62
6.1.2 Step 2: Identify the Problems 63
6.1.3 Step 3: Enhance the Problems 64
6.1.4 Step 4: Find Problem Causes 64
6.1.5 Step 5: Find Therapy 67
6.1.6 Step 6: Write out the Conclusions 69
6.2 Example 2 73
6.2.1 Finding the Correct Causes for the Knee Problems 73
6.3 Conclusions 80
7. Performance of Dr. Gait-2 **82**
7.1 Testing Process 82
7.2 Testing Results 82
7.2.1 Identification of Problems 82
7.2.2 Finding Causes of Problems 83
7.2.3 Therapy Selection 83
7.3 Conclusions About Performance 84
8. Conclusion **85**
8.1 Summary 85
8.2 Future Directions 85
8.3 Final Remarks 87
I. **93**

List of Figures

Figure 2-1: Gait Cycle (Reprinted from page 26 of [Inman 81] with permission of the publisher) 16
Figure 2-2: Three Orthogonal Planes for Gait Motion (Reprinted from page 34 of [Inman 81] with permission of the publisher) 18
Figure 2-3: Motions of Gait (Reprinted from page 35 of [Inman 81] with permission of the publisher) 19
Figure 2-4: Stick Figures and Force Vectors 22
Figure 2-5: Flexion/Extension Graph 22
Figure 2-6: EMG Data 23
Figure 3-1: Motion Data Transformation 27
Figure 3-2: EMG Data Transformation 27
Figure 4-1: Dr. Gait-1's Organization 30
Figure 4-2: Example of Motion Class Object 32
Figure 4-3: Time Hierarchy 33
Figure 4-4: Example of Event Object 33
Figure 4-5: Example of Phase Object 33
Figure 4-6: Example of Body Segment Class Object 34
Figure 4-7: Example of Joint Class Object 35
Figure 4-8: An Example of a Muscle Class Object 35
Figure 4-9: Rule which Selects Abnormal-Dorsiflexion Paragraph Template 42
Figure 4-10: Abnormal-Dorsiflexion Paragraph Template 42
Figure 4-11: A Rule which Fills in a Slot 44
Figure 4-12: The Filled in Paragraph Template 44
Figure 5-1: Dr. Gait-2 Organization 52
Figure 5-2: Tree of Torques for Knee in SLS 55
Figure 5-3: IQ Algebra 57
Figure 6-1: Torque Tree for Knee in Second-half-swing 65

List of Tables

Table 3-1:	Scale for Motion Data	28
Table 4-1:	Motion Data for John Patient	38
Table 4-2:	Diagnoses for Joint Motions of John Patient	41
Table 4-3:	Hip Motion Common Patterns	41
Table 6-1:	Motion Data for Example 1	62
Table 6-2:	Example 1: Knee Muscle Data	64
Table 8-1:	Muscle Data for Dr. Gait	93

Preface

This report is an updated version of my thesis, which was submitted to the Department of Electrical Engineering and Computer Science at MIT on May 16, 1980 for the degree of Master of Science. Since the thesis was written the therapy module of the program has been implemented. This report discusses the actual implementation. Thus sections 5.2.3 and 6.1.5 have been changed. In additon the system has been tested on a set of cases. The results are discussed in chapter 7.

Acknowledgements

No thesis is completed without the help and encouragement of all the people around you. I would like to thank my thesis supervisor, Peter Szolovits, for all his help during my research for this thesis. I am especially thankful for all those pep talks when everything seemed to be moving in the wrong direction. I would also like to thank all the present and past members of the Clinical Decision Making Group for giving me such a wonderful environment to work in. I appreciate all the help the workers, physical therapists and doctors at The Gait Analysis Laboratory provided in my quest to understand gait analysis. Jim Dzierzanowski was very helpful in getting me started and in setting up the GENIE system on the computer at the Gait Analysis Lab. A special thank you goes to Dr. Sheldon Simon of The Gait Analysis Lab. Our numerous meetings which often continued late into the night provided me with the necessary understanding of gait and gait analysis in order for this research to succeed. Everyone would be better off if we could all have Dr. Simon's enthusiasm, dedication and understanding. I would also like to thank Rosemary Hegg for editing the final version of this report. With the realization that life is not all work, I must thank Dennis Fogg who on numerous occasions forced me to take some time off from working to go out and enjoy what Boston has to offer.

This report describes research performed at the Laboratory for Computer Science of the Massachusetts Institute of Technology and at the Gait Analysis Laboratory at Children's Hospital in Boston. Support for this research has been provided in part by the National Institute for Handicap Research of the Department of Education under contract G008300074 and also in part by the National Institutes of Health Grant No. R24 RR01320 from the Division of Research Resources. A special word of thanks is also due to Professor Randy Davis for the use of the computer facilities supported by a grant from Digital Equipment Corporation.

An Expert System for Diagnosing Gait in Cerebral Palsy Patients

by

David Edward Hirsch

Submitted to the
Department of Electrical Engineering and Computer Science
on May 16, 1986 in partial fulfillment of the requirements
for the Degree of Master of Science

Abstract

Many first generation expert systems in medicine assumed that a single fault was the cause of the patient's problems. However, this is not always so and in the domain of gait analysis this is usually ***not*** the case. This work looks at an expert system for performing gait analysis on cerebral palsy patients. The system is able to handle cases where there are many interacting faults causing the patient's gait deviations. A qualitative-torque model of gait is used as a framework for analyzing the patient's disorders. This model allows the effects of individual faults to be examined and also explicitly captures the interaction between various faults. The incorporation of this model into an expert system leads to a more complete and robust expert system for gait analysis than any existing previously.

The qualitative-torque model of gait models gait by using torque equations represented as incremental qualitative (IQ) algebra equations. Because everything is not known about the torques, the system has to make assumptions to arrive at some values. These assumptions are kept consistent by using an assumption-based truth maintenance system. The only part of the model specific to cerebral palsy is the set of heuristics used to score competing assumptions. Thus, the gait model developed for this system could easily be adapted for use in an expert system diagnosing gait disorders of any given etiology.

Thesis Supervisor: Dr. Peter Szolovits
Title: Associate Professor of Electrical Engineering and
Computer Science

1. Introduction

The field of Artificial Intelligence (AI) has grown over the past decade from relative obscurity to prominence in the world of business. Although this growth in attention has been accompanied by the creation of new ideas and techniques, the art[1] of building AI programs is still a difficult task. There are two paths which lead to advances in this field. One path is that of the continuing invention of new techniques which can solve previously unsolvable problems. The other is the application of existing techniques which help the researchers better understand and further refine these techniques. The work presented in this thesis follows the second path.

This thesis examines the design and implementation of a knowledge based system, Dr. Gait, which interprets measurements of a cerebral palsy patient's gait pattern. The program, Dr. Gait, is given joint motion data, EMG data, force plate data and physical exam data which it interprets to find the underlying deviations in the patient's gait pattern, their causes and possible therapies.

First, a prototype system, Dr. Gait-1, was built. This system tries to emulate the reasoning of the expert physicians. It takes a very simple approach to gait analysis and was built in order to better understand what a gait analysis program has to do. The prototype uses empirical associations represented as production rules as well as domain specific knowledge stored in frames to arrive at its conclusions.

Based on what was learned from the prototype system, a final system, Dr. Gait-2, was built. This system is based on causal reasoning. The causal knowledge is stored in incremental qualitative algebra (IQ) equations. The values constrained by the IQ equations are kept consistent by means of an assumption based truth-maintenance system. In addition, Dr. Gait-2 uses frames, rule bases and discrimination trees to store other domain knowledge.

We will see that the naive approach taken by the prototype system, Dr. Gait-1, has several

[1] The field is still to young to be accurately called a science.

problems. The system does not give enough attention to the muscles, which have an important effect on a person's gait. It has only limited abstraction abilities and in addition, it makes many implicit assumptions which may lead it to incorrect conclusions in abnormal cases. The final system, Dr. Gait-2, corrects many of these problems by having a qualitative model of gait. The model gives the system a framework in which to organize the problems and a means to study the

2. A Primer on Human Gait and Gait Analysis

In this chapter we will first examine human gait. In so doing, we will be able to distinguish and better understand the important characteristics of a person's walking motion. Interested readers might want to look at Inman's book [Inman 81] for a more in depth discussion of human walking. After examining gait we will then go over some of the details of gait analysis. We will discover how the different forms of data are gathered and how they are then presented to the attending physician. All this will give the reader sufficient background to understand the matters with which the program, Dr. Gait., is concerned.

2.1 Human Gait

Everyone faces the problem of how to get from one place to another. Furthermore, each of us would like to do this with minimum effort, adequate stability and acceptable appearance. Each person's unique solution to this problem represents his gait (walking) pattern. Remarkably enough, most of us learn to walk with reasonable facility and surprising efficiency.

2.1.1 Gait Cycle

Usually, when talking of gait, one thinks of a cyclic movement of body parts which is repeated over and over again, step after step. A basic assumption is that this cyclic pattern is the same from cycle to cycle. While this is not always true, it is a good approximation. Thus, most analyses and descriptions of gait deal with what happens over one gait cycle.

A *gait cycle* consists of the time between two consecutive heel strikes of the same foot.[6] The distance between these heel strikes constitutes the person's stride length. The gait cycle can also be divided into phases which are determined by significant events. We will use the following significant events to determine our phases:

1. *Right Heel Strike* (RHS) - When the heel of the right foot first touches the ground.

[6] In this thesis we will always consider a gait cycle as going from *right* heel strike to *right* heel strike.

2. *Left Toe Off* (LTO) - When the toes on the left foot have their last contact with the ground.

3. *Left Heel Strike* (LHS) - When the heel of the left foot first touches the ground.

4. *Right Toe Off* (RTO) - When the toes on the right foot have their last contact with the ground.

Using these significant events to divide the gait cycle we get the following phases:[7]

1. ***Weight Acceptance*** (WA) From RHS to LTO, 0 to 16% of gait cycle. During this phase the body weight is being transferred from the left leg to the right leg. Since the body is being supported by both legs during this phase it is also called double limb support.

2. ***Single Limb Stance*** (SLS) From LTO to LHS, 16-50% of gait cycle. During this phase the body is only being supported by the right leg. As the body raises itself up and over the right leg, the left leg is being swung forward.

3. ***Weight Release*** (WR) From LHS to RTO, 50-66% of the gait cycle. During this phase the body weight is being transferred back to the left leg. Again this phase is also called double limb support since the body is being supported by both legs.

4. ***Swing*** (swing) From RTO-RHS, 66-100% of the gait cycle. During this phase the right leg is free to swing forward. While the right leg is swinging forward the body is being supported by the left leg.

Figure 2-1 shows these different phases of the gait cycle.

2.1.2 Components of Gait

Now that we have definitions for a gait cycle and its different phases we have some idea of how to break up a person's walking motion in order to examine it. However, we must still ask what factors determine the form of a person's gait. A person's gait pattern is the result of several torques acting upon the body. There are internal torques which are produced by the muscles pulling on the limb segments. Then there are external torques caused by gravity acting upon the body mass, momentum and counteraction of the floor's reaction force. The result of all these torques is the particular motions of the different joints.

Even though trunk movement and arm swing play an important role in walking, we can

[7] These phases are from the point of view of the *right* leg.

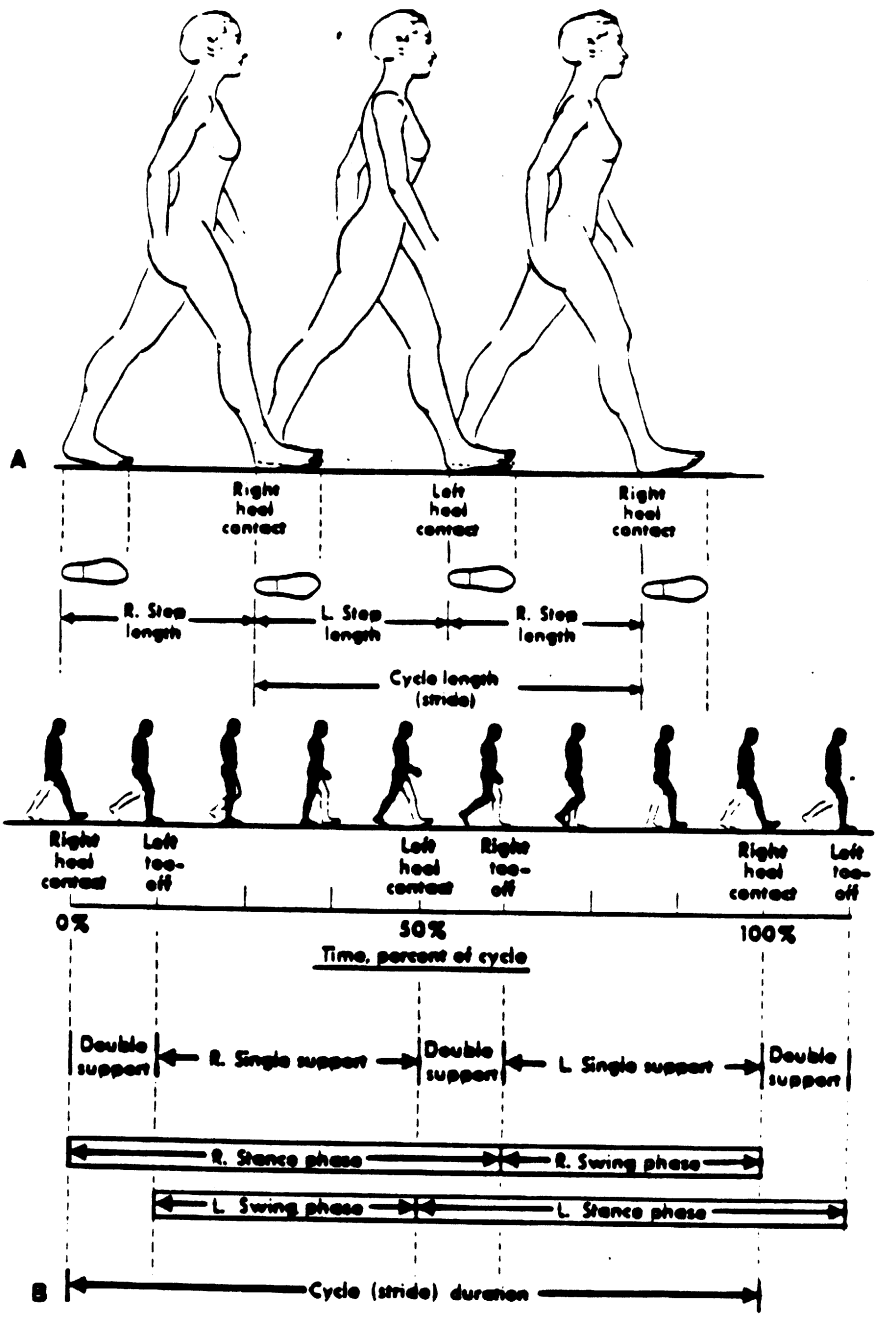


Figure 2-1: Gait Cycle
(Reprinted from page 26 of [Inman 81] with permission of the publisher)

understand much about human gait by looking only at the lower extremities. Thus, to examine gait, we look at the joints and muscles in the legs. The three joints of concern in each leg are the hip, knee and ankle. The three dimensional motion of each of these joints can be broken into projections onto three orthogonal planes. These three planes are named coronal, transverse and sagittal. See figure 2-2 for an illustration of these planes. The motions of the joints in these planes have names such as flexion, extension, adduction, internal rotation, etc. See figure 2-3 for illustrations which define each of the motion terms.

The muscles with which will be concerned all act upon the legs. We will take the simplistic view that when a muscle is on, it shortens and thus pulls on the joint in the corresponding direction (or else is opposing a torque in the opposite direction). When a muscle is off, it is at rest. Thus, an inactive muscle does not pull on the joint or resist pulls that oppose it. Appendix A gives a list of all the muscles we are concerned with and what actions they each perform.

2.1.3 Gait Disorders

A gait disorder is any abnormality in a person's gait pattern. An abnormality could be the improper positioning of a limb or improper activity of a muscle. These abnormalities hinder a person's gait and thus make it more difficult for him to walk.

A gait disorder can either be temporary or chronic. A temporary gait disorder usually results from an accident or an acute illness. These disorders are usually alleviated through the passage of time as the body heals itself. Sometimes physical therapy is needed to help in rehabilitation. Chronic gait disorders usually occur because of a chronic condition such as cerebral palsy, or arthritis, or a stroke. Chronic gait disorders are usually more serious and are more difficult to treat.

There are three types of gait disorders: muscular problems such as muscle weakness or tightness, joint problems such as tight or inflamed joint capsules, and neurological problems (which are the most difficult). A patient with a neurological disorder is not able to activate his muscles or therefore control movements of his limbs. This leads to poor gait which is difficult to

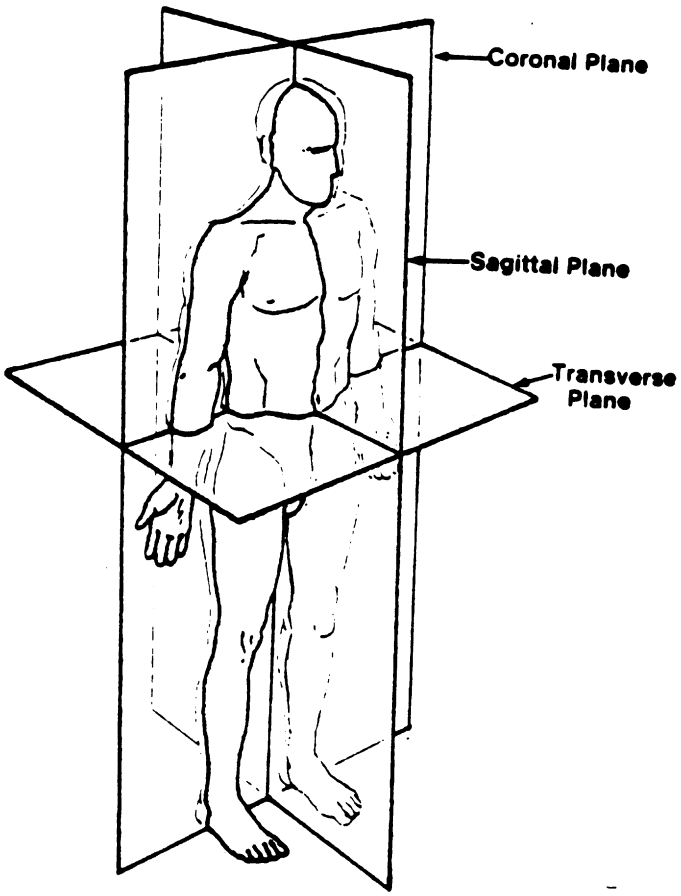


Figure 2-2: Three Orthogonal Planes for Gait Motion
(Reprinted from page 34 of [Inman 81] with permission of the publisher)

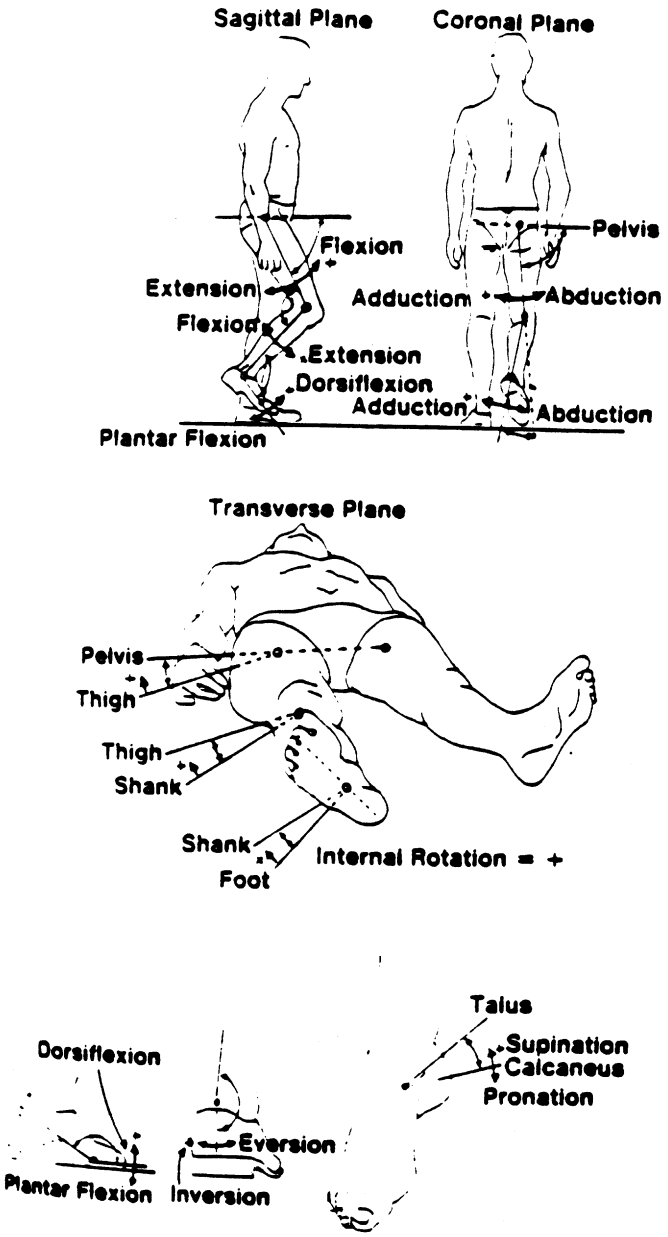


Figure 2-3: Motions of Gait
(Reprinted from page 35 of [Inman 81] with permission of the publisher)

treat beacause it cannot be treated directly. One must treat the manifestations of the control disorder instead of treating the control problem itself.

2.1.4 Treatments of Gait Disorders

Physical therapy, bracing and surgery are the three ways to treat gait disorders. It is used for minor problems. Physical therapy could include stretching to loosen up tight muscles or exercises to strengthen weak muscles. For more severe problems there is a choice between bracing and surgery. Bracing is used either to restrict the range of motion of a joint or else to keep the joint in a fixed position. For example, if the patient has a dropped foot in the swing phase, an ankle brace could be used to keep the ankle in the neutral position (90 degrees) and thus prevent the dropped foot in swing.

The most severe treatment is surgery. The types of surgery available are numerous. Most of them involve changing the muscles. There is muscle lengthening used to offset the affect of overactive muscles, muscle shortening used to counteract weak muscles, and muscle transfer used to change how the muscle effects the joint.

2.2 Gait Analysis

2.2.1 History

For many years there has been a need to help people with gait disorders and thus a need for gait analysis in order to know how to better correct their disorders. However, until recently the analysis of gait was done mainly by casual observation. The doctor would watch the patient walk by several times and then make an educated guess as to how to correct any observed problems. This led to problems, in that doctors were unable to make accurate and complete diagnoses of the gait disorders. Furthermore, doctors could not be sure if the treatments they prescribed would improve the patient's gait. Many times a treatment would help one deviation, which the doctor had observed, but in doing so would aggravate another problem that the doctor had missed. The outcome of all of this was that some patient's were no better off than before, even after six or seven surgeries.

2.2.2 Current Gait Analysis Techniques

Over the last several years half of the problem has been solved. Now there are monitoring systems which can gather data about a patient's gait. For example at the Gait Analysis Laboratory at Children's Hospital in Boston,

> ...a patient's gait pattern is monitored with the simultaneous measurements obtained of the motions of all four limb segments and trunk with three 16mm high speed cameras, the motor activity of the major muscle groups in the lower extremities with EMG electrodes and foot floor reaction forces generated by means of a force plate.[8]

All this data is sent to a computer where it is stored so that it can later be processed to yield useful information about a patient's gait.

By placing special anatomical markers at key points on the body, these points can then be traced on the film. After the film is digitized, frame by frame, various calculations can be made regarding a patient's gait. Such parameters as limb segment positions, shown as stick figures, and joint angles for each joint in each of the three planes, shown as flexion/extension graphs, are calculated. See figures 2-4 and 2-5 for examples of stick figures and flexion/extension graphs. In addition, from the film various parameters such as velocity, stride length, step width and the time spent in each phase can be calculated.

The electromyogram (EMG) electrodes gather a record of the muscle motor activity. After performing much complex filtering of this data to eradicate the noise, a graph of the motor activity for each monitored muscle is obtained. See figure 2-6 for an example of EMG data.

The force plates provide data on the foot-floor reaction forces, whose magnitude and direction as a function of time can be calculated. The calculated forces are shown as vectors superimposed on the stick figures. See figure 2-4.

Currently the attending physician looks over all these graphs and parameters, in addition to taking into account the patient's past medical history to document the patient's gait deviations, determine the cause of the deviations and recommend treatment. Doctors now have more

[8]From a Gait Analysis Report, The Gait Analysis Lab, Children's Hospital, Boston, MA

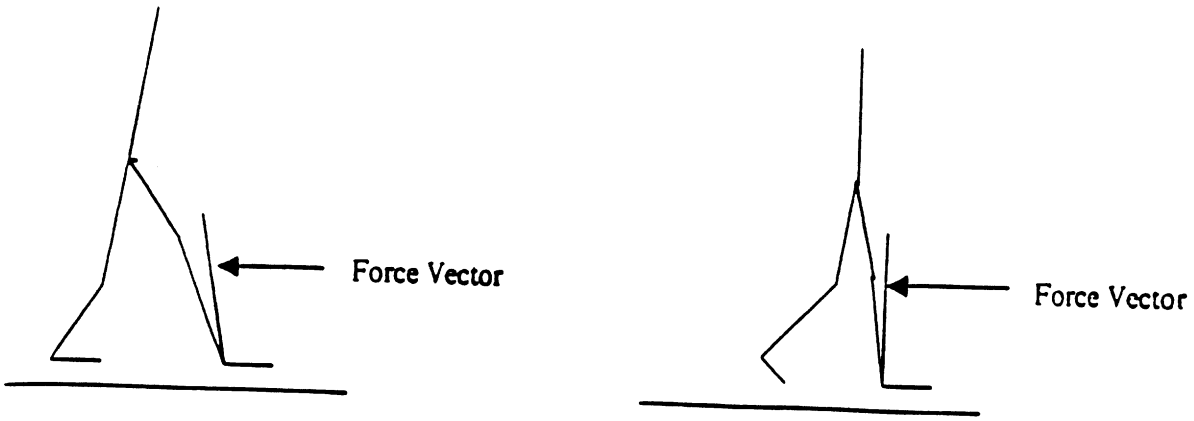


Figure 2-4: Stick Figures and Force Vectors

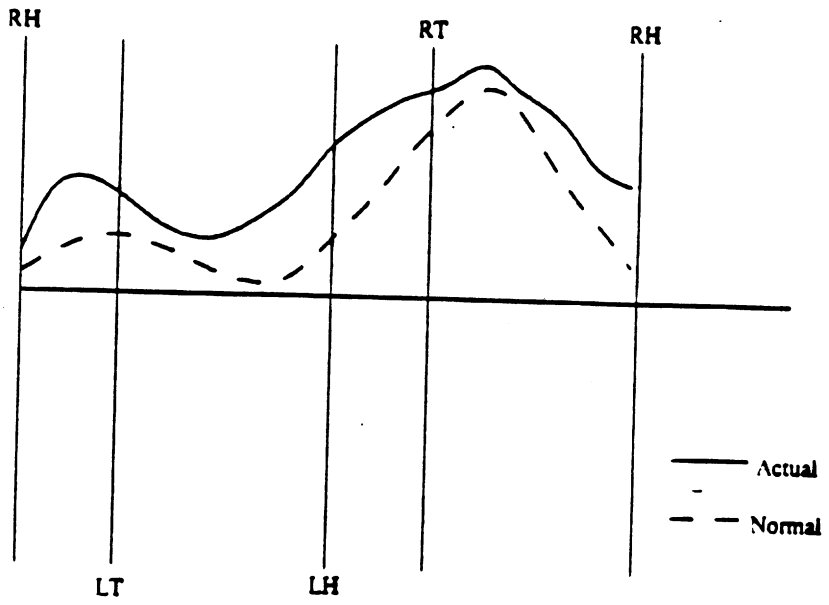


Figure 2-5: Flexion/Extension Graph

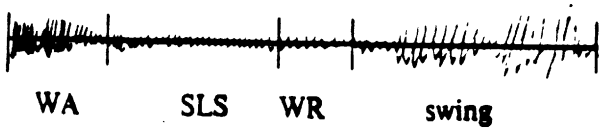


Figure 2-6: EMG Data

complete data to use in their decisions than in the past when only subjective observations were available. However, they are faced with the task of sifting through a large amount of interrelated data in order to diagnose and treat patients. This implies that if a computer system could be written to screen this data and start fitting the pieces together, it would make life much easier for the attending doctor. Dr. Gait is such a program.

3. Dr. Gait's Task

This chapter explains the specifications of the Dr. Gait system. First, we will see where Dr. Gait fits into the gait analysis process. This will give us a good idea of what Dr. Gait has to do. Next we will look at the data given to Dr. Gait. We will then explain the three tasks that Dr. Gait must accomplish. Finally, the goal we are trying to achieve by building this system will be stated and criteria listed to use in checking to see if the final system reaches the desired goal.

3.1 How the Dr. Gait System Will be Used

To understand how and when Dr. Gait will be used we need to look at the steps currently taken in a gait analysis of a patient. Current analysis takes the following steps:

1. *Data is gathered* using a gait recording system and is stored in the computer.
2. The film of a patient walking is *digitized* frame by frame to give the computer more information about the patient's gait.
3. All the data in the computer is *processed* by various programs to produce the graphs and figures discussed in section 2.2.2.
4. The graphs and figures are *examined and analyzed* by the doctors and physical therapists who then write up a gait analysis report.

The most difficult part of this process is the last step where the doctors and therapists have to look at vast amounts of data and then draw some conclusions. The intention is to add an additional step between steps three and four in which the data is also passed to the Dr. Gait program. The Dr. Gait program would produce a preliminary gait analysis report. The doctors and therapists in step four would then not only have the graphs and figures but also a preliminary report to work with. This should make it much easier to produce the final gait report.

One thing that is important to note is that there should be no need for the doctors to interact with Dr. Gait. Dr. Gait gathers its data directly from the computer. This eliminates the need for complex human interfaces and worries about the speed of operation. The goal is to have the program run automatically when each gait study is processed without any need for human interaction, but this has not been fully achieved. Currently, although much of the data is

gathered directly from the computer, some of the data still must be input by the user. The input specifications in the next section tell which data currently comes from the computer and which from the user. After more patient data is put on line it should be possible to write additional interfaces to gather the rest of the data directly.

Taking all of this into account the new way to perform a gait analysis is:

1. Data is gathered using a gait recording system and is stored in the computer.

2. The film of a patient walking is digitized frame by frame to give the computer more information about the patient's gait.

3. All the data in the computer is processed by various programs to produce the graphs and figures discussed in section 2.2.2.

4. *The data in the computer in addition to the data produced by step three are fed into Dr. Gait. Dr. Gait analyzes all of this data and then produces a preliminary report.*

5. The *preliminary report along with the graphs and figures* are examined and analyzed by the doctors and physical therapists who then write up a *final analysis report.*

3.2 Inputs

The inputs to Dr. Gait consist of various kinds of measurements. Most of this data is an encoding of the data presented in section 2.2.2. This input data includes:

1. ***Motion Data*** - The motion data presented in the flexion/extension graphs for the different joints is scaled by degree of deviation from normal. The scale ranges from markedly decreased (-5) to markedly increased (+5). See table 3-1 and figure 3-1 for the scaling definitions and an example. Thus, the absolute joint position is *not* used by the program, only its *deviation* from normal is used.

   Our representation of the motion data scales a continuous quantity to a set of discrete qualitative values immediately. Others argue that the data should be left in its continuous form and only transformed to discrete values when needed. They claim that information is lost in this transformation. This is indeed the case. There is the problem that several different joint motions can be transformed to the same set of discrete qualitative values. Currently, if any of this lost information is later needed, the user is asked for this information. In the future if routines could be written to determine these additional characteristics directly from the data there would be no reason to ask the user for the information. The scaled qualitative information is used to help identify motion problems and to check the value of torques in the causality module. For both of these tasks the qualitative information is well suited. The causality module is using qualitative reasoning so that the qualitative values for the motions fit in perfectly. Furthermore, the experts naturally classify the motion data as some discrete descriptive value and then use this in their reasoning. If this value does not capture some information they later need they will refer back to the graph. Thus, the way we treat the motion data is the same as the expert appears to use it.

2. ***Interpretation of Motion Data*** - Also from the flexion/extension graphs other information is gathered. This information is currently gathered by asking the user to answer questions about the graphs. For example, "Is the hip's rate of increase of flexion in the beginning of swing increased, normal or decreased?" or "Is there a flattening of knee flexion in single limb stance?" Each question is only asked if it is relevant to the program's current investigation. It is hoped that in the future much of this information could be calculated directly from the motion data in the computer.

3. ***EMG Data*** - The EMG signal from each muscle is used to determine if the given muscle was off or on in a particular phase. See figure 3-2 for an example. Presumably more sophisticated interpretations could be made, but, first of all, more research into EMG interpretations is needed.

4. ***Information About the Patient's Medical History*** - The user is asked about particular past surgeries and use of walking-assistive devices. This includes such inquiries as, "Has there been previous use of a below the knee orthosis?" or "Has there been a previous hamstring lengthening?". Again a particular question is asked only if it is relevant to the current case. As more and more patient data gets stored in an online data base, it should be possible to get answers to there questions directly from the data base instead of asking the user.

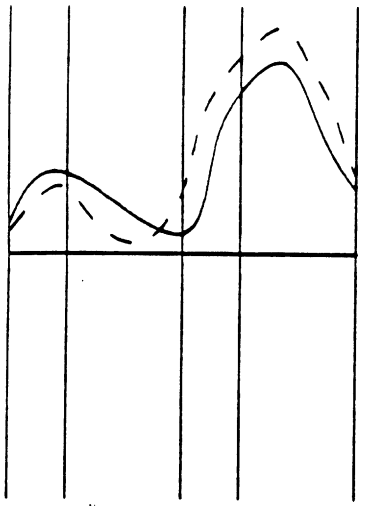
3.3 Output

A gait analysis report has the following sections: an introductory section discussing the patient's medical history and reasons for referral to the gait lab, another documenting a patient's motion and muscle activity deviations, a further one discussing the possible explanations for the patient's gait problems, and then possibly a final section recommending therapies to help alleviate the noted problems. Because we want the program to be a first step in producing such a report, the program should output similar information. Thus, the program needs to accomplish three tasks:

1. ***Identify*** all the deviations present and their consequences (e.g. decreased hip flexion is present in swing and this causes scraping of the toe).

2. For each identified deviation determine its ***cause*** (e.g. the increased knee flexion in single limb stance is caused by excessive hamstring activity).

3. ***Recommend treatments*** which will help alleviate the identified deviations (e.g. suggest hamstring lengthening to reduce excessive knee flexion).

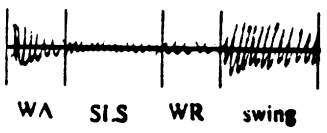


Right Knee Flexion

Phase	Motion
WA	mildly increased
first half SI.S	increased
second half SI.S	mildly increased
WR	decreased
first half swing	decreased
second half swing	mildly decreased

Scaled Motion Data

Figure 3-1: Motion Data Transformation



Right Anterior Tibialis EMG Signal

Phase	EMG
WA	on
first half SI.S	off
second half SI.S	off
WR	off
first half swing	on
second half swing	on

EMG Data

Figure 3-2: EMG Data Transformation

Motion Description	*Scaled Magnitude*	*Measured Joint*
severely increased	+5	>25 degrees increased from normal
markedly increased	+4	20-25 degrees increased from normal
very increased	+3	15-20 degrees increased from normal
increased	+2	10-15 degrees increased from normal
mildly increased	+1	1-10 degrees increased from normal
WNL	0	normal - no deviation
mildly decreased	-1	1-10 degrees decreased from normal
decreased	-2	10-15 degrees decreased from normal
very decreased	-3	15-20 degrees decreased from normal
markedly decreased	-4	20-25 degrees decreased from normal
severely decreased	-5	>25 degrees decreased from normal

Table 3-1: Scale for Motion Data

3.4 Goals and Success Criteria

Our goal is to have Dr. Gait perform its three tasks successfully. This means that Dr. Gait has to identify the motion deviations, determine their causes and recommend treatments. The gold standard that we will be using is the set of results obtained from the expert. Thus, we need to see if Dr. Gait identifies the same gait deviations, discovers the same underlying causes and recommends the same treatments as does the expert on the same cases.

3.5 Summary

Dr. Gait's task is to produce a preliminary gait analysis report which the doctors can use in writing up a final report. To produce this report, Dr. Gait must identify the deviations, determine their causes and recommend therapies. Dr. Gait is given motion data represented as scaled deviations from normal, EMG data represented as on/off values and various other data obtained by asking the user specific questions. Thus, all we need now explain is how Dr. Gait takes this data and accomplishes its three tasks. In Chapter 4 we will see a naive approach taken by the prototype system, Dr. Gait-1. Later, in chapter 5, we will see a more sophisticated approach taken by Dr. Gait-2.

4. Dr. Gait-1

This chapter examines the design and implementation of the prototype system, Dr. Gait-1. A brief overview of the structure of the program is presented, after which there will be an explanation of Dr. Gait-1's analysis process. An example will be presented to illustrate the operation of the program and also to indicate some of the weaknesses of the naive approach. Finally, corrections to these problems are discussed.

The prototype system was built to mimic the expert's reasoning as closely as possible, in part to identify his knowledge. This led to a system based on empirical associations, with no deep understanding of how gait works but only associations between patterns of symptoms and causes.

A CP patient usually has multiple deviations with different causes. Thus, the single fault assumptions that many expert systems make is not valid for this domain. The prototype system deals with these multiple deviations and causes by finding patterns of symptoms which represent a set of deviations and then attributing some causes to this pattern of symptoms. Therefore, each deviation and cause is not dealt with in isolation. The resulting system is then limited in the cases it can handle. Only patterns of symptoms that are encoded into the system are found. If the patient's symptoms do not quite match these patterns, nothing is concluded.

4.1 Dr. Gait-1's Structure

The Dr. Gait-1 system is made up of five parts: some top level code which controls the system, the GENIE inference engine which is used to operate on frames and rule bases, a patient frame which stores the current information about the patient, a medical knowledge base which captures the essential medical knowledge necessary to perform gait analysis and several rule bases which perform the analysis. The system organization is shown in figure 4-1.

Below the important parts of the Dr. Gait-1 System are discussed:

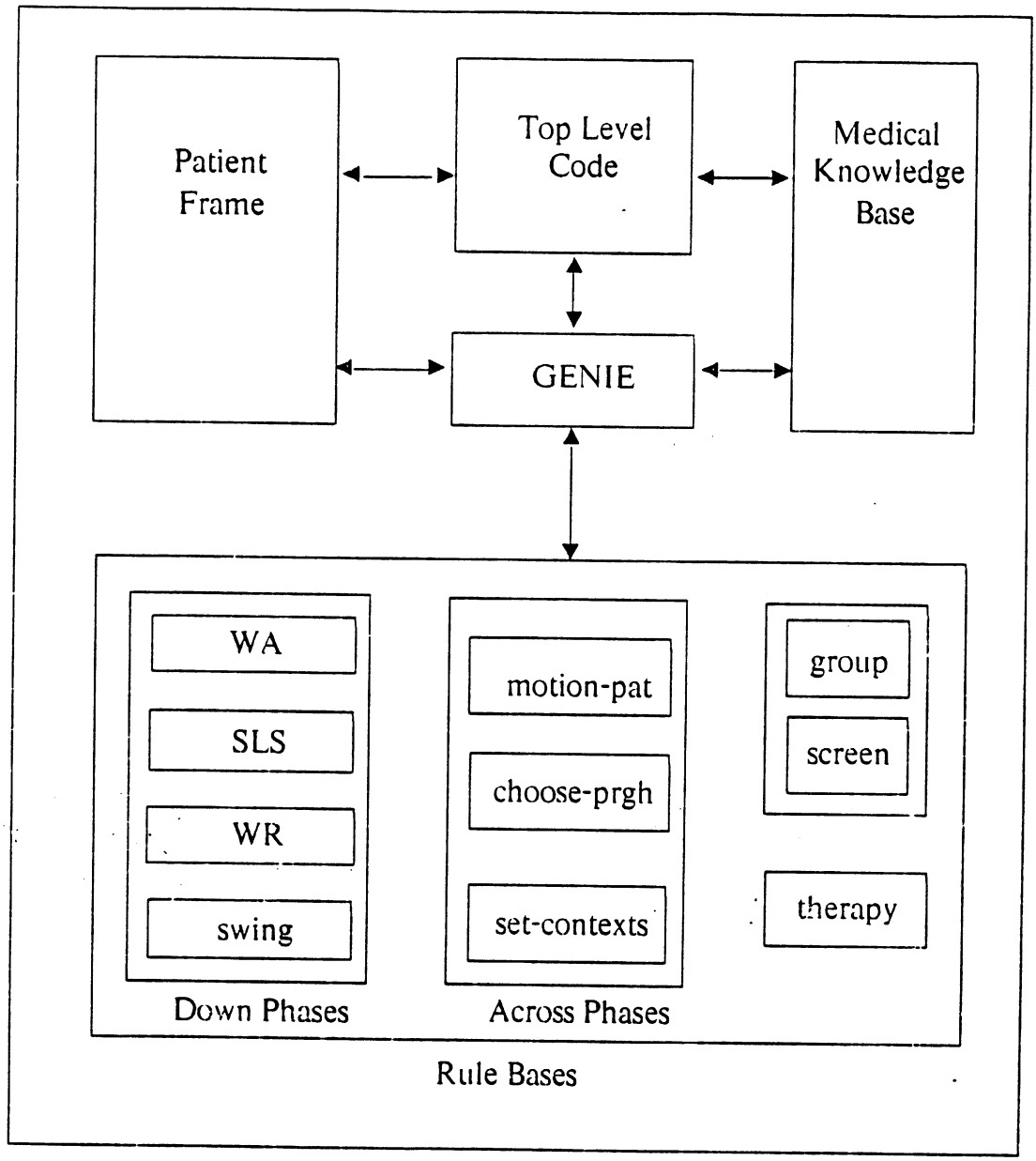


Figure 4-1: Dr. Gait-1's Organization

4.1.1 GENIE

GENIE [Sandell 84] is a general purpose knowledge engineering tool, developed at Vanderbilt University, for creating expert systems based both on rule bases and frames. Systems built with GENIE can use frames throughout the system as a uniform data structure in which to store static knowledge. Production rules can be partitioned into several rule bases which can then be applied individually. Context-dependent variables[9] can be included in the rule clauses, thus allowing rules with general knowledge to be applied to specific situations. GENIE uses the MYCIN [Shortliffe 76] scheme of certainty factors to deal with uncertainty. The rule interpreter of GENIE allows rules to be applied using backward chaining, forward chaining or data-directed rule application. We have modified the system so that whenever a rule causes a value to be placed in a frame slot, a record of dependency information is stored with it. Efficient control of the expert system can be obtained by combining the different rule-base application strategies to fit the problem domain. The GENIE system also includes facilities for explaining its reasonings through the display of rules, rule application order and frame contents.

Dr. Gait-1 uses the GENIE system to run rule bases and to access and create frames. GENIE provides a convenient tool to help build many of the components of Dr. Gait-1.

4.1.2 Medical Knowledge Base

The medical knowledge base captures the essential anatomical and physiological knowledge necessary so that the accompanying expert system, in this case Dr. Gait-1, can perform gait analysis. The knowledge base is built up of objects (usually nodes in a hierarchy), each of which is represented as a frame. The objects in the data base are divided into three classes: motion, time-measure and anatomical-entity classes.

[9] We modified the system so that specially designated global variables could be used in rule clauses.

4.1.2.1 Motion Class Object

The motion class objects give the system information about the different types of motion that are used in describing the motion of a joint. These motions include abduction, adduction, flexion, extension, internal rotation, external rotation, dorsiflexion and plantarflexion. Each motion tells which plane the motion occurs in, whether the motion is defined as the positive or negative direction and the name of its opposing motion. Figure 4-2 gives an example of a motion object.

```
dorsiflexion
    is-a
        motion
    in-plane
        sagittal
    direction
        -
    opposing-motion
        plantarflexion
```

Figure 4-2: Example of Motion Class Object

4.1.2.2 Time-Measure Class Object

There are two different entities in this class, *event objects* and *phase objects*. *Event objects* represent the significant events defined in section 2.1. These events serve as landmarks in the gait cycle and are ordered so that the system will know which event proceeds or follows another one. They include information on which phases or subphases they initiate and which ones they terminate. Figure 4-4 shows a typical event object.

Phase objects define names for the intervals between event objects in the gait cycle. These are exactly the phases defined in section 2.2. Furthermore, these phase objects are hierarchically organized so that some phases also have subphases. Figure 4-3 shows the time hierarchy. The phase objects keep track of the events that initiate and terminate them. Thus, the phases can be ordered using the knowledge of the ordering of the significant events. A phase object is shown in figure 4-5.

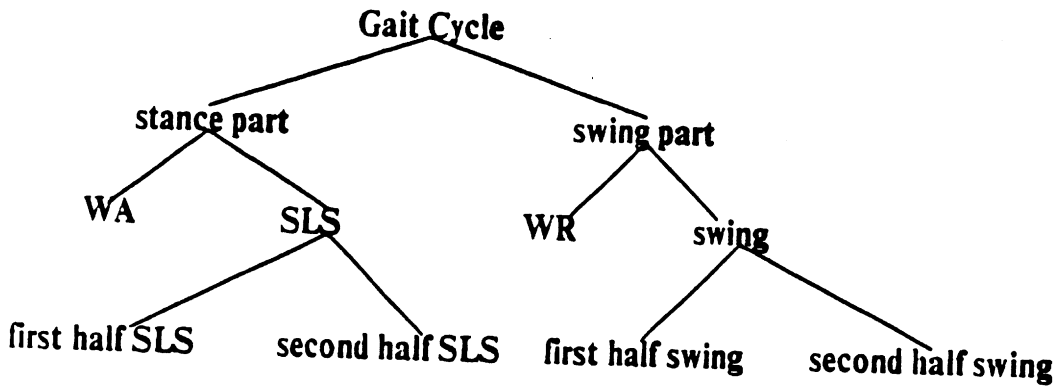


Figure 4-3: Time Hierarchy

```
ipsilateral foot strike
    is-a
       event
    initiates
        gait-cycle
        WA
    ends
        gait-cycle
        swing
        second-half-swing
```

Figure 4-4: Example of Event Object

```
SLS
    is-a
        stance-part
    subphases-of
        first-half-SLS
        second-half-SLS
    from
        opposite toe off
    to
        opposite heel strike
    motion-rule-base
         SLS-rb
```

Figure 4-5: Example of Phase Object

4.1.2.3 Anatomical-Entity Class Objects

There are three types of anatomical-entity class objects: *body segment objects, joint objects and muscle objects. Body-segment objects* define connectivity information about the various segments of the body which include the thigh, shank, foot and HAT (head, arms and trunk). Figure 4-6 shows a typical body segment object.

Joint objects store pertinent information about each of the joints in the leg. We can learn about the possible motions for each joint, which muscles act upon the joint in question, which joints are above or below, and which body segments are directly above or below. An example is in figure 4-7.

Muscle objects have information about a muscle's physiology which acquaints us with which joints the muscle acts upon and how the muscle affects each such joint when it is active. In addition, each muscle object has information about when the muscle is normally on. See figure 4-8 for an example.

```
thigh
    is-a
        body-segment
    proximal-joint
        hip
    distal-joint
        knee
```

Figure 4-6: Example of Body Segment Class Object

4.1.3 Patient Frame

This part of the system acts like the blackboard in the HEARSAY [Lesser 77] system to record information about the current patient and any results obtained or intermediate conclusions reached by any part of the system. The various rule bases can pass information to each other via the patient frame.

```
knee
    is-a
        joint
    motions
        flexion
        extension
    muscles
        quadricep
        hamstring
        gastroc/soleus
    proximal-body-segment
         thigh
    distal-body-segment
         shank
    joints above
        hip
    joints below
        ankle
```

Figure 4-7: Example of Joint Class Object

```
knee-flexor
    is-a
        muscle
    acts-on
        knee
            causing
               flexion
    stronger-than
        knee-extensor
    instances-of
        hamstring
        gastroc/soleus
```

Figure 4-8: An Example of a Muscle Class Object

4.1.4 Rule Bases

One advantage of the GENIE system over a system like MYCIN [Shortliffe 76] is that one can build an expert system with several rule bases and thus partition the rules into smaller sets. Dr. Gait-1 has nine rule bases which perform various functions. These rule bases are:

1. ***Group*** - This rule base uses backward chaining to classify the patient into one of three velocity groups (high velocity group, middle velocity group or low velocity group). This is a very small rule base. Rules are of the form:

```
group-rule2
  ifall
    (patient velocity) > 0
    (patient velocity) <= 35
  then
    conclude: (patient group)
       with value: low-velocity
```

2. ***WA, WR, SLS, Swing*** - Each of these rule bases is applied to a particular phase and tries to classify the motion at each joint as either a *primary cause*, a *compensation* for other deviations, a demonstration of *lack of compensation* for the deviations, a *passive problem* or a *passive compensation*. This classification is carried out by pattern matching combinations of observed motions with those in the predicates in rules. The rules in each of these rule bases is searched by backward chaining through the rules. A typical rule looks like this:

```
WR-rule7
 ifall
  (patient psm hip (context current-phase) motion-data)
     at-least mildly-increased
  not
    (patient psm ankle (context current-phase) motion-data)
       at-most mildly-decreased
  not
    (patient psm knee (context current-phase) motion-data)
       at-most mildly-decreased
 then
   conclude: (patient hip (context current-phase) Dx)
      with value: primary-cause
```

The (context current-phase)'s in the rule clauses allow the current value of the named frame slot, in this case current-phase, to be spliced into the name of the path to the parameter. This allows one to write a rule base, say for SLS, that can apply to both the first-half of SLS and the second-half of SLS subphases. To apply it to one or the other one need only set the frame slot appropriately. The above rule states that if the hip shows increased flexion, while the knee does not show decreased flexion and the ankle does not show decreased plantarflexion, then the hip's motion is diagnosed to be a primary cause.

3. ***Motion Pattern*** - This rule base tries to match the patterns of motions across the phases for a particular joint. These patterns were arrived at by having the expert examine many past case studies and abstracting the most common patterns. Currently, there are about ten patterns for each joint. To match the patterns the rule base uses backward chaining.

4. ***Choose Paragraphs*** - Using the motion patterns determined by the motion pattern rule base along with other information about the patient's gait, this rule base uses forward chaining to identify global problems that are affecting the patient's gait. In effect, what this rule base does is select applicable paragraphs which describe particular deviations. These paragraphs are selected by matching predetermined situations to the one at hand. Once again we are using superficial knowledge to arrive at conclusions. This moderately sized rule base has forty rules.

5. ***Set Contexts*** - The paragraphs selected by the choose paragraphs rule base are more like templates than full paragraphs. This rule base is used to fill in the blanks in the chosen paragraph templates. By doing so one specializes the selected paragraph to fit the case at hand. Since there are over one hundred slot, each of which can be filled in in several ways, this leads to a large rule base of over one hundred and fifty rules. The rule base is forward chained. GENIE's simple forward-chaining mechanism causes every rule to be tried once. This means that the rule base is very slow and inefficient. Furthermore, the slots can only be filled in in a few predetermined ways. This sometimes leads to cases where a paragraph gets chosen but not all the slots can be filled in with the appropriate information.

6. ***Therapy*** - The therapy rule base is much like the choose paragraphs rule base. Using information about matched motion patterns, chosen paragraphs emg data and motion data, this rule base selects paragraphs describing therapy recommendations. It also uses the empirical pattern matching approach. This rule base is only partially complete. Most of its information deals only with the ankle. Thus, there are very few therapy recommendations for problems at the hip or knee. Even those rules for the ankle do not work in many situations. The main problem with this rule base is that there are no intermediate concepts to guide the search. It tries to match the situation *exactly* to the conclusions it knows how to reach.

4.2 Dr. Gait-1's Analysis Process

The top level control gathers some background information about the user and then proceeds to run the rule bases in the order presented above. Thus the process has the following eight steps:

1. Gather initial patient information.

2. Apply the group rule base to classify the patient into a velocity grouping. This grouping gives one a good idea of the overall ability of the patient.

3. Apply the WA, SLS, WR and Swing rule bases to classify the motion of each of the three joints in each phase as a primary cause, compensation, etc.

4. Apply the motion pattern rule base to see if any common motion patterns are present across the phases for any of the joints.

5. Apply the choose paragraphs rule base to find local and global problems that are demonstrated by the patient.

6. Apply the set contexts rule base to fill in the slots in the paragraph templates just selected.

7. Apply the therapy rule base to find possible therapies to help the patient.

8. Write out all conclusions.

4.3 Example

Below we will present Dr. Gait-1 analyzing a patient.[10] This example will demonstrate the reasoning process of Dr. Gait-1. It will also illustrate some of the problems with the Dr. Gait-1 system. These problems will be indicated and ultimately some suggestions for alleviating these problems will be discussed.

4.3.1 The Case

Dr. Gait-1 is going to analyze John Patient's gait. John Patient is a CP patient who has difficulty walking. His velocity is only 55% of normal but his coordination is fairly good. His problem is that he has several weak muscles. It should be noted that muscles that are usually weak are not the causes of CP patients' problems. This case causes problems for Dr. Gait-1. Later we will see that Dr. Gait-2 has no problems with this case.

The motion data given to the program is shown in table 4-1.

	Hip	Knee	Ankle
WA	decreased	increased	WNL[11]
first-half SLS	WNL	decreased	decreased
second-half SLS	WNL	decreased	decreased
WR	WNL	WNL	mildly decreased
first-half swing	WNL	WNL	WNL
second-half swing	increased	increased	WNL

Table 4-1: Motion Data for John Patient

4.3.2 The Analysis

Step 1: Dr. Gait-1 gathers information on the session number, patient name, etc. It also gathers the patient motion data shown in table 4-1 and velocity data.

Step 2: Backward chaining through the group rule base to classify the patient. The patient can be placed in one of three groups: low-velocity if velocity <= 35% of normal, middle-velocity if 35% < velocity <= 65% or high-velocity if velocity > 65%.

`Determining which group the patient falls into.`

[10] The program's output will be in normal type and comments will be in *italics*.

[11] WNL = within normal limits i.e. motion is within normal range

```
It was determined that the patient was in group: middle-velocity.
```

Step 3: Apply the rule bases to the phases and classify the motions as either primary-cause, compensation, lack-of-compensation, etc. The system needs to look at each joint in each phase. For each joint and phase, it examines the joint's motion relative to the other joints during the same phase and then classifies the joint's motion.

The system now begins to scan each joint in each phase. We will look at some of them in detail and then summarize the classifications reached by the system.

...

It is now trying to classify the increased knee flexion in WA.

```
Running the WA rule base.
In order to DX the knee.
```

The system begins to backward chain through the WA-rb. It finds that WA-rule-12 fires successfully concluding that the increased knee flexion is a primary-cause. Rule WA-rule-12 is:

```
WA-rule-12
 ifall
  (patient psm knee (current context-phase) motion-data)
    at-least mildly-increased
  (or (patient psm ankle (context current-phase) motion-data)
        same-as WNL
      (patient psm ankle (context current-phase) motion-data)
        at-least increased)
 then
  Conclude: (patient psm knee (context current-phase) dx)
    with value: primary-cause
```

This rule says that if the knee shows increased flexion and the ankle shows normal or increased plantarflexion then the knee flexion should be classified as a primary-cause

```
I am concluding for (patient psm knee WA Dx) that it should
be (primary-cause). This was done by rule WA-rb-rule12
```

...

The system now tries to classify the knee in the first-half of SLS.

```
Running the first-half-SLS rule base.
In order to DX the knee.

How would you characterize the quadricep's activity
during first-half-SLS? [unknown]

     1)  on
     2)  off
     3)  unknown
```

```
  Please enter choice(s):
     # 2
```

However, we see below that it cannot classify the deviation. There is no rule which handles the case where the knee has decreased flexion while the quadricep is off. Since the hip positioning is normal, the only way to account for the lack of flexion in the knee is to then assume that the gastroc/soleus is producing less of a flexion torque than normal. However, this is not what one would first guess as the cause of the knee's motion deviation.

Dr. Gait-1 asks the user for the answer since no rules were successful.

```
What is the Dx of the specified joint 's motion ?

patient
   psm
      knee
         first-half-SLS
            Dx

Please enter value:
     >> indeterminant

...
```

The same situation is evident at the knee for the second-half of SLS. Thus, once again the user must supply the information.

```
Running the second-half-SLS rule base.
In order to DX the knee.

...

What is the Dx of the specified joint 's motion ?

patient
   psm
      knee
         second-half-SLS
            Dx

Please enter value:
     >> indeterminant

...
```

Similarly, we could go through the other fifteen combinations of joints and phases to arrive at their diagnoses. Table 4-2 shows the result of all these rule-bases.

	Hip	*Knee*	*Ankle*
WA	primary cause	primary cause	normal
first-half SLS	normal	indeterminant	primary cause
second-half SLS	normal	indeterminant	primary cause
WR	lack of compensation	lack of compensation	primary cause
first-half swing	normal	normal	normal
second-half swing	primary cause	primary cause	normal

Table 4-2: Diagnoses for Joint Motions of John Patient

Step 4: Now the system needs to look across the phases to find global problems. The first step in looking across the phases is to find any common motion patterns for each of the joints. A motion pattern is a pattern of motion deviations of a joint across all the different phases. Table 4-3 shows the different motion patterns for the hip.

	WA	***1st-half SLS***	***2nd-half SLS***	***WR***	***1st-half Swing***	***2nd-half Swing***
pattern 1	N[12]	N	N	N	I	N
pattern 2	I	N	N	N	I	I
pattern 3	I	N	N	D	D	N
pattern 4	D	N	N	D	D	D
pattern 5	I	I	I	I	I	I
pattern 6	N	N	I	I	I	N
pattern 7	I	N	N	N	N	I

Table 4-3: Hip Motion Common Patterns

To try to find motion patterns Dr. Gait-1 does backward chaining through the motion-pattern-rb rule base. The rules are of the form:

```
motion-pattern-rb-rule11
  if-all
     (patient psm knee WA motion-data) at-least mildly increased
     (patient psm knee first-half-SLS motion-data) same-as WNL
     (patient psm knee second-half-SLS motion-data) same-as WNL
     (patient psm knee first-half-swing motion-data) same-as WNL
     (patient psm knee second-half-swing motion-data)
         at-least mildly increased
   then
     conclude: (patient psm knee motion-pattern)
       with value: pattern-1
```

Thus, to try each rule, the program must try to satisfy all the clauses in the if part of the rule. The problem with this approach is that a value of a parameter might have to be repeatedly found and checked. For example, the parameter (patient psm knee WA knee motion-data) had to be found and compared seven times in determining the knee motion pattern. This is obviously very inefficient. A better approach than using rules

[12] In the table, N = normal, I = increased, D = decreased.

would have been to use a discrimination tree. At the time that Dr. Gait-1 was built there was no facility to allow the making of discrimination trees so the technique adopted was chaining through rules in a rule base.

Backward chaining through the motion-pattern-rb three times, once for each joint, Dr. Gait-1 finds that the knee has knee pattern-3 and the ankle has ankle pattern-10. The hip motion did not match any pattern.

```
Finished analyzing down the phase - Now analyze across the phases.

I am concluding for (patient psm knee motion-pattern) that it
should be (pattern-3).
This was done by rule motion-pattern-rb-rule13.


I am concluding for (patient psm ankle motion-pattern) that it
should be (pattern-10).
This was done by rule motion-pattern-rb-rule23
```

Step 5: The next step in the analysis is to select paragraph templates which describe the problems of the patient's gait. This is done by having a rule base, choose-paragraphs-rb, which has forty rules. Each rule that succeeds in firing selects one or more paragraph templates. If a rule is being tried which requires more information than is known to make a decision, then the user is asked for further information. This rule base is forward chained and the names of all the paragraph templates chosen are stored in the patient frame. In the case at hand, one paragraph template that is chosen is called abnormal-dorsiflexion. The rule which selects this template is shown in figure 4-9. The template itself is shown in figure 4-10.

```
choose-paragraphs-rb-rule22
  ifany
     (patient psm ankle motion-pattern) eq pattern-6
     (patient psm ankle first-half-SLS motion-data)
       at-most mildly decreased
     (patient psm ankle second-half-SLS motion-data)
       at-most mildly decreased
  then
    conclude: (patient psm ankle paragraphs)
      with value: abnormal-dorsiflexion
```

Figure 4-9: Rule which Selects Abnormal-Dorsiflexion Paragraph Template

```
Abnormal dorsiflexion during ------ single limb stance is noted.
Gastroc/soleus activity ------  ------ is unable to counteract
body weight ankle dorsiflexion torque ------ ------.
```

Figure 4-10: Abnormal-Dorsiflexion Paragraph Template

As the choose-paragraphs-rb rule base proceeds it asks the user for some more information and selects several more paragraph templates in addition to the abnormal-dorsiflexion paragraph template.

```
How would you characterize the hamstring's activity
during SLS? [unknown]

     1)  on
     2)  off
     3)  unknown

  Please enter choice(s):
     # 2


I am concluding for (patient psm knee paragraphs) that it should
be (thrust). This was done by rule choose-paragraphs-rb-rule13

...

I am concluding for (patient psm hip paragraphs) that it should
be (hip-increase-swing2).
This was done by rule choose-paragraphs-rb-rule34.
```

Step 6: Now for all the paragraph templates the system has to fill in the blanks. Because each blank has a name, rules can refer to the different blanks. The rule base, set-contexts-rb, is the rule base which tries to fill in all the blanks of the chosen templates correctly. The rule which attempts to fill in the first slot of the abnormal-dorsiflexion is shown in figure 4-11. This rule states that if the paragraph template abnormal-dorsiflexion

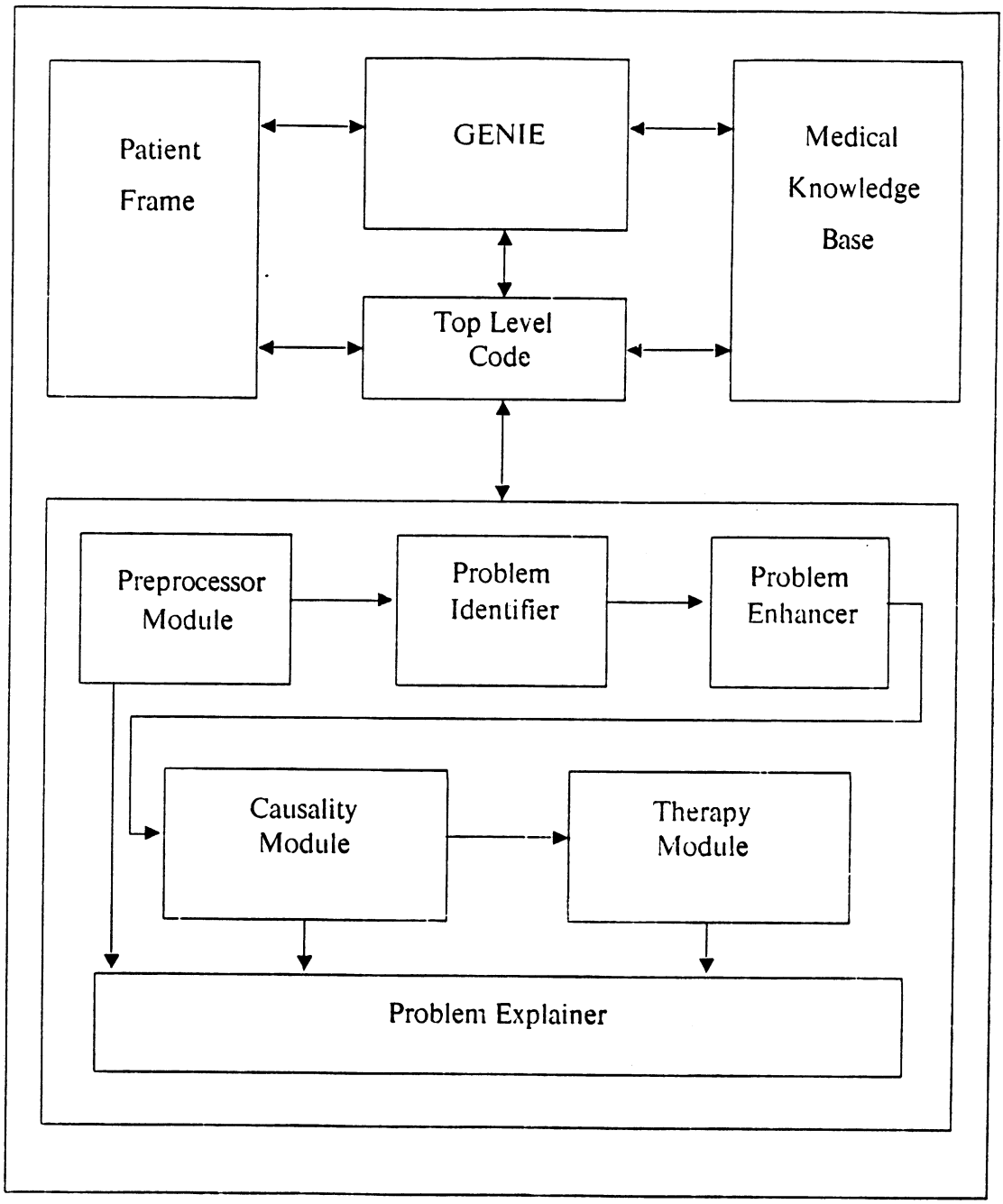


Figure 5-1: Dr. Gait-2 Organization

concluded the corresponding assertion or else a set of assumptions that support the assertion. The assumption system will be explained later in this chapter.

5.2.4 Preprocessor Module

The preprocessor's function is to set up the system so that it can begin its tasks of identifying, understanding and treating problems. The preprocessor consists of several parts. There is a program that takes the motion data stored in the data base[14] and converts it into scaled motion deviations. This transformation was shown in figure 3-1. There is also code which gathers the muscle and patient background data from the data base and the user. This module also invokes the same group rule base that classifies the patient into one of three velocity groups. Thus, after this module is run, the system has gathered all the necessary data and knows what joints in which phases of the gait cycle it should instruct the rest of the system to analyze.

5.2.5 Problem Identifier Module

By examining the gathered data, this module attempts to find the problems present in each patient. Each frame or class of problems has a knowledge frame associated with it. From these problem frames, we can obtain information on how to identify and set up a particular program, find both causes and explanation of it, as well as identify other necessary data. The reader should notice that the approach for dealing with problems has moved from a rule based approach of MYCIN [Shortliffe 76] to a more frame like approach of the Present Illness Program(PIP) [Pauker 76].

Currently there are three classes of problems: *limited range of motion class*, which deals with problems associated with very restricted ranges of motion throughout the gait cycle by any of the joints; *contracture class*, which deals with motion problems caused by tight muscles or tight joint capsules; and *motion deviation class*, which deals with any noticed motion deviation. These three classes have been sufficient to capture all of the problems encountered so far.

To identify a problem, Dr. Gait-2 looks in the problem-identify slot of the procedure's

[14] This is in the form of absolute joint angles for each 2% of the gait cycle.

corresponding frame. This slot either holds a procedure to run (usually for a whole class of problems) or a set of conditions necessary for the problem to exist. If the module determines that the problem is present, then the its set up procedure is called and it is added to the list in the patient frame.

5.2.6 Problem Enhancer Module

Once a problem has been identified, the system might want to take other actions. Such actions could include identifying other problems, refining the identity of a problem or anything else which adds to the understanding of the problem. Currently, none of the problems take advantage of this feature.

5.2.7 Causality Module

This module is the heart of the system; it does most of the analysis and takes advantage of Dr. Gait's knowledge of torques. To go back to an idea that was stated earlier, the motion at a joint is the result of the combination of torques acting upon this joint. Looking at the knee, for example, we can say that its motion is determined by all of the torques acting upon the knee. These torques are the hamstring-torque, the quadricep-torque, the gastroc/soleus-torque, the bodyweight-torque and the momentum-torque. If the knee's position is abnormal, then at least one of these torques must also be abnormal. Determining which torques are abnormal helps us to better understand the cause of the knee's motion deviation. If the knee shows increased flexion and we discover that the hamstring-torque is increased, we then know that the hamstring is contributing to the knee's problem.

We could say that the knee's torque must satisfy the equation:

knee-torque = hamstring-torque + quadricep-torque + gastroc/soleus-torque + bodyweight-torque + momentum-torque

However, knowing that the knee-torque is abnormal means that we now have to look at five torques at once to assign the blame. In most problems involving search and combination it helps to organize the entities. To organize our torques we need two concepts. First, we will say that a torque is a *flexion torque* if it normally causes flexion or analogously an *extension torque* if it

normally causes extension. Hence, for the knee, the quadricep-torque is an extension torque and the hamstring-torque and the gastroc/soleus-torque are flexion torques. The classification for the momentum-torque and the bodyweight-torque depends on which phase we are considering. Second, we will say that torque is *internal* if it is produced locally or *external* otherwise. Thus, all muscle torques are internal while the bodyweight and the momentum torques are external. Using these two concepts we can classify our torques and then organize them in a tree.[15] See figure 5-2.

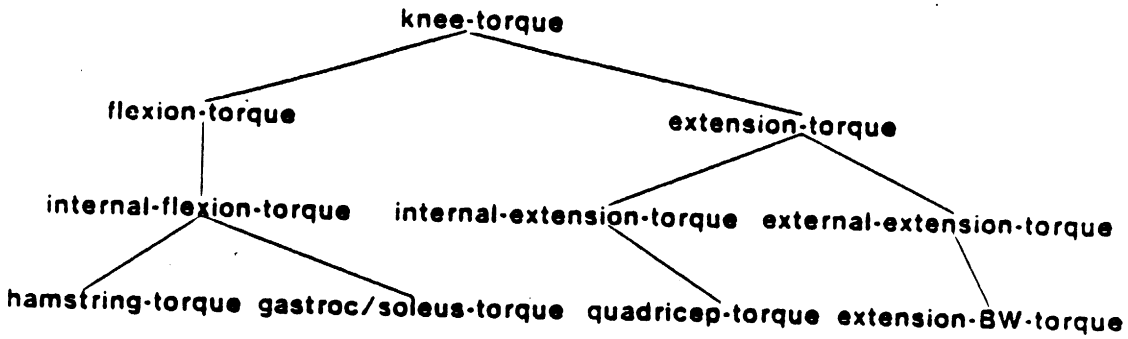


Figure 5-2: Tree of Torques for Knee in SLS

Using the tree as our guide we can get a set of torque equations for a joint by saying that a node equals the sum of its children. For the knee is SLS we get the equations:

[15] In Dr. Gait-2 the bodyweight-torque includes the effect of the body-weight and also momentum. Thus there is no momentum-torque.

```
knee-torque                = flexion-torque -
                             extension-torque
flexion-torque             = internal-flexion-torque
extension-torque           = internal-extension-torque +
                             external-extension-torque
internal-flexion-torque    = hamstring-torque +
                             gastroc/soleus-torque
internal-extension-torque  = quadricep-torque
external-extension-torque  = extension-BW-torque[16]
```

Now that we have these equations, how do we use them? We describe all our motion data as scaled deviations from normal. In a sense, we just say whether the joint's position is increased, decreased or normal. If we use the same descriptions for the torques then a torque will have a value of *increased*, *decreased* or *normal*. The next question is how to use these values in an equation. Fortunately others have used similar equations. They are called incremental qualitative (IQ) equations [de Kleer 79]. The rules for IQ equations are shown in figure 5-3.

Looking at figure 5-3, we see that in all but two cases we can get an answer for C. In the two unknown cases we need more information about A and B to find C. An alternative is that we can obtain a value for C if we make assumptions about A and B. We introduce the relations $<diff$, $>diff$ and $=diff$. The statement $A <diff B$ says that "A has a smaller deviation from normal then B". Thus, if A is increased, B is decreased and A <diff B, then we can conclude that C is decreased. Using these new relations we get a modified IQ algebra whose rules are shown in figure 5-3.

Since we do not have a measurement of the various torques, we do not know for certain the relationships between the torques. However, we can make assumptions about what the relationships might be. We can then obtain a value for C depending on which assumptions seem more appropriate. The strategy that will be used is to assume anything which is physically reasonable and then at the end, to use heuristic knowledge about the particular disease the patient has, to choose the best set of hypotheses. Thus, by making assumptions and using domain specific knowledge we can obtain a value for all torques.

[16] This is a bodyweight torque which is causing extension.

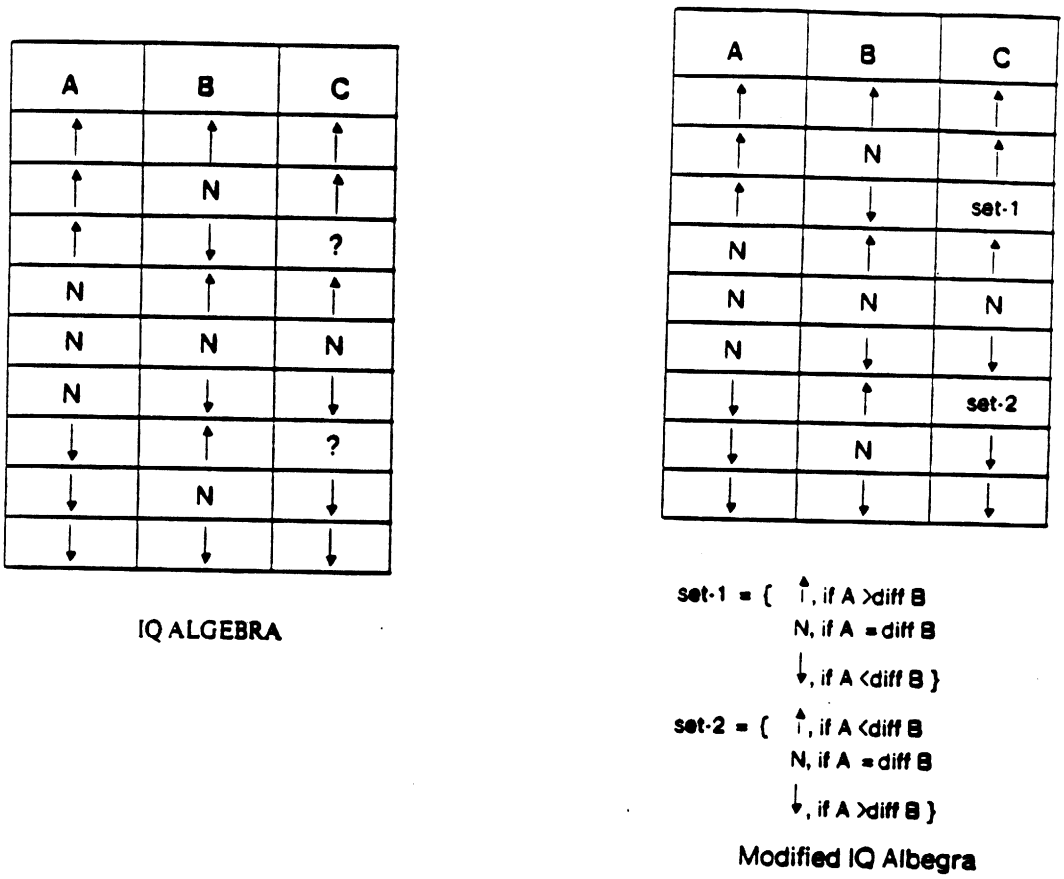


Figure 5-3: IQ Algebra

As we have just seen, to deal with the torque equations, we need to make assumptions. If the system is making assumptions ine in different places, we need to make sure that they are consistent. We do not want the system at some point to assume the quadriceps are overactive in WA and then later to assume that they they are underactive. To keep the assumptions organized, a support system was built up on de Kleer's idea of an assumption-based truth maintenance system [de Kleer 84]. This system allows one to keep a set of alternative computations on hand and compare them. The values maintained by the system [illegible]

6.1.2 Step 2: Identify the Problems

Dr. Gait-2 looks at the triggering conditions in the problem frames to identify the patient's gait problems. It then prints out the name of each problem it tries to identify and whether or not it succeeds in doing so. We see that Dr. Gait-2 identifies right-ankle-LROM which stands for "limited range of motion of the right ankle." The triggering condition for this problem is:

```
(patient data right ankle sagittal range-of-motion) >= 15
(patient data right ankle saggital range-of-motion) < 20
(direction? (patient data right ankle sagittal WA motion-data)
            'plantarflexion)
(direction? (patient data right ankle sagittal first-half-SLS
             motion-data)
            'plantarflexion)
(direction? (patient data right ankle sagittal second-half-SLS
             motion-data)
            'plantarflexion)
(direction? (patient data right ankle sagittal WR motion-data)
            'plantarflexion)
(direction? (patient data right ankle sagittal first-half-swing
             motion-data)
            'plantarflexion)
(direction? (patient data right ankle sagittal second-half-swing
             motion-data)
            'plantarflexion)
(patient data right gastroc/soleus activity-type) eq continuous
(patient data right anterior-tibialis activity-type)
    eq continuous
(patient conclusion group) neq low-velocity
```

The program will conclude that there is limited range of motion of the ankle if (1) the range of motion is between 15 and 20 degrees, (2) the deviations in motion are all in the direction of plantarflexion, (3) the gastroc/soleus and anterior tibialis show continuous activity, and (4) the patient is not classified in the low velocity group. This is all true for Jane Patient; consequently, the program concludes that she has limited range of motion of her right ankle.

Any motion deviation from normal is found by the identify motion class problem's procedure and it is entered into the system. Contracture class problems are identified in the same way that the limited range of motion class problems are identified. It turns out that Jane does have a right ankle contracture. Taking this all into account, Dr. Gait-2 identifies the following problems in Jane's gait:

```
Finished finding the problems.
They are
    right-ankle-LROM
    right-ankle-contracture
    right-hip-sagittal-WA-flexion[19]
    right-hip-sagittal-first-half-swing-flexion
    right-hip-sagittal-second-half-swing-flexion
    right-knee-sagittal-WA-flexion
    right-knee-sagittal-second-half-swing-flexion
    right-ankle-sagittal-WA-plantarflexion
    right-ankle-sagittal-first-half-SLS-plantarflexion
    right-ankle-sagittal-second-half-SLS-plantarflexion
    right-ankle-sagittal-WR-plantarflexion
    right-ankle-sagittal-first-half-swing-plantarflexion
    right-ankle-sagittal-second-half-swing-plantarflexion.
```

6.1.3 Step 3: Enhance the Problems

All that is done here is to see if the problem should be considered in more detail. In Jane's case all the problems are deemed worthy of further consideration.

6.1.4 Step 4: Find Problem Causes

The limited range of motion problem and the contracture problem of the ankle do not go through the causality module since their causes are known just by knowing that they exist. Furthermore, the information stored with the contracture problem causes the ankle's deviations in both halves of SLS to be reclassified as static problems. This means that the cause of the increased plantarflexion in these phases is the joint contracture and not the dynamic muscle activity. Thus, any problem that is classified as static will not be put through the causality module. The remaining motion problems must be processed by the causality module. We will go through the problem right-knee-sagittal-second-half-swing in detail. The relevant muscle data for the knee is shown in table 6-2.

Muscle	Usual Activity	Actual Activity	Strength
gastroc/soleus	off	on	unknown
hamstring	on	on	unknown
quadricep	off	off	not weak

Table 6-2: Example 1: Knee Muscle Data

[19] This problem name identifies increased right hip flexion in WA.

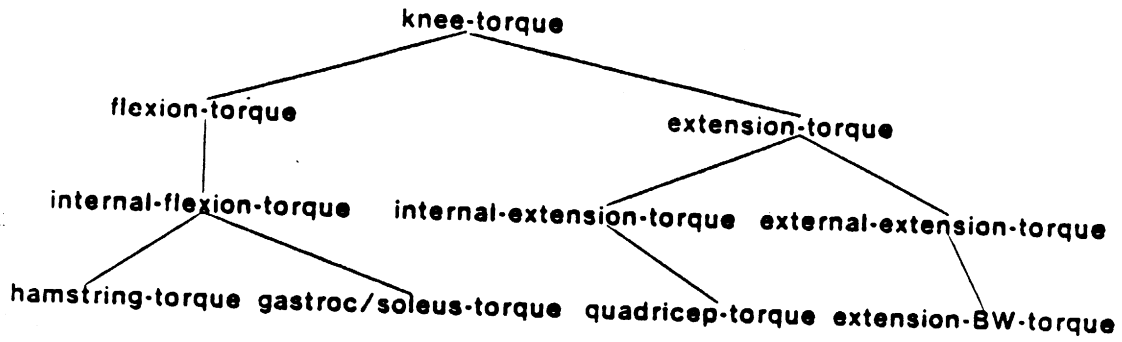


Figure 6-1: Torque Tree for Knee in Second-half-swing

The torque tree in figure 6-1 is instantiated. This gives us the following torque equations:

```
level1: knee-torque              = flexion-torque -
                                   extension-torque
level2: flexion-torque           = internal-flexion-torque +
                                   external-flexion-torque
        extension-torque         = internal-extension-torque +
                                   external-extension-torque
level3: internal-flexion-torque = hamstring-torque +
                                   gastroc/soleus-torque
        internal-extension-torque = quadricep-torque
        external-extension-torque = extension-BW-torque
level4: hamstring-torque  <---- from discrimination nets
        gastroc/soleus-torque <---- from discrimination nets
        quadricep-torque      <---- from discrimination nets
        extension-BW-torque   <---- from body weight rules
```

The system starts with the equations at level 4 and works its way up the tree to level 1.

```
Using the discrimination nets and body weight rule base Dr. Gait-2
gets the following value for the level 4 torques:
hamstring-torque = <increased, {hamstring not weak}>,
```[20]

[20]The torque values are of the form <V, {A1,A2,A3...}> where V is the assertion of the torque's value and the Ai's are the assumptions in the assumption set.

```
                    <decreased, {hamstring weak}>

gastroc/soleus-torque = <increased, {}>

quadricep-torque  = <increased, {}>

extension-BW-torque = <decreased, {}>
```

Using the modified IQ algebra and the torque equations we move up to level 3 and get:

```
internal-flexion-torque
         = <increased, {hamstring not weak}>,
           <increased, {gastroc/soleus-torque >diff
                        hamstring-torque, hamstring weak}>,
           <normal, {gastroc/soleus-torque =diff
                     hamstring-torque, hamstring weak}>,
           <decreased, {gastroc/soleus-torque <diff
                        hamstring-torque, hamstring weak}>

internal-extension-torque = <normal, {}>

external-extension-torque = <decreased, {}>
```

Again using the torque equations we move up to level 2, where we get the following values:

```
flexion-torque = <increased, {hamstring not weak}>,
                 <increased, {gastroc/soleus-torque >diff
                              hamstring-torque, hamstring weak}>,
                 <normal, {gastroc/soleus-torque =diff
                           hamstring-torque, hamstring weak}>,
                 <decreased, {gastroc/soleus-torque <diff
                              hamstring-torque, hamstring weak}>

extension-torque = <decreased, {}>
```

We use the torque equations a final time to get values for the knee-torque:

```
knee-torque
     = <increased, {gastroc/soleus-torque =diff hamstring-torque,
                    hamstring weak}>,
       <increased, {hamstring not weak}>
       <increased, {gastroc/soleus-torque >diff hamstring-torque,
                    hamstring-weak}>,
       <increased, {flexion-torque <diff extension-torque,
                    gastroc/soleus-torque <diff hamstring-torque,
                    hamstring weak}>,
       <normal,    {flexion-torque =diff extension-torque,
                    gastroc/soleus-torque <diff hamstring-torque,
                    hamstring weak}>,
       <decreased, {flexion-torque >diff extension-torque,
```

```
                    gastroc/soleus-torque <diff hamstring-torque,
                    hamstring weak}>
```

Imposing the known constraint that the knee-torque is increased causes the system to remove inconsistent values. Thus the remaining values for the knee torque[21] are:

```
knee-torque
     = <increased, {hamstring not weak}>
       <increased, {gastroc/soleus-torque =diff
                    hamstring-torque, hamstring weak}>,
       <increased, {gastroc/soleus-torque >diff hamstring-torque,
                    hamstring weak}>,
       <increased, {flexion-torque <diff extension-torque,
                    gastroc/soleus-torque <diff hamstring-torque,
                    hamstring weak}>
```

This leaves us with four possible explanations for the increased knee flexion in the second half of swing. The four corresponding assumption sets are sent to a scoring routine which uses heuristic knowledge about Jane's disease (CP in her case) to determine which assumptions are more likely and thus score each assumption and assumption set. The top scoring assumption set is "{hamstring not weak}"; it is chosen as the best possible explanation.

All of the other problems are fed to the causality module in a similar manner and an similar analysis is done.

6.1.5 Step 5: Find Therapy

This section illustrates how the therapy module processes this case. First the problems are ranked according to the phase in which they occur, the joint where they occur, and the severity of the problem. This gives the following ranking of the problems:

```
right-ankle-sagittal-first-half-swing-plantarflexion
right-hip-sagittal-second-half-swing-flexion
right-knee-sagittal-second-half-swing-flexion
right-ankle-sagittal-second-half-swing-plantarflexion.
right-ankle-sagittal-WA-plantarflexion
right-knee-sagittal-WA-flexion
right-hip-sagittal-WA-flexion
right-ankle-sagittal-WR-plantarflexion.
right-ankle-LROM
```

[21] Any inconsistent values for the lower torques are also removed.

```
right-ankle-contracture
```

The first pass of the therapy analysis must find the possible therapies and classify them for each of the problems. The program first examines the problem right-ankle-sagittal-first-half-swing-plantarflexion. In doing so it finds that possible bracings are rigid ankle brace and a flexible ankle brace. Each of these bracing therapies check against the appropriate rules to see how they should be classified. Both of these therapies are classified as ok. The possible therapies and their classifications for this problem are shown below. Similarly, possible therapies and classifications are obtained for the other problems.

```
Possible physical therapies: ankle-physical-therapy (ok)
Possible bracings: ankle-rigid-brace (ok)
                   ankle-flexible-brace (ok)
Possible surgeries: gastroc/soleus-lengthening (ok)
                    peroneal-lengthening (ok)
                    peroneal-split-transfer (ok)
                    peroneal-complete-transfer (ok)
```

Each therapy type for each joint receives a score that is determined by the classification of the therapy types for each joint. If a therapy type is "ruled out" or "not desired" then the score is decreased but if a therapy type is "ok" then the score is increased. Each therapy type has four scores. One score for each joint and a global score which is the total of these other three scores.

In the second pass of the analysis the system must pick the desired therapy for each problem. The first step is to pick the preferred therapy type for each joint. The preferred therapy is the therapy type with the highest score for the joint. The global scores are used to break any ties. If there is still a tie after examining the global scores then the least severe therapy is chosen. For the ankle the program chooses surgery, for the knee or the hip physical therapy. Now it needs to select the actual therapies. Looking at the problem right-ankle-sagittal-first-half-swing-plantarflexion it now needs to pick from among the four surgeries that were listed above. It looks at how each of the four surgeries effects the patient's other problems. The gastroc/soleus lengthening is found to be the most helpful and thus it is selected. The system then realizes that this treatment will help all of the ankle and knee problems. Thus, the only problem left to treat

is the increased hip flexion in weight acceptance and the second half of swing. Since physical therapy is the desired therapy type for hip problems, the program selects hip physical therapy to help improve the hip's increased flexion.

In summary, the system selects gastroc/soleus lengthening to improve the ankle plantarflexion which occurs through out the gait cycle and the knee flexion which occurs during WA and the second half of swing. The system also recommends physical therapy to help improve the increased hip flexion which occurs during WA and the second half of swing.

6.1.6 Step 6: Write out the Conclusions

```
Problem name: right-ankle-LROM
Problem summary:
  The right ankle shows moderate limited range of motion.
  This appears to be due to co-contracture of
   gastroc/soleus anterior-tibialis which is modified by body-weight.


Problem name: right-ankle-sagittal-first-half-SLS-plantarflexion
Problem summary:
  The right ankle has increased plantarflexion during first-half-SLS.

The primary cause of this problem is:

  Problem name: right-ankle-contracture
  Problem summary:
    The right ankle has a joint contracture which is the cause
    of the noticed deviation.


Problem name: right-ankle-sagittal-second-half-SLS-plantarflexion
Problem summary:
  The right ankle has increased plantarflexion during second-half-SLS.

The primary cause of this problem is:

  Problem name: right-ankle-contracture
  Problem summary:
    The right ankle has a joint contracture which is the cause
    of the noticed deviation.


Problem name: right-ankle-sagittal-second-half-swing-plantarflexion
Problem summary:
  The right ankle has increased plantarflexion during
  second-half-swing.
```

```
Assuming the following:
1) (patient hypotheses right ankle sagittal second-half-swing
    dorsiflexion-torque)
      has a smaller deviation from normal than
   (patient hypotheses right ankle sagittal second-half-swing
    plantarflexion-torque).

The PRIMARY CAUSE(s) of this problem is(are):

increased gastroc/soleus-torque. which is due to abnormal-firing
of a functionally-spastic gastroc/soleus.
increased peroneal-torque. which is due to abnormal-firing
of a functionally-spastic peroneal.

The COMPENSATION(s) for this problem is(are):

increased anterior-tibialis-torque. which is due to normal-firing
of a non-functionally-spastic anterior-tibialis.


Problem name: right-ankle-sagittal-first-half-swing-plantarflexion
Problem summary:
 The right ankle has increased plantarflexion during first-half-swing.
Assuming the following:
1) (patient hypotheses right ankle sagittal first-half-swing
    dorsiflexion-torque)
     has a smaller deviation from normal than
   (patient hypotheses right ankle sagittal first-half-swing
    plantarflexion-torque).

The PRIMARY CAUSE(s) of this problem is(are):

increased gastroc/soleus-torque. which is due to abnormal-firing
of a functionally-spastic gastroc/soleus.
increased peroneal-torque. which is due to abnormal-firing
of a functionally-spastic peroneal.

The COMPENSATION(s) for this problem is(are):

increased anterior-tibialis-torque. which is due to normal-firing
of a non-functionally-spastic anterior-tibialis.


Problem name: right-ankle-sagittal-WR-plantarflexion
Problem summary:
  The right ankle has increased plantarflexion during WR.
Assuming the following:
1) (patient hypotheses right ankle sagittal WR dorsiflexion-torque)
      has a smaller deviation from normal than
   (patient hypotheses right ankle sagittal WR plantarflexion-torque).

The PRIMARY CAUSE(s) of this problem is(are):
```

```
increased gastroc/soleus-torque. which is due to abnormal-firing
of a functionally-spastic gastroc/soleus.
increased peroneal-torque. which is due to abnormal-firing
of a functionally-spastic peroneal.

The COMPENSATION(s) for this problem is(are):

increased dorsiflexion-BW-torque.
increased anterior-tibialis-torque. which is due to abnormal-firing
of a non-functionally-spastic anterior-tibialis.


Problem name: right-ankle-sagittal-WA-plantarflexion
Problem summary:
  The right ankle has increased plantarflexion during WA.
Assuming the following:
1) (patient hypotheses right ankle sagittal WA dorsiflexion-torque)
      has a smaller deviation from normal than
   (patient hypotheses right ankle sagittal WA plantarflexion-torque).

The PRIMARY CAUSE(s) of this problem is(are):

increased gastroc/soleus-torque. which is due to abnormal-firing
of a functionally-spastic gastroc/soleus.
increased peroneal-torque. which is due to abnormal-firing
of a functionally-spastic peroneal.

The COMPENSATION(s) for this problem is(are):

increased dorsiflexion-BW-torque.
increased anterior-tibialis-torque. which is due to normal-firing
of a non-functionally-spastic anterior-tibialis.


Problem name: right-knee-sagittal-second-half-swing-flexion
Problem summary:
  The right knee has increased flexion during second-half-swing.
Assuming the following:
1) (patient data right hamstring muscle-strength) equals notweak.

The PRIMARY CAUSE(s) of this problem is(are):

increased hamstring-torque. which is due to normal-firing
of a functionally-spastic hamstring.
increased gastroc/soleus-torque. which is due to abnormal-firing
of a functionally-spastic gastroc/soleus.

The AUXILIARY CAUSE(s) of this problem is(are):

decreased extension-BW-torque.

Problem name: right-knee-sagittal-WA-flexion
```

```
Problem summary:
  The right knee has increased flexion during WA.
Assuming the following:
1) (patient hypotheses right knee sagittal WA extension-torque)
      has a smaller deviation from normal than
   (patient hypotheses right knee sagittal WA flexion-torque).

2) (patient data right hamstring muscle-strength) equals notweak.

The PRIMARY CAUSE(s) of this problem is(are):

increased hamstring-torque. which is due to normal-firing
of a functionally-spastic hamstring.
increased gastroc/soleus-torque. which is due to abnormal-firing
of a functionally-spastic gastroc/soleus.

The COMPENSATION(s) for this problem is(are):

increased quadricep-torque. which is due to normal-firing
of a non-functionally-spastic quadricep.

Problem name: right-hip-sagittal-second-half-swing-flexion
Problem summary:
  The right hip has increased flexion during second-half-swing.
Assuming the following:
1) (patient hypotheses right hip sagittal second-half-swing
    extension-torque)
     has a smaller deviation from normal than
   (patient hypotheses right hip sagittal second-half-swing
    flexion-torque).

2) (patient data right hamstring muscle-strength) equals notweak.

The PRIMARY CAUSE(s) of this problem is(are):

increased iliopsoas-torque. which is due to abnormal-firing
of a functionally-spastic iliopsoas.

The COMPENSATION(s) for this problem is(are):

increased hamstring-torque. which is due to normal-firing
of a non-functionally-spastic hamstring.


Problem name: right-hip-sagittal-WA-flexion
Problem summary:
  The right hip has increased flexion during WA.
Assuming the following:
1) (patient hypotheses right hip sagittal WA extension-torque)
      has a smaller deviation from normal than
   (patient hypotheses right hip sagittal WA flexion-torque).
```

```
2) (patient data right gluteus-maximus muscle-strength)
   equals notweak.

3) (patient data right hamstring muscle-strength) equals notweak.

The PRIMARY CAUSE(s) of this problem is(are):

increased flexion-BW-torque.
increased quadricep-torque. which is due to normal-firing
of a functionally-spastic quadricep.

The COMPENSATION(s) for this problem is(are):

increased gluteus-maximus-torque. which is due to normal-firing
of a non-functionally-spastic gluteus-maximus.
increased hamstring-torque. which is due to normal-firing
of a non-functionally-spastic hamstring.
```

6.2 Example 2

Our second example deals with John Patient. His case is more difficult than Jane's because his major problems are related to muscle weakness, which is unusual for a CP patient. Instead of looking at all of the steps of the analysis in detail, we will examine the parts of the analysis that highlight how Dr. Gait-2 deals with the problems that Dr. Gait-1 could not handle properly.

6.2.1 Finding the Correct Causes for the Knee Problems

Dr. Gait-1 had a lot of trouble in dealing with John's knee. It could not classify the knee's decreased flexion in SLS and it concluded incorrectly that overactive hamstrings were the cause of the increased knee flexion in WA and the second-half of swing. The program also incorrectly suggested hamstring lengthening to help fix the increased knee flexion. We saw that these problems occurred because Dr. Gait-1 did not confirm its implicit conclusions with the muscle data, and its "hard wired" causes made the most likely cause the only possible cause.

Dr. Gait-2 does not suffer from these deficiencies. It uses the muscle data and through the torque equations makes sure it is consistent with the motion data. In this example, if we had stated that the quadricep was definitely not weak, then the program would have found that the data was inconsistent for the knee in SLS. However, we make no such restriction here and we will thus see that a weak quadricep is the cause of the knee's problems in SLS.

As Dr. Gait-2 starts going through the problems to find their causes, it begins with the decreased hip flexion in WA. Several reasonable explanations for the hip's deviation are found. It scores all of them, and we discover that assuming the quadricep is *not weak* yields the most likely explanation. This explanation says that a spastic gluteus maximus is the primary cause of the problem with compensation from a weak hamstring and a decreased flexion body weight torque.

This works well until the increased knee flexion in WA is examined. Then we see the following:

...

The system discovers that assuming the quadricep is not weak does not allow for the increased knee flexion in WA.

```
removing the assumption set
  ((eq (patient data right quadricep muscle-strength) notweak))
```

...

We see that the system wants to assume that the quadricep is weak. However, this contradicts the assumption that the quadriceps are not weak which was made in examining the hip in WA. Thus, some adjustments need to be made.

```
Cannot add desired environment to right-knee-sagittal-WA-flexion.

The environment is
((>diff (patient hypotheses right knee sagittal WA extension-torque)
        (patient hypotheses right knee sagittal WA flexion-torque))
 (eq (patient data right quadricep muscle-strength) weak))
Will try to make appropriate adjustments.

Looking at other problems to see if can change their
environments to maintain consistency.

I added the environment
((>diff (patient hypotheses right knee sagittal WA extension-torque)
        (patient hypotheses right knee sagittal WA flexion-torque))
 (eq (patient data right quadricep muscle-strength) weak))
to right-knee-sagittal-WA-flexion

I added the environment
((>diff (patient hypotheses right gluteus-maximus WA
         gluteus-maximus-torque)
```

```
        (patient hypotheses right hamstring WA hamstring-torque))
 (eq (patient data right quadricep muscle-strength) weak))
to right-hip-sagittal-WA-extension
```

We see that the chosen environment for the hip in WA problem has been changed to maintain consistency. Thus, the hip in WA also is assuming that the quadricep is weak.

Dr. Gait-2 has no problem with the knee in the first-half and second-half of SLS.

```
I am concluding for
  (patient hypotheses right knee sagittal first-half-SLS
   extension-BW-torque value)
that it should be (decreased). This was done by rule
body-weight-rb-rule11.
I found the bodyweight to be abnormal

(patient hypotheses right knee sagittal first-half-SLS
 extension-ext-torque)
 has the value(s)
 (decreased).

(patient hypotheses right knee sagittal first-half-SLS
 extension-int-torque)
 has the value(s)
 (wnl).

(patient hypotheses right knee sagittal first-half-SLS
 extension-torque)
 has the value(s)
 (decreased).

(patient hypotheses right knee sagittal first-half-SLS
 flexion-ext-torque)
 has the value(s)
 (zero-torque).

(patient hypotheses right knee sagittal first-half-SLS
 flexion-int-torque)
 has the value(s)
 (increased decreased).

(patient hypotheses right knee sagittal first-half-SLS flexion-torque)
 has the value(s)
 (decreased increased).

(patient hypotheses right knee sagittal first-half-SLS knee-torque)
 has the value(s)
 (decreased wnl increased).
```

It now removes the assumptions that lead to improper values for the knee torque.

...

Similarly it continues finding the causes of the remaining problems.

The system now writes out its conclusions.

```
Problem name: right-ankle-sagittal-WR-dorsiflexion
Problem summary:
  The right ankle has increased dorsiflexion during WR.
```

This problem was not examined in detail because it is only a mild deviation

```
This problem was not examined in detail.


Problem name: right-ankle-sagittal-second-half-SLS-dorsiflexion
Problem summary:
  The right ankle has increased dorsiflexion during second-half-SLS.
Assuming the following:
1) (patient hypotheses right gastroc/soleus second-half-SLS
    gastroc/soleus-torque)
       has a greater deviation from normal than
   (patient hypotheses right peroneal second-half-SLS
    peroneal-torque).

2) (patient data right gastroc/soleus muscle-strength) equals weak.

The PRIMARY CAUSE(s) of this problem is(are):

increased dorsiflexion-BW-torque.
increased anterior-tibialis-torque. which is due to abnormal-firing
of a functionally-spastic anterior-tibialis.

The AUXILIARY CAUSE(s) of this problem is(are):

decreased gastroc/soleus-torque. which is due to normal-firing
of a non-functionally-spastic gastroc/soleus.

The COMPENSATION(s) for this problem is(are):

increased peroneal-torque. which is due to normal-firing
of a non-functionally-spastic peroneal.


Problem name: right-ankle-sagittal-first-half-SLS-dorsiflexion
Problem summary:
  The right ankle has increased dorsiflexion during first-half-SLS.
Assuming the following:
1) (patient hypotheses right gastroc/soleus first-half-SLS
    gastroc/soleus-torque)
       has a greater deviation from normal than
   (patient hypotheses right peroneal first-half-SLS
    peroneal-torque).
```

```
2) (patient data right gastroc/soleus muscle-strength) equals weak.

The PRIMARY CAUSE(s) of this problem is(are):

increased dorsiflexion-BW-torque.
increased anterior-tibialis-torque. which is due to abnormal-firing
of a functionally-spastic anterior-tibialis.

The AUXILIARY CAUSE(s) of this problem is(are):

decreased gastroc/soleus-torque. which is due to normal-firing
of a non-functionally-spastic gastroc/soleus.

The COMPENSATION(s) for this problem is(are):

increased peroneal-torque. which is due to normal-firing
of a non-functionally-spastic peroneal.


Problem name: right-knee-sagittal-second-half-swing-flexion
Problem summary:
  The right knee has increased flexion during second-half-swing.
Assuming the following:
1) (patient hypotheses right knee sagittal second-half-swing
    extension-ext-torque)
        has a greater deviation from normal than
   (patient hypotheses right knee sagittal second-half-swing
    extension-int-torque).

2) (patient hypotheses right knee sagittal second-half-swing
    extension-torque)
      has a greater deviation from normal than
   (patient hypotheses right knee sagittal second-half-swing
    flexion-torque).


The AUXILIARY CAUSE(s) of this problem is(are):

decreased extension-BW-torque.

The COMPENSATION(s) for this problem is(are):

decreased hamstring-torque. which is due to normal-firing
of a weak hamstring.
increased quadricep-torque. which is due to abnormal-firing
of a non-functionally-spastic quadricep.


Problem name: right-knee-sagittal-second-half-SLS-extension
Problem summary:
  The right knee has increased extension during second-half-SLS.
Assuming the following:
```

```
1) (patient hypotheses right knee sagittal second-half-SLS
    extension-torque)
       has a smaller deviation from normal than
   (patient hypotheses right knee sagittal second-half-SLS
    flexion-torque).

2) (patient data right gastroc/soleus muscle-strength) equals weak.


The AUXILIARY CAUSE(s) of this problem is(are):

decreased gastroc/soleus-torque. which is due to normal-firing
of a non-functionally-spastic gastroc/soleus.

The COMPENSATION(s) for this problem is(are):

decreased extension-BW-torque.


Problem name: right-knee-sagittal-first-half-SLS-extension
Problem summary:
  The right knee has increased extension during first-half-SLS.
Assuming the following:
1) (patient hypotheses right knee sagittal first-half-SLS
    extension-torque)
       has a smaller deviation from normal than
   (patient hypotheses right knee sagittal first-half-SLS
    flexion-torque).

2) (patient data right gastroc/soleus muscle-strength) equals weak.


The AUXILIARY CAUSE(s) of this problem is(are):

decreased gastroc/soleus-torque. which is due to normal-firing
of a non-functionally-spastic gastroc/soleus.

The COMPENSATION(s) for this problem is(are):

decreased extension-BW-torque.


Problem name: right-knee-sagittal-WA-flexion
Problem summary:
  The right knee has increased flexion during WA.
Assuming the following:
1) (patient hypotheses right knee sagittal WA extension-torque)
       has a greater deviation from normal than
   (patient hypotheses right knee sagittal WA flexion-torque).

2) (patient data right quadricep muscle-strength) equals weak.
```

The AUXILIARY CAUSE(s) of this problem is(are):

decreased quadricep-torque. which is due to normal-firing
of a non-functionally-spastic quadricep.

The COMPENSATION(s) for this problem is(are):

decreased flexion-BW-torque.
decreased hamstring-torque. which is due to normal-firing
of a weak hamstring.

Problem name: right-hip-sagittal-second-half-swing-flexion
Problem summary:
  The right hip has increased flexion during second-half-swing.

The PRIMARY CAUSE(s) of this problem is(are):

increased quadricep-torque. which is due to abnormal-firing
of a functionally-spastic quadricep.

The AUXILIARY CAUSE(s) of this problem is(are):

decreased hamstring-torque. which is due to normal-firing
of a weak hamstring.

Problem name: right-hip-sagittal-WA-extension
Problem summary:
  The right hip has increased extension during WA.
Assuming the following:
1) (patient hypotheses right gluteus-maximus WA

7. Performance of Dr. Gait-2

7.1 Testing Process

To test the performance of Dr. Gait-2, the domain expert selected a set of cases which would cover a wide range of the types of patients the program might be expected to deal with. The expert selected 22 cases. We then entered the data for these 22 cases into the computer and made Dr. Gait-2 analyze each of them. To analyze the results we decided to compare Dr. Gait-2's output to the gait lab's written reports. When the reports were incomplete[23] we asked the expert to determine if the conclusions reached by the program were reasonable or not.

Dr. Gait-2 has three tasks to accomplish - identify problems, determine their causes and recommend treatments. We decided to look first at how well Dr. Gait-2 performed on each of these three tasks and then draw our final conclusions about the performance of the program. The only problems, causes and treatments examined were those dealing with the events in the sagittal plane since this is the only plane that Dr. Gait-2 currently deals with.

7.2 Testing Results

7.2.1 Identification of Problems

In reviewing the 22 written reports we found a total of 170 problems. Of these problems, Dr. Gait-2 identified 151, which is almost 89% of the problems. Most of the problems that Dr. Gait-2 omitted were limited range of motion problems. It appears that the triggering conditions for this class of problems is too restrictive. Such problems as plantarflexion thrusts and attempts at dorsiflexion were also omitted. The system would need to gather additional data and have problem frames with triggering conditions to be able to identify these omitted problems.

Dr. Gait-2 also identified 46 problems which the reports did not mention. Some of these additional problems were minor. The others were probably not mentioned because they were either not found or were perceived to be insignificant.

[23] Not all problems had causes listed with them. Some reports did not list any therapies.

It appears that Dr. Gait-2 is very successful at identifying problems. With some refinements to the knowledge base it is possible that its performance on this task could be made even better.

7.2.2 Finding Causes of Problems

Dr. Gait-2 is also adept at identifying the causes of the problems. For the 196 problems it identified it only found incorrect causes nine times. Thus, Dr. Gait-2 found the correct causes 96% of the time. Six of the nine mistakes took place because the system thought the quadricep has more power to flex the hip than it actually does. The system knows how muscles act on a joint but not to what degree they can move the joint. Two of the nine mistakes occurred when it failed to recognize body weight as the cause of the problem and when there was excessive knee flexion at the knee in WA. This suggests that there is a missing rule in the body weight rule base. The only remaining problem that was misdiagnosed is the attributing of too much flexing power to the gastroc/soleus at the knee during swing. We thus see that the system does very well at determining the causes of the identified problems. Most of its mistakes occur because it doesn't know to what degree particular muscles can influence the various joints.

7.2.3 Therapy Selection

After analyzing the 22 cases, Dr. Gait-2 recommended the correct therapy ten times, partially correct therapy ten times, and incorrect therapy once.[24]

There are two situations in which Dr. Gait-2 has problems in prescribing therapies. In six of the cases it suggests gastroc/soleus-lengthening when it is inappropriate. The problem is that the program doesn't check for the effect of a therapy on *normal* motions. In all of these cases the patient had normal ankle motion in single limb stance and thus lengthening the gastroc/soleus will hurt the patient's gait severely. The surgery would prevent the patient from keeping his ankle from flexing during stance.

The other problems concern bracing. If the patient has a joint contracture which is severe enough to prevent him from moving his joint into a neutral position then he cannot wear a brace.

[24] An astute reader might notice that this only covers 21 cases. One case was not considered because it had faulty data.

This situation occurred a few times. To alleviate the problem of the system incorrectly recommending bracing in this case, the system could ask an additional question to assess the severity of any observed joint contracture and eliminate bracing when it is inappropriate.

Thus, Dr. Gait-2 did reasonably well in prescibing therapies but we noted two deficiencies that hinder its performance: not considering the effect of contractures on braces, and not considering the effect of therapies on normal motions. These two deficiencies would have to be corrected in order for Dr. Gait-2's performance to be improved.

7.3 Conclusions About Performance

Overall the system performs well at all three of its tasks. Most of its problems seem to be due to lack of knowledge. One problem is the the lack of understanding of the relative influences of different causes on a problem. The system might find that the cause of the knee flexion is due to the gastroc/soleus, hamstring and body weight but it does not realize that the hamstring and body weight are the major culprits. The other problem is not checking the effect of a considered therapy on normal motions. One could change the therapy module so it not only checks the effect of a therapy on all of the problems but also on the normal motions as well It is hard to say how difficult it would be to do this.

8. Conclusion

8.1 Summary

Dr. Gait-2 is an expert system that analyzes human gait. Using a torque model of gait, it is able to capture enough principles of gait to do causal reasoning. This torque model leads to a system which is more robust than previous gait analysis systems. We saw that it can deal with cases that the prototype system, Dr. Gait-1, cannot. Furthermore, its model of gait is not disease specific. Thus, the gait model developed can be used in systems which diagnose gait disorders of a different etiology.

The first AIM systems made the assumption that there was only a single cause of the patient's problems. Later systems, such as ABEL, tried to relax this constraint and allow for multiple causes of the patient's problems. We have seen that in doing gait analysis for cerebral palsy patients, there are almost always multiple causes. We saw that Dr. Gait-2 is able to deal with these multiple problems by using a model of gait. The qualitative torque model of gait gives Dr. Gait-2 a framework to keep track of the interactions of the various problems. The body weight torques and the torques of the two joint muscles provides the mechanism for capturing the interactions. By keeping the interactions and assumptions explicit the system is able to arrive at possible causes for each motion deviation. By then using disease specific knowledge the system can choose the most probable cause from among the possible ones.

8.2 Future Directions

Dr. Gait-2 is not the final solution for a gait analysis program. There are still several limitations that need to be overcome and future directions that merit further exploration.

Dr. Gait-2's treatment of time is very elementary. The gait cycle is divided up into a fixed number of phases and each phase is treated as a single point in time. Dr. Gait-2's only knowledge about these phases is that one phase precedes another. Thus, the system merely looks at the gait data at six instances. This leads to conclusions that describe only what is happening in each

phase. The system's limited notion of time makes it hard for it to produce a more global summary. A possible solution to this is to allow the significant phases to be determined dynamically as opposed to statically as it is currently implemented. The system could determine which events in the patient's gait cycle are significant and then make a phase out of each time period between significant events. Furthermore, the time tag for a conclusion should be allowed to grow if possible. For example, if the system determines that hamstring lengthening should be recommended for the knee in WA the first and second half of SLS, then the system should say that hamstring lengthening is recommended for the knee to correct the deviations in stance.

The qualitative torque model of gait needs to be improved. The model's knowledge of how torques are produced by the muscles should be enhanced. Currently, it only takes the emg signal and the muscle strength into account. However, the torque of the muscle is also affected by the position of the muscle relative to the joint. This positioning changes as the joints change their relative positions. The system also could use knowledge of the relative strengths of competing muscles at a joint. It would be helpful to know that the gastroc/soleus contributes about 85% of the plantarflexion torque while the peroneals only contribute about 15% of the plantarflexion torque. Thus, if the ankle shows increased plantarflexion and both the gastroc/soleus and the peroneals are spastic, then it should know that it would not be very beneficial to correct the peroneal and not the gastroc/soleus.

The model cannot handle all situations. In Dr. Gait-2, if the system determines that there is a joint contracture, then it reclassifies the appropriate motion deviations as static and they do not go through the causality module. A better solution would be to encompass the contracture in the model of gait. The contracture would have to be some special torque that turned on only when the joint angle was below a certain value and when on could not be overcome by any opposing torque. Currently, the gait model also cannot deal with patients who are using assistive devices such as a walker or crutches. One answer to this would be to change the body weight rule base so that it would take assistive devices into account when figuring out the body weight torque.

Gait analysis is a difficult skill to learn. It is possible that this gait analysis system could be modified for use in teaching. Such a modification would require extensive work on the system. The need for a good user interface would then be essential. The system would also need to formulate strategies on how to teach gait analysis. The system would have to have a model of the student to know better how to tailor the teaching to the particular student. The control of the modeling of gait also might have to be changed. In Dr. Gait-2 the torques are arranged in a tree. The torques get values by first filling in the values for the leaves and then percolating the values up to the root. One could imagine that if a student were using the system he might try to hypothesize about a value for a particular torque and then see its effect. This means that the system should allow any torque value to be set or retracted at any time. To do this the assumption based truth maintenance system would have to include justifications for each assertion, as was originally intended by de Kleer. Including justifications and modifying some of the access routines to the values maintained by the assumption-based TMS should allow for the appropriate characteristics.

8.3 Final Remarks

The work presented in this thesis took the existing AI techniques of knowledge based systems, modeling systems with qualitative equations and assumption-based truth maintenance systems, and applied them to the domain of gait analysis. The resulting system is a contribution to the field of gait analysis because it uses a qualitative model of gait which makes it more powerful than the previous gait analysis programs. The resulting system is also a contribution to the field of Artificial Intelligence in Medicine because it demonstrates that a qualitative model that uses assumptions to account for incomplete data is a feasible way to perform data analysis in the medical domain.

I.

Table I-1: Muscle Data for Dr. Gait

Muscle Name	When Usually On	Muscle Action
adductor	WA, WR	adducts thigh
anterior tibialis	WA, swing	Dorsiflexes and inverts foot
gastroc/soleus	SLS	Plantarflexes ankle, flexes knee
gluteus maximus	WA	extends and abducts thigh
gluteus medius	WA, SLS	abducts thigh
hamstring	WA, second-half swing	extends hip, flexes knee
iliopsoas	WR, first-half swing	flexes hip
peroneals	SLS	plantarflexes ankle
posterior tibialis	SLS	plantarflexes ankle
quadriceps	WA	extends knee and flexes hip

References

[Aikens 83] Aikens, J., Kunz, J., Shortliffe, E. and Fallat, R.
PUFF: An Expert System for Interpretation of Pulmonary Function Data.
Computers in Biomedical Research 16:199-208, 1983.

[Clancey 81] Clancey, M. and Letsinger,R.
NEOMYCIN: Reconfigurating a Rule-Based System for Application to Teaching.
In *Proceedings of the Seventh International Conference on Artificial Intelligence*, pages 829-836. Joint Conference on Artificial Intelligence, 1981.
In volume 2.

[de Kleer 79] de Kleer, Johan.
The Origin and Resolution of Ambiguities in Causal Arguments.
In *Proceedings of the Sixth International Conference on Artificial Intelligence*, pages 197-203. Joint Conference on Artificial Intelligence, 1979.

[de Kleer 84] de Kleer, Johan.
Choices without Backtracking.
In *Proeedings of the National Conference on Artificial Intelligence*, pages 79-85. 1984.

[Dzierzanowski 84]
Dzierzanowski, J.
Artificial Intelligence Methods in Human Locomotor Electromyography.
PhD thesis, Vanderbilt University, 1984.

[Fagan 80] Fagan, L.
VM: Representing time-dependent relations in a medical setting.
PhD thesis, Stanford University, 1980.

[Inman 81] Inman, V., Ralston, H., Todd, F.
Human Walking.
Williams and Wilkins, Baltimore, MD, 1981.

[Lesser 77] Lesser, V. and Erman, L.
A Retrospective View of the Hearsay-II Architecture.
In *Proceedings of the Fifth International Conference on Artificial Intelligence*, pages 790-800. Joint Conference on Artificial Intelligence, 1977.
In volume 2.

[Miller 82] Miller, R., Pople, H. and Myers, J.
Internist-1, An Experimental Computer-Based Diagnostic Consultant for General Internal Medicine.
New England Journal of Medicine 307(8):468-476, August, 1982.

[Patil 81] Patil, R.S.
Causal Representation of Patient Illness for Electrolyte and Acid-Base Diagnosis.
PhD thesis, Massachusetts Institute of Technology, October, 1981.

[Pauker 76] Pauker, S. G., Gorry, A., Kassier, J. and Schwartz, W.
Towards the Simulation of Clinical Cognition: Taking the Present Illness by Computer.
American Journal of Medicine 60:981-996, 1976.

[Rajagopalan 84] Rajagopalan, Raman.
Qualitative Modeling in the Turbojet Engine Domain.
In *Proeedings of the National Conference on Artificial Intelligence*, pages 283-287. 1984.

[Sandell 84] Sandell, H.
GENIE User's Guide and Reference Manual.
Technical Report 84-003, Electrical and Biomedical Engineering Vanderbilt University, July, 1984.

[Shortliffe 76] Shortliffe, E.
Computer-Based Medical Decision Making: MYCIN.
American Elsevier, New York, 1976.

[Simon 78] Simon, S., Mann, R., Hagy, J., Larsen, L.
Role of the Posterir Calf Muscles in Normal Gait.
The Journal of Bone and Joint Surgery 60-A(4):465-472, June, 1978.

[Tracy 79] Tracy, K., Montague, E., Gabriel, R., Kent, B.
Computer-Assisted Diagnosis of Orthopedic Gait Disorders.
Physical Therapy 59(3):268-277, March, 1979.

[vanMelle 80] van Melle, W.
A domain-independent system that aids in contructing knowledg-based consultation programs.
PhD thesis, Stanford University, 1980.

[Weiss 78] Weiss, S., Kulikowski, C., Amarel, S. and Safer,A.
A Model-Based Method for Computer-Aided Medical Decision Making.
Artificial Intelligence 11:145-172, 1978.

[Winarski 74] Winarski, D.
A Computer Simulation of the Human Gait Cycle for Medical Applications

description is *declarative* in nature: it describes the relationships among the flows. An implemented data processing system is *procedural* in nature: it must describe in detail how the flows are computed. The flow equations must be reinterpreted as basic computation steps (with an output flow and one or more flows as inputs) and constraints on the order in which these computations can be performed (the computation producing a flow must be performed before any computations using that flow) must be made explicit.

Design:[4]

The implementation will make use of files of data to be processed by job steps which will in turn create other files. Each file will contain the information represented by one or more flows; each job step will perform the processing to satisfy one or more flow equations. The design of each file (information contained, organization, storage device, record sort order) and of each job step (equations implemented, loop structure, accessing methods used) should be made in such a way as to minimize some overall cost measure (e.g. dollars-and-cents cost, time used, number of secondary storage I/O events) for the execution of the data processing system. Typically this requires dynamic (behavioral) analysis of tentative design configurations.

Code Generation:

The system's design must be coded in a supported high-level language so that it can be executed.

1.4 Data Driven Loops

Each flow equation represents a *computation* whose implementation is essentially iterative in

[4] In ProtoSystem I the design process is performed by the Optimizing Designer module.

nature. That is, the realization of the relation described will assist iteration over the records of the flow(s) in the flow expression on the right-hand side, performing the indicated operations in order to generate the records of the flow on the left-hand side. In traditional high-level languages this iteration is effected through the use of a loop mechanism.

A single loop (possibly with nested loops) can be used to implement a flow equation. That is, a loop can be devised that will produce the entire flow appearing on the left-hand side of the flow equation from the flows appearing on the right-hand side. In functional terms, the flows on the right-hand side of the flow equation are called the inputs to the corresponding loop and the flow appearing on the left-hand side is called the output of the loop.

The body of the loop specifies (by cases) whether, for a given index value, a record of the output is to be generated and, if so, how. For example, consider the (fully explicit) HIBOL flow equation:

```
S  IS    A + B   IF A PRESENT AND B PRESENT
       ELSE  A      IF A PRESENT
```

The body of the corresponding loop will distinguish two cases and corresponding courses of action:

1. Records in both A and B are present for the current value of the loop index, in which case a corresponding record of the flow S is produced whose datum is the sum of those in the records of A and B.

2. Only A contains a record corresponding to the current value of the index, in which case a corresponding record of S is produced that is identical to A's record.

If neither of these cases obtains, no output record is produced.

Clearly, in a correct implementation the body of the loop must be performed for every index value for which records of the input(s) exist that will be used to produce an output record. We call

[5] APL is an exception and for this reason (among others) it is often considered a VHLL.

any set of values for a particular index an *index set* and we distinguish two special kinds of index sets:[6]

The set of index values for which a flow F contains a record is called the *index set* of F (denoted IS(F)).

The set of index values for which an input flow F_I contains a record that will be used in generating a record of the output flow F the *critical index set of* F_I *with respect to* F (denoted $CIS_F(F_I)$).

These two should not be confused. $CIS_F(F_I)$ for some flow F will often be a proper subset of $IS(F_I)$.[7]

The problem we face is that of finding some way of enumerating the critical index sets of each input so that loop can be properly driven.[8] It is generally impractical to use the set of all possible (legal) index values for which an input might have a record. For one thing this set may be unbounded. Even if it is finite and enumerable, it will often be much larger than the critical index set and thus grossly inefficient. In the DEMAND flow equation example given above, for instance, the critical index set of the input flow CURRENTORDER is likely to be orders of magnitude smaller than its maximum possible size (the case where every store has orders for every item).

A much more efficient way of enumerating a set of index values that is assured to cover the critical index sets of the inputs is to use the union of the index sets of the input flows. This will work because a record of the output can be produced *only* if there is some input flow in which that

---

[6] Unfortunately, this terminology is at variance with that used by Baron in his thesis [1]. Baron uses the term "critical index set" to mean what we call the "index set".

[7] On no account, of course, can it be other than a proper or improper subset of $IS(F_I)$.

[8] This statement is somewhat oversimplified, but it will suffice for now. A fully precise statement of the problem is given by the Fundamental Driving Constraint in Part IV.

record is present. Moreover, the union of these sets is easily enumerated simply by reading the input flows (which have to be read anyway). A loop that is thus driven by index values supplied by its inputs is said to be a *data driven loop* and the flows used to drive the loop are termed its *driving flows* (the set of flows that drive a loop are called its *driving flow set*). The structure and implementation of data driven loops is the subject of this paper.

It will not always be necessary to use all of inputs of a loop to cover their critical index sets. From the flow equation for S above it can be deduced that the index set of A alone is sufficient to cover $CIS_S(A)$ and $CIS_N(B)$. This is so because it can be shown that a record of B will be needed only if there is corresponding record in A. Therefore, the loop can be driven by A alone; that is, it is sufficient to perform the body of the loop only for those index values for which A has a record. In general, for the sake of efficiency it is of interest to find the minimum subset or subsets of the inputs that will be sufficient to drive the loop.

Part II: *Structure of Data Driven Loops*

Before a general treatment of data driven loops can be developed it is necessary to examine the structures of the loops encountered in the HIBOL system. We begin by presenting a taxonomy of computation types and their corresponding loop implementations.

II.1 Loop Terminology

Before discussing loop structures it is useful to establish some terminology. By the term *loop* we mean a control construct which somehow enumerates a set of values for a *loop-index* and which performs a fixed sequence of statements (its *body*), once for each value of the loop-index. A loop may contain one or more loops within its body. The inner loops are said to be *nested* within the outer (enclosing) loop and the structure as a whole is called a *nested loop structure*. Each enclosure defines a different *level* of the nested loop structure. The degenerate case of a nested loop structure, where there is no loop in the body of the outer loop, is called a *single-level loop*, since there is only one loop level.

A *totally nested loop* is a nested loop structure whose component loops are totally ordered under enclosure (i.e. for any two loops L_1 and L_2 either L_1 is inside L_2 or L_2 is inside L_1).

II.2 Kinds of Computations and Their Loops

Each run (computation, job step, program) in the implementation produced for a HIBOL description of a data processing system is essentially a loop that iterates over the records of its input files to generate records of its output file(s). The structure of this loop depends on the nature of the computation being performed. We will begin with computations that directly implement single HIBOL flow equations of various types. Then we will consider computations that implement more than one flow equation (aggregated computations) simultaneously.

II.2.1 Simple Computations

A simple computation computes a non-reduction flow expression involving a single flow. Thus, it takes one flow as input and produces one flow as output. An example of such a computation is described by the HIBOL flow equation:

```
PAY IS        HOURS * 3.00                IF HOURS PRESENT AND
                                              NOT HOURS > 40

        ELSE 120 + (HOURS - 40) * 4.5   IF HOURS PRESENT
```

This assumes that there is a flow HOURS indexed by (employee id), each of whose records contains as its datum the number of hours the indicated employee worked. Each employee is paid $3.00/hour, with time-and-a-half for overtime. PAY is a flow containing the gross pay for each employee who worked (i.e. had a corresponding record in the HOURS flow).

The loop implementing this kind of computation is of the simplest kind. It is a single level loop having a single input file which is the driver. On each iteration the next record of the input is read and used to provide both a datum and index value for the body, which calculates the datum of the corresponding output record, assembles that record from this datum and the current index value, and outputs it.

In the SEAL language (see Appendix I) this would look like:

```
for each (employee-id) from HOURS

   get HOURS(employee-id)

   PAY(employee-id) =

      if defined(HOURS(employee-id))
         and not(HOURS(employee-id) > 40)

         then HOURS(employee-id) * 3.0

         else if defined(HOURS(employee-id))

                  then 120.0 + (HOURS(employee-id) - 40) * 4.5

                  else undefined

   if defined(PAY(employee-id))
      then write PAY(employee-id)

end
```

The for-end construct represents the basic iteration over values of the index employee-id. It specifies that the values for the index are obtained from the HOURS flow. For each index value, the corresponding record of HOURS is read, the corresponding record of PAY is generated, and (if generation was successful) that record is written out. Notice that the PAY calculation is a direct translation from the HIBOL flow equation.

For reasons of exposition the loop implementation presented here is of the most general form. An actual implementation would incorporate various efficiency enhancing improvements.[9] Nevertheless, we shall continue to use such forms to show explicitly where I/O and testing occur *conceptually*.

[9] For instance, since the for has to read the next record of the driver to get the current index value, the get could be omitted. Furthermore, the defined tests in the PAY calculation could be omitted since they are testing the presence of record which must be present. Finally, in this computation, the check before output could also be omitted.

II.2.2 Matching Computations

A matching computation computes a non-reduction flow expression involving two or more flows. Thus it is similar to a simple computation, but instead of operating on a single record of a single input flow to produce an output record, it operates on a set of corresponding records, one from each input flow. Correspondence is established by common index values. The name "matching computations" derives from the necessity of matching up the records of the inputs by index values before they can be operated on.

Two sub-classes of matching computations can be distinguished depending on whether all of the inputs have indices with identical key-tuples or not.

II.2.2.1 Expressions Involving Flows with a Uniform Index

Consider the a pay calculation similar to that given above, but where employees are paid various hourly rates. Let RATE be a flow, indexed by (employee-id), each of whose records has as its datum the hourly pay rate for the employee indicated by its index value. The pay calculation then becomes

```
PAY IS   HOURS * RATE                    IF     HOURS PRESENT
                                            AND RATE PRESENT
                                            AND NOT HOURS > 40

      ELSE
        RATE * 40 +
        (HOURS - 40) * 1.5 * RATE        IF     HOURS PRESENT
                                            AND RATE PRESENT
```

HOURS and RATE have identical indices, each consisting of the single key "employee-id". The loop that implements such a computation has a single level.

Because a record of the output is generated only if there is a record in the HOURS file, that

file alone is sufficient to drive the loop. (Alternatively, by similar reasoning, the RATE file could be used to drive the loop.) This is the simplest case of a matching computation because only one input is needed to drive the loop. (The computation of the flow S above is also of this type.) On each iteration the next record of the HOURS file is read, the corresponding RATE record is fetched, and the computation of gross pay performed.

This loop is represented in the SEAL language thus:

```
for each (employee-id) from HOURS

   get HOURS(employee-id)

   get RATE(employee-id)

   PAY(employee-id) =

      if defined[HOURS(employee-id)]
         and defined[RATE(employee-id)]
         and not(HOURS(employee-id) > 40)

         then HOURS(employee-id) * RATE(employee-id)

         else if defined[HOURS(employee-id)]
                 and defined[RATE(employee-id)]

                 then RATE(employee-id) * 40 +
                      (HOURS(employee) - 40) * RATE(employee-id) * 1.5

                 else undefined

   if defined[PAY(employee-id)]
      then write PAY(employee-id)

end
```

Again, the defined checks on the driver, HOURS, are superfluous. But those on RATE are necessary (to determine whether the corresponding get was successful) and the defined check on PAY is necessary (so that a record is written if and only if a datum was generated).

Now consider the HIBOL flow equation for the DEMAND flow given above:

IF CURRENTORDER PRESENT
AND BACKORDER PRESENT
ELSE CURRENTORDER IF CURRENTORDER PRESENT
ELSE BACKORDER IF BACKORDER PRESENT

Again, CURRENTORDER and BACKORDER have identical indices, both consisting of the component keys item-id and store-id, so again the implementing loop has a single level. In this case both inputs are necessary to drive the loop.

When there is more than one driver for a computation some of the drivers will have records for a given index while the others do not.[10] If the drivers have their records sorted in the same order (say, alphabetically) by index values, the loop may be performed by making only one pass through the inputs in the following manner:[11]

0. Read the first record of each input.

1. Use the smallest index value among those that is the current one and perform the loop body, fetching other non-driver records if necessary.

2. Discard all driver records with the same value that was used, and read the next record of every driver whose record was discarded.

3. Repeat from 1.

If the only sorting constraint on the inputs is that the drivers be sorted in the same order (i.e. no constraints on the order of the records in the other inputs), the records of each non-driving input will have to be fetched by random access (if the organization of the file allows random access) or by search; but if any non-driving input is sorted in the same order as the drivers, its records can be fetched by *sequential* reading, which is generally more efficient.

---

[10] If this were not the case, [illegible]

[11] If they are not similarly sorted, more inefficient means must be used.

These details are implicit in the SEAL representation of the loop which is simply:

```
for each (item-id, store-id) from CURRENTORDER, BACKORDER
   get CURRENTORDER(item-id, store-id)
   get BACKORDER(item-id, store-id)
   DEMAND(item-id, store-id) = ...
   if defined[DEMAND(item-id, store-id)]
      then write DEMAND(item-id, store-id)
end
```

II.2.2.2 General Discussion of Expressions Involving Flows with Mixed Indices

The treatment of mixed-index flow expressions in this paper will be restricted to those that are legal in HIBOL. The restrictions that HIBOL imposes are made for good reasons. A brief discussion of the various conceivable types of mixed-index flow expressions is presented here in order to show the motivation behind these restrictions.

The various cases where the flows in a flow expression have mixed indices (i.e. their indices have different key-tuples) can be distinguished by the set interrelationships among the key-tuples.

Consider the case where flows have disjoint key-tuples (e.g. (w, x) and (y, z)). Correspondence among records of such flows is meaningless, so we do not allow them to appear in the same flow expression.

Now consider the more general case where there is intersection among index key-tuples, but *the union of their pair-wise intersections is not identical to their (simple) union.* In this case correspondence is always ambiguous. For example, consider the two flows: A with index (x, y) and B with index (y, z). Suppose that there are records in A for the particular index values (x_1, y_1) and

(x_2, y_1) and that there are records on B for index values (y_1, z_1), (y_1, z_2) and (y_1, z_3). Which of A's records correspond to which of B's records?[12]

For correspondence to be meaningful and unambiguous it must be the case that the union of the pair-wise intersections of the key-tuples of the indices involved is identical to their union. This is always the case when there exists an index among the flows involved whose key-tuple is a superset of all the key-tuples of the other flows.