NEW MODELS FOR EXPERT SYSTEM DESIGN

JAMES STUART AITKEN

Submitted for the degree of Doctor of Philosophy to the Department of Electronics and Electrical Engineering of Glasgow University.
March 1990

Acknowledgements

I thank Dr James Macleod for his help and advice on all aspects of my studies. My work has benefited from his constructive criticism.

The clinical study of lung sound would not have been possible without the cooperation of staff at Glasgow Royal Infirmary. Dr Kenneth Anderson was responsible for proposing the histamine challenge study and for finding suitable patients. My thanks are also due to Dr Roger Carter for his assistance with the histamine challenge study and for providing information on the interpretation of pulmonary function data. I have received much assistance from Rosemary McCusker on the operation of the computer system of the Department of Respiratory Medicine. Anne MacKinnon and Dugald Campbell also gave advice on computing problems.

The willingness of Dr Francis Moran to support the collaborative investigation of lung sound and to provide material assistance is greatly appreciated.

The assistance of Dr Macleod with the preparation of this thesis has been invaluable. The author is responsible for any remaining grammatical errors or spelling mistakes.

Table Of Contents

List Of Figures

List Of Tables

Publications

Anderson K. Aitken S. Macleod J.E.S. Moran F. Lung-Sound Transmission And Reliability Of Chest Signs. Lancet 1988 ii:228

Aitken S. Anderson K. Moran F. Macleod J.E.S. The Correlation Of Breath Sound Power Spectrum And FEV1. Proceedings XVI World Congress Of Diseases Of The Chest, Boston, November 1989 :258

Anderson K. Aitken S. Carter R. Macleod J.E.S. Moran F. Variation Of Breath Sound And Airway Calibre Induced By Histamine Challenge.
to appear in the American Review Of Respiratory Disease

Summary

This thesis presents new work on the analysis of human lung sound. Experimental studies investigated the relationship between the condition of the lungs and the power spectrum of lung sound detected at the chest wall. The conclusion drawn from two clinical studies was that the median frequency of the lung sound power spectrum increases with a decrease in airway calibre. The technique for the analysis of lung sound presented in this thesis is a non-invasive method which may be capable of assessing differences in airway calibre between different lobes of the lung.

An expert system for the analysis of lung sound data and pulmonary function data was designed. The expert knowledge was expressed in a belief logic, a system of logic which is more expressive than first order logic. New automated theorem proving methods were developed for the belief logic. The new methods were implemented to form the 'inference engine' of the expert system. The new expert system compared favourably with systems which perform a similar task. The use of belief logic allows introspective reasoning to be carried out. Plausible reasoning, a type of introspective reasoning which allows conclusions to be drawn when the database is incomplete, was proposed and tested. The author concludes that the use of a belief logic in expert system design has significant advantages over conventional approaches.

The experimental results of the lung sound research were incorporated into the expert system rule base: the medical and expert system research were complementary.

Chapter 1 Introduction

Introduction

This thesis is a contribution to the study of human lung sound. The authors work builds on previous studies of lung sound carried out by researchers in the Department of Electronics and Electrical Engineering of Glasgow University. In common with Urquhart [1] and Luk [2] the author examined the power spectrum of lung sound. Features derived from the power spectrum were correlated with flow and volume measurements. The aim was to determine the relationship between the power spectrum of lung sound and the physiological condition of the lungs and more specifically, with the condition of the airways.

The use of an expert system to analyse the lung sound and pulmonary function data was proposed. The expert knowledge required for the interpretation of this data is explicitly expressed as logical rules. The rules assess the degree to which the lung function of the patient in question is impaired. Expert systems are able to explain their reasoning by showing how a fact has been deduced. Conventional programming methods usually lack these features.

Any relationship between lung sound and airway condition found experimentally can be expressed as a set of logical rules and incorporated into the expert system rule base. Consequently the medical and expert systems research are complementary.

The range of concepts which can be represented and manipulated by an expert system is determined by the logical language in which the expert rules are expressed. Logics which include modal operators were studied, these logics being more expressive than first order logic. From these observations the author concluded that the implications of using a belief logic as the language for knowledge representation in an expert system should explored with the aim of improving expert system design ('belief' is a modal operator). In contrast with many approaches to expert system design [3], the use of belief logic provides a well understood semantic theory (or model theory) and the inference procedure is an explicit component of the model theory. The expert rules are viewed as the beliefs of an expert agent and because facts and rules are expressed as beliefs, introspective reasoning may be carried out.

Introspective reasoning requires an agent to examine what can be deduced from its beliefs and what cannot. The problem of making logical deductions when the knowledge about a problem is incomplete (known as the problem of reasoning with incomplete information) can be tackled by introspective reasoning.

The design of an expert system to interpret lung sound data which was based on an established technique would not represent an advance in expert system design. The use of belief logics in the design of an expert system represents the development of new models for expert system design.

A guide to this thesis.

The study of human lung sound is introduced in Chapter 2. This chapter begins with a review of literature which reports experimental results relating lung sound and clinical measurements in normal subjects. A detailed analysis of the published data is developed (Section 2.2). Experimental equipment for the detection and spectral analysis of lung sound was designed by the author (Section 2.3). This equipment was used in two clinical studies which are described in Sections 2.4 and 2.5. Some general conclusions on the possible clinical uses of lung sound are presented in the final section of Chapter 2 (Section 2.6).

Chapter 3 presents the theory and background material on which the author's work on expert systems is based. Chapter 3 begins with an introduction to formal systems of logic and to the resolution theorem proving method (Section 3.2). The various systems of modal logic are described, as are the automated theorem proving methods for these logics (Section 3.3). The notions of knowledge and belief are examined and several systems of logic which model belief are reviewed in Section 3.4. The 'deduction model of belief' is presented in greatest detail as the author's work is concerned with enhancing and applying this logic. The final section of Chapter 3 is a review of literature on the problem of reasoning with incomplete information (Section 3.5).

The authors original work on expert system design is presented in Chapter 4.

New resolution theorem proving methods for the deduction model of belief are presented in Section 4.2. The new resolution methods were implemented and employed as the inference engine of the 'Inspire' expert system. The design of the expert system and the method of refining a set of expert rules for the interpretation of pulmonary function data are presented. The performance of the expert system and the role of the logical language are discussed (Section 4.3). A new theory for the deduction of plausible beliefs when the database is incomplete is presented (Section 4.4) and evaluated in practice (Section 4.5). The problems of reasoning with uncertainty, plausible reasoning and the formulation of these problems in modal logic are discussed in Section 4.6. Some general conclusions on the use of belief logic and its applications are outlined in the final section of Chapter 4 (Section 4.7).

A general discussion of directions for future work on the study of human lung sound and the development of the Inspire expert system is presented in Chapter 5.

References

1. Urquhart R.B. Some New Techniques For Pattern Recognition And Lung Sound Analysis. PhD Thesis Glasgow University 1983

2. Luk A. Some New Results In Nearest Neighbour Classification. PhD Thesis Glasgow University 1987

3. Oxman R. Gero J.S. Using An Expert System For Design Diagnosis And Design Synthesis. Expert Systems 1987 Vol 4 No 1 :4–15

Chapter 2 Lung Sound

2.1 Introduction

The stethoscope was the first instrument capable of detecting lung sounds and from its use an impression of sound amplitude and pitch can be obtained. Sounds commonly associated with disease such as crackles and wheezing can be identified. Abnormal sounds can be used to diagnose the early stages of lung disorder. However, observations made using the stethoscope are sometimes difficult to quantify.

The development of sophisticated diagnostic techniques, including airflow and lung volume measurements, has superseded auscultation as the primary means of diagnosing lung disorder. Measurement of lung volume and the expiratory flow rates a patient can achieve are important in themselves. The results are quantifiable by standard procedures and may be compared to the predicted values obtained from studies of a normal population.

The stethoscope bell can be placed over any of the lobes of the lung to assess the regional ventilation of the right or left lung. In contrast the flow–volume measurements provide an assessment of the lungs as a whole. The potential exists for lung sound to be used to measure regional ventilation in the lung, complementing the flow–volume tests. To achieve this the relationship between lung sound as detected at the chest wall and the airflow, airway geometry and chest wall characteristics which generate and transmit the sound must be investigated.

This chapter begins by examining previous research relating lung sound to airflow and lung volume. In the first experiments sound was characterised by its amplitude or intensity only, but with the development of suitable technology the frequency spectrum of lung sound was later studied. The author has designed a computer based system for the logging and spectral analysis of lung sound. This equipment described and its use in two investigations are described in the later sections of this chapter. In the first study the airway calibre of asthmatic subjects was reduced by the administration of histamine and the effect on lung sound was observed. The second study assessed the effect on lung sound of a reduction in lung function caused by disease.

2.2 Literature Review And Analysis

2.2.1 Introduction

This review examines recent research in the field of breath sounds and lung sounds. Breath sounds are those sounds audible at the mouth during the breath cycle, while lung sound is always detected at the chest wall. The latter may also be referred to as breath sound if no extraneous lung sounds are present. The experimental results and the theories developed in a selection of papers to explain the amplitude and frequency spectrum of lung sound are discussed. This requires study of the differences in regional ventilation and the relationship with lung sound. There is considerable disagreement amongst researchers about the origin of lung sound, but the basic physiology of the lung and the mechanics of breathing are more established and are presented in the following section.

2.2.2 The Mechanics of Breathing

The main function of the lungs is the exchange of respiratory gases. There are four main components in this process: ventilation, gas transfer, pulmonary blood flow and blood gas transport [1]. The present study is concerned with factors affecting the ventilation of the lungs and their impact on inspiratory breath sounds.

In normal subjects, during quiet breathing, inspiration is produced by the action of the diaphragm and the intercostal muscles. Expiration is produced by the elastic recoil of the lung. The airway can be divided into upper and lower sections. The upper respiratory tract warms and humidifies the inspired air and the lower respiratory tract conducts air to and from the alveoli. The airway from the larynx to the bifurcation into left and right main bronchi is known as the trachea. The length of the trachea and the angle of the bifurcation are not fixed but variable [2].

The main bronchus on the left side divides into left upper lobe and left lower lobe bronchi. These supply the upper and lower segments of the left lung. Further subdivisions occur and while the walls of the airways contain cartlige they are known as bronchi. Thereafter the conducting airways are known as bronchioles. The bronchi divide between eight and thirteen times, the smallest having a diameter of approximately 1 mm. There are between three and four subdivisions of bronchioles before the terminal bronchiole is reached. The alveoli occur in the walls of these airways. At each branching of the airway the combined cross section of the branches is greater than that of the stem from which they arose.

The left lung is divided into two lobes, and the right lung into three lobes. Sketches of these divisions are shown in Figure 2.2.1. Each lobe is ventilated by a system of bronchi originating from the main bronchi. Pathological changes are often confined to a single lobe, indicating a degree of isolation between lobes [3]. Both lungs are surrounded by a fluid filled membrane called the pleura. The thoracic cavity is referred to as the chest.

It is necessary to examine the forces which act on the airways during the breath cycle. The mechanical properties of the lungs and the effect of gravity must be considered.

There are both static and dynamic forces which oppose inspiration. The lungs have elastic properties: their compliance can be defined as the increase in lung volume per unit increase in distending pressure. Compliance varies as a non-linear function of lung volume. The elastic properties of the lung as a whole are dependent on the compliance of the tissues, the structure of the lung and the surface tension of the liquid lining the alveoli. The surface tension of the air liquid interface accounts for approximately half of the elastic recoil.

The dynamic forces which oppose inspiration are flow dependent. For air to flow through the airways a pressure difference must exist between mouth and alveoli. As the cross sectional area of the airways decreases the greater the pressure difference must be to maintain the same flow rate hence the greater the resistance to flow. When flow is laminar the resistance to flow is constant but where airflow is turbulent resistance is increased. In the thorax and large upper airways flow is turbulent whereas in the peripheral airways flow is laminar [4]. It has been shown that airways with a diameter of less than 2 mm contribute little to airway resistance.

During quiet breathing, expiration is produced by the elastic recoil of the lung

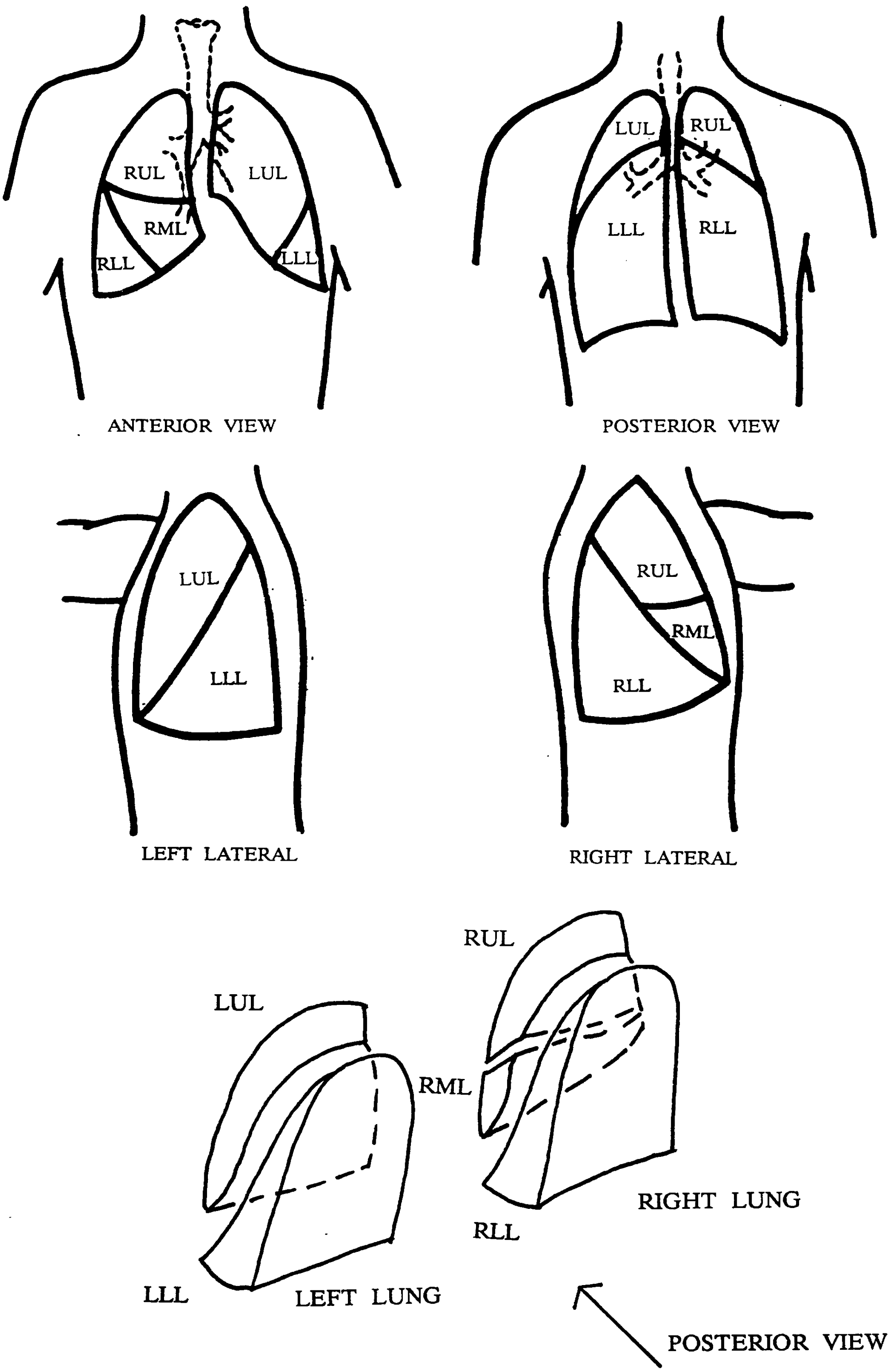


Figure 2.2.1. The respiratory system.

tissue. The recoil pressure of the expanded lung forces air through the respiratory passages. The lateral pressure on the walls of the airways plus the outward traction on the airway walls (due to the recoil pressure) is greater than the intra-thoracic pressure so that all airways remain open. During deep breathing the greater distension of the lung provides a greater elastic recoil. During forced expiration the lateral pressure falls due to the increased flow rate and the muscular effort required increases the intra-thoracic pressure. These effects can cause the dynamic compression of bronchi. Any increase in muscular effort and hence in intra-thoracic pressure tends to compress the peripheral airways further and does not increase the flow rate. The maximum flow rate is then effort independent [1]. As a consequence the flow-volume measurements and FEV1 (the volume of air which can be expelled in 1 second) are also effort independent.

FEV1 is reduced when airway calibre is reduced. If the recoil pressure is reduced, as in emphysema, dynamic compression will be more pronounced and FEV1 will be reduced.

Regional differences in ventilation within the lung can be caused by local mechanical factors. There is a vertical gradient in recoil pressure due to gravity. This pressure is greater at the apex than at the base in the upright position. Superior and inferior regions are the upper and lower regions with respect to the vertical axis. These regions may not correspond directly to segments of the lung.

2.2.3 The Distribution of Pulmonary Ventilation.

It is often assumed that lung sound intensity is a measure of ventilation of the underlying lung. There was no experimental research to support this view until the 1970's. Examination of regional differences in lung sound is important as it provides evidence about the processes of sound generation and transmission in the lung and their relationship to lung physiology. Five papers which present data from normal subjects, by well known authors, are reviewed.

In a paper on the relation between lung sounds and the distribution of pulmonary ventilation, Leblanc found that the maximum intensity of breath sound during inspiration was always found at low lung volumes over the superior regions of the lung [5]. The sound amplitude signal (after being rectified, filtered and integrated), airflow and volume were recorded simultaneously. Sound amplitude (Leblanc used the term 'intensity') was shown to increase with flow rate, and to vary with lung volume. Amplitude was plotted against lung volume for eight subjects inspiring from residual volume at a constant flow rate, in the upright position. Amplitude over the apex decreased between 10 and 70% of vital capacity (VC), while amplitude over the base increased from 10 to 40% VC and decreased thereafter, in all cases. Similar results were obtained when the microphone was positioned on the left apex and recordings were made with the subjects in the right lateral position and also in the left lateral position. In the right lateral position the microphone was over a superior part of the lung and the sound amplitude decreased as lung volume increased from 10 to 70% of VC. In the left lateral position the microphone was over an inferior region of the lung. In this case sound amplitude increased between 10 and 40% of VC and decreased thereafter. These changes were not due to changes in pitch of breath sound with lung volume.

Leblanc concluded that the amplitude of breath sounds is not uniform throughout the lung but varies with lung volume, flow, body position and site of recording. Studies by other researchers in the 1960's had used radioactive Xenon to show that ventilation was uneven and gravity dependent. This was considered to be due to the gradient of pleural pressure. Leblanc argued that regional amplitude of breath sound varies in the same way as the regional distribution of ventilation, but did not show this quantitatively.

If the lung inflated uniformly then the relation between lung volume and sound amplitude would be the same for all regions of the lung. Leblanc's study shows there are regional differences. These are differences in the phase of the breath cycle at which the ventilation of a segment reaches a peak. In this study lung volume refers to the total volume of air inspired at a particular instant. Lung sound at a particular location on the chest wall is not a function of volume directly. The volume, as the percentage of vital capacity, is a measure of the phase of the respiratory cycle. As the lung expands from residual volume to vital capacity the inspiratory airflow goes from zero through a peak value and back to zero. The peak flow into the superior segment occurs early in the breath cycle (VC< 10%), the peak flow into the inferior segment occurs later (VC= 40%). Leblanc considered lung sound to generated peripherally. If sound is generated centrally then there must be regional components to generation and transmission.

A paper was published by Ploysongsang (1977) which attempted to show the relationship between breath sounds and regional ventilation in a quantitative way [6]. Breath sounds were measured by using the sound detected at the anterior apex of the right lung as a reference. The variation of breath sounds between apex and base was measured using an oscilloscope with the x and y channels connected to the reference and measurement signals respectively (the microphone outputs were rectified, filtered and integrated). The oscilloscope trace was approximately linear during inspiration, a phase difference between the two signals was observed. The slope of the trace was independent of chest volume, airflow was maintained between 1 and 2 l/s. The differences in transmission paths between reference and measurement sites was deduced by introducing white noise at the mouth and measuring the slope of the oscilloscope trace as before. From these gradients the breath sound index (Ib) and transmission index (Tn) were calculated. By dividing Ib by Tn the compensated breath sound index was obtained. This was intended to be a measure of the actual sound generated. The recording procedure was carried out at four sites on the chest wall for 15 normal subjects in upright and supine positions. For ten of these subjects ventilation at the four sites was calculated using a method whereby radioactive Xenon was inhaled. Breath sound indices were correlated with the ventilation at each site (relative to the reference site). The correlation between compensated breath sounds and regional ventilation was significant in upright and supine positions (r= 0.54, r= 0.41). The correlation between uncompensated breath sound and ventilation was significant in the upright position only (r= 0.56).

Ploysongsang concluded that uncompensated breath sounds cannot be used confidently in both upright and supine positions to assess regional ventilation. He argued that sound was generated peripherally, in airways of less than 3mm in diameter. Ploysongsang's study confirms that ventilation of the lung is not uniform and that lung sounds reflect this. The calculations of the compensated breath sound index confirm that sound is generated regionally and that the intensity of this component of breath sound varies between regions. The study does not show how much sound is generated centrally and hence detectable at both reference and measurement locations. Not accounting for centrally generated sound leads to an under estimation of the breath sound indices. Ploysongsang comments that the use of a white noise source to calculate the transmision index has several limitations. Changes in transmission path length from mouth to chest wall, between apex and base, and the possibility that sound is conducted through body tissue as well as through the airways are sources of error. If sound is generated over a series of airways and not at a point source then this method may not compensate correctly.

Measurements were made on the right anterior chest wall, so that the base of the lung was an inferior region when the patient was upright. When the patient moves to the supine position a section of the right lower lung (RLL) becomes an inferior region, but the RLL extends to the anterior base, a superior region. The data presented by Ploysongsang shows that breath sounds increase less from apex to base when the subject moves to the supine position. This may be explained by noting that the RLL is not simply an inferior segment in the supine position (due to the gradient in pleural pressure) and this may affect airway geometry and sound generation.

The question of whether breath sounds can be correctly compensated by using a transmission index was studied by Kraman (1983). The method used was similar to that of Ploysongsang [7]. Breath sounds were quantified by compensating the mean sound intensity calculated over a 25 ms period by the instantaneous flow rate at the mouth. Twelve of these values were taken during inspiration and averaged to obtain a single value. Breath sounds were recorded at 2 cm intervals from apex to base on both left and right sides of the anterior and posterior thorax. To calculate the transmission index white noise was introduced at the mouth. This index was defined by the ratio of the mean sound pressure at the study site to that at the reference site. The index was calculated at FRC (Functional Residual Capacity, the volume of air which remains in the lung after a normal expiration) at the same locations that breath sound was recorded from. In this way maps of breath sound amplitude and transmission were built up. A computer was used to store and analyse the data.

Kraman comments that the right anterior sound and transmission patterns are similar to those reported by Ploysongsang. There was considerable variation in both measurements over small distances. The sound intensity detected at right and left posterior locations was similar, however the transmission index over the right lung was twice that over the left lung. If compensated sound intensity corresponds to ventilation then it must be concluded that ventilation of the left lung is twice that of the right lung. This is unlikely and casts doubt on the validity of compensating breath sounds by the above method. Kraman explained this finding as being due to the direct transmission of sound to the right lung as the trachea is in contact with the right mediastinal pleural surface, whereas direct sound transmission to the left lung is impeded.

In Kraman's study breath sound intensity was compensated for variations in flow rate and the transmission index was measured at a fixed lung volume. In contrast Ploysongsang calculated both indices throughout inspiration, possibly a better method. The correlation between uncompensated breath sound and ventilation remains valid.

Both Leblanc and Ploysongsang assumed that breath sound was generated in the small airways of the lung but provided no conclusive evidence for this. From theoretical calculations of the Reynolds number in the small airways it is very likely that the flow is laminar and silent. Others conclude that breath sound is generated in the upper airways by turbulent flow [4].

To try to discover the site of production of respiratory sounds a technique called subtraction phonopneumography was developed by Kraman (1980)[8]. This involved recording breath sounds simultaneously at two sites a known distance apart. Sound signals were recorded during inspiration and expiration for microphone distances between 0 and 10 cm laterally across the lower left lung. The aim was to find the site of sound generation by finding the degree to which reference and measurement signals were equal in frequency and phase characteristics. The cancellation index (SII) was defined as the ratio of the difference to the sum of the detected signals. A distant sound source would show the same phase/frequency relationship at two equidistant locations hence differences between the signals would tend to zero (as would SII), providing the amplitude spectra are also equal. The sum of the reference and measurement signals was displayed on an oscilloscope and defined by the peak to peak amplitude of the trace. The difference signal was similarly obtained by inverting the phase of the measurement signal before summing.

Measurements were made over the posterior left lower lung by incrementing the distance between microphones by 1 cm over the range 0 to 10 cm for five normal subjects. The cancellation index increased with increasing distance between microphones. This relationship was significant during both inspiration and expiration (using linear regression $r=0.82$, $r=0.56$). The rate of increase of SII with distance was greater during inspiration than during expiration. An increase in SII corresponds to a reduction in the degree to which the signals have common components. The curves flatten out at a cancellation index of 1.0. In different group of three subjects who had pleural friction rub the cancellation index was 1.00 with the microphones 2 cm apart.

Kraman concludes that the sounds produced during expiration are generated more centrally than those generated during inspiration. This claim is based on the evidence that the slope of the SII/distance curve is smaller during expiration. That is, while the common components of reference and measurement signals decrease with distance in both expiration and inspiration, the rate of decrease is less during expiration indicating a more central source of generation. With the microphones 2 cm apart the cancellation index was always less than 0.45 during inspiration in normal subjects. In comparison a value of 1.00 was obtained for sound generated at the pleura. This suggests that normal inspiratory sound is further from the chest wall than the pleura.

There are several limitations to subtraction phonopneumography. Kraman states that the variation in SII from breath to breath and the fact that the subtraction signal occasionally disappeared into baseline noise were problems. The method of estimating the peak to peak signal amplitude may also have been a source of error. Kraman used a balance control to equalise the amplitudes of the breath sound signals prior to addition and subtraction. This would not equalise the amplitude/frequency spectrum as would be required for the cancellation index to be a measure of phase difference only (no compensation for transmission differences was made either). SII may be overestimated as a result.

Kraman states that these results are consistent with an intralobar source for inspiratory sound and an upper airway source for expiratory sound. Kraman's study comfirms the theory that sound is not generated at the periphery of the lung. Centrally generated components of the breath sound were detected at the chest wall but these did not dominate regionally generated components. The contribution of regional components varied across the small distances studied. It is not possible to specify the generation(s) of bronchi where the centrally generated sound originates but it can be concluded that regionally generated sound must originate in later generation(s) of bronchi. The results of this study are consistent with the theory developed previously that breath sounds detected at the chest wall are not the result of centrally generated sound alone, lung sounds reflect ventilation.

It has been shown that breath sound amplitude varies with the phase of the respiratory cycle. Peak air flow, and therefore peak ventilation and breath sound occurs first in superior regions of the lung. Ploysongsang (1983) quantified the difference in phase between sound amplitude at the apex and the base for 19 normal subjects and 15 smokers [9]. The methods used were similar to those described earlier. The signals from reference and measurement microphones were displayed on the y and x channels of an oscilloscope screen, after being rectified and integrated. The microphones were located anteriorly at the apex and 15 cm below the apex, respectively. The phase difference between the signals was calculated from the oscilloscope trace.

Ploysongsang found that the mean phase angle for the non-smokers was 3.3° which was significantly lower than that of the smokers of 8.8°. The phase angle was shown to increase as the dynamic compliance decreased. This indicates that differences in phase of regional ventilation are due to dynamic compliance, a bulk property of the lung, as well to the gradient in pleural pressure.

The main findings of this section of the review can be summarised as follows. Sound intensity at the chest wall increases as airflow at the mouth increases. The actual flow rate into a region of the lung during inspiration is determined by pleural pressure, airway geometry and compliance. The gradient in pleural pressure ensures superior regions are ventilated ahead of inferior regions. As sound is generated by turbulent flow, breath sound amplitude correlates with regional ventilation. Lung sound is composed of centrally and regionally generated components. Sound is not generated from a point source or at the periphery of the lung. The connection between lung sound and lung physiology has been shown by the interpretation of experimental evidence.

2.2.4. Lung Volume, Airflow and Breath Sound.

Research into the relationship between the amplitude and frequency characteristics of breath sound, lung volume, airflow and FEV1 is reviewed.

In a paper published in 1971, Forgacs attempted to relate breath sound intensity to flow rate, both measured at the mouth [10]. The influence of bronchial calibre on FEV1 and inspiratory sound was also assessed. The subjects breathed through a pneumotachograph which had a microphone mounted 2 cm from the open end. The sound amplitude and flow rate signals were fed to the x and y channels of an oscilloscope. These signals were not rectified or integrated. The oscilloscope trace was recorded photographically during periods of steady flow. Peak sound amplitude increased linearly with flow rate in normal subjects. The ratio of the gradient to the average gradient for normal subjects defined a measure of 'intensity' (Is). Forgacs found Is correlated with FEV1 and peak expiratory flow rate in chronic bronchitis and in asthma (r values > 0.60).

Forgacs concludes that in chronic bronchitis inspiratory sounds are generated in intrathoracic airways. Breath sounds were silenced when helium was inhaled indicating that turbulent flow is the source (as turbulence is density dependent). The experiment demonstrates that sound intensity (Is) depends on the inspiratory calibre of the first two or three generations of bronchi, and Forgacs notes that there is no *a priori* reason why Is should show a good correlation with any of the usual tests of airway obstruction. The forced expiratory tests, which assess the degree of airway obstruction, depend on the flow rate and lung volume which cause the dynamic compression of the large airways. Forgacs' findings suggest that the narrowing of the central airways, detected by sound measurements does not stop at the central bronchi but also involve the more peripheral generations of airways as well i.e. those whose calibre is reduced during forced expiration.

The use of peak sound amplitude to calculate the gradient of the amplitude/flow curve takes no account of spurious peaks which may occur. No attempt was made to remove these by integration or numerical averaging. However in subjects with bronchitis Forgacs recorded gradients ten times greater than normal and such differences cannot be attributed to experimental error. There is no doubt that sound amplitude increases with flow rate, but it is difficult to assess whether this relationship is precisely linear due to the experimental method used.

The relationship between airflow, lung volume and mean sound intensity at the anterior chest wall was investigated by Leblanc as part of the study of regional ventilation [5] described in the previous section. Leblanc concludes that sound intensity increases linearly with flow rate when measured at a specific lung volume (at the same point on the respiratory cycle) during successive inspirations. When the flow rate is kept constant, lung sound amplitude varies throughout the respiratory cycle. In this experiment, all measurements were recorded throughout the breath cycle, and all data points were instantaneous values estimated at specific volumes or flow rates.

The relationship between total airflow at the mouth and breath sounds recorded at the chest was investigated by Shykoff and Ploysongsang in a paper published in 1988 [11]. Their study made extensive use of a computer based measurement system. Airflow at the mouth was measured using a pneumotachograph, and the output signal was filtered before being digitised. When the patient was in the upright position one microphone was fixed to the right anterior chest wall over the right upper lung. The second microphone was fixed on the right posterior chest wall over the right lower lung. The microphone output signals were filtered prior to sampling by a bandpass filter (bandwidth 100Hz–800Hz). All three signals were sampled at 2 kHz. Subjects attempted to maintain a specified constant flow rate for two seconds during inspiration and expiration. Data was recorded over a period of thirty seconds.

The average root mean squared (RMS) value of the sound signal over a 10 ms interval was calculated and correlated with the mean flow rate over the same period. Analysis showed sound amplitude to vary with the square of the flow rate. To confirm the exponential relationship, the slope of the regression of the logarithm of sound intensity and flow rate was calculated. In over 80% of the 54 subjects studied the slope of this curve did not differ significantly from 4. This appears to confirm that sound amplitude (the square root of sound intensity) is proportional to the square of flow rate.

Shykoff and Ploysongsang conclude that the amplitude of sound pressure during inspiration is highly correlated with the square of simultaneous flow. As kinetic energy is a function of the square of velocity, the proportionality may be between kinetic energy and sound amplitude.

Shykoff and Ploysongsang state that a quadratic curve could be fitted by eye to the data presented by Leblanc [5]. This claim cannot be substantiated for the published data. The conclusion that sound amplitude is proportional to the square of flow rate conflicts with the results of Leblanc and Forgacs [10] who found sound amplitude to be proportional to flow rate. The experimental method employed by Leblanc accounted for the fact that flow rate varies with the phase of the breath cycle. Leblanc measured sound amplitude at known points of the respiratory cycle. In contrast, Shykoff and Ploysongsang did not make any allowance for the variation of regional flow rate with respiratory phase (or lung volume). Consequently it is difficult to compare the results of these studies. It is possible that the quadratic relationship found by Shykoff and Ploysongsang could be approximated by a linear relationship over part of its range, in this way the conflict could be be resolved.

2.2.5 The Spectral Characteristics of Breath Sounds.

Three papers which study the spectral characteristics of normal breath sounds by the Fast Fourier Transform (FFT) method are reviewed. The first was published in 1981 by Noam Gavriely [12]. Breath sounds were recorded from four sites on the chest wall, over right and left lower lungs on the posterior chest wall and over the right upper lung on the anterior and posterior chest wall, and also from the throat. At each location breath sounds were recorded onto magnetic tape for a period of thirty seconds. The microphone used in this study had a linear gain/frequency response. A high pass filter was used to attenuate signals below 75Hz. The recorded signal was digitized at 4000 samples per second.

Each sequence of 4096 samples was windowed in the time domain, using a cosine function, prior to transformation into the frequency domain by the FFT. A single inspiratory spectrum was obtained by averaging all of the sequences which contained inspiratory sounds only (4–6 sequences). A single expiratory spectrum was similarly obtained.

Power was found to decline exponentially with frequency. The frequency beyond which 'zero power' was detected was defined as the maximal frequency (fmax). Each power spectrum was defined by the slope of the log. amplitude/ log. frequency curve (A) and fmax, when this curve was shown to be linear by regression (non-linear spectra were rejected).

Gavriely found no significant change in gradient between inspiration and expiration. The difference in gradients between right and left lower lungs was minimal. There was a significant difference between gradients calculated for the right anterior site and those for the lower posterior sites. The measured power of the inspiratory segments was found to be significantly greater than that of the expiratory segments. Breath sounds detected over the trachea were found to have an approximately flat power spectrum, extending from 75Hz to an average frequency of 920Hz.

Gavriely concludes that these results are consistent with sound generation at the larynx, trachea or bronchial tree. It was assumed that the tracheal breath sound represented the sounds produced by the generator and that the spectral characteristics of the sounds picked up over the chest wall represent the attenuation of this sound. As the spectral slopes of sounds picked up over a specific location did not vary during the respiratory cycle, this implies that the transfer function is independent of the phase of respiration but is characteristic of the pickup location with its underlying tissues. Sound intensity over the trachea did not vary in the same way as intensity at the chest wall. This suggests that sound is generated in the bronchial tree and this results in a greater input to the lung parenchyma. Gavriely states that fmax and A would be expected to alter under pathophysiological conditions and also that new sound generators may distort the exponential pattern.

There are a number of possible objections to Gavrielys' conclusions. It is now believed that the mechanisms responsible for inspiratory and expiratory lung sounds are different [13] which tends to undermine Gavrielys' analysis. Moreover if the broad band sound spectrum detected at the throat is taken to represent the sounds as generated then this spectrum appears to be independent of airflow and airway geometry, but there is no evidence for this independence in Gavrielys' study. The method of characterising breath sound is not applicable to abnormal breath sounds as they would not be purely exponential hence the test for linearity would fail and the gradient could not be calculated.

The effects of lung volume and airflow on the frequency spectrum of lung sound was investigated by Kraman (1986) [14]. The amplified signals from two microphones and the signal from a pneumotachograph were recorded onto magnetic tape. One microphone was located on the upper right anterior chest wall over the right upper lung, the second was located over the right lower lobe on the posterior chest wall. Two inspiratory manoeuvres were studied, the first required the subject to inhale at a constant flow rate from residual volume to total lung capacity (the variable volume manoeuvre). The second required a set peak inspiratory flow rate to be reached 1.0–2.3 l/s (the varaible airflow manoeuvre).

When the tape was replayed signals below 100Hz were attenuated. Sound spectra calculated by the FFT method from the variable volume recordings had a frequency resolution of 8Hz (frequency range 0–1000Hz). The resolution for the variable airflow spectra was 20Hz (frequency range 0–2500Hz). In each case the spectra were quantified by calculating the quartile frequencies $Q_1 - Q_3$. The median frequency (Q_2) defines the frequency below which 50% of the total energy lies. To analyse the variable volume data the quartile frequencies were found from spectra calculated over six time intervals during inspiration. The time intervals were determined by the points in time when the vital capacity had increased by one sixth. The variation of each quartile frequency with volume was analysed at posterior and anterior sites. There was a significant decrease in Q_1 with volume at the right anterior site ($r = -0.43$, $p < 1\%$), and no other significant changes were observed. The variable airflow data was analysed by calculating the quartile frequencies when peak airflow was detected. The increase in Q_3 with airflow was significant at one site ($r = -0.285$, $p < 1\%$), no other significant trends were observed. When the data for each subject was examined individually, significant changes in quartile frequencies were found but they followed no overall trend. The log. power/ airflow relationships were linear.

Kraman concludes that the frequency spectrum of lung sounds of healthy subjects is minimally affected by submaximal variations in airflow or lung volume.

By dividing the power spectrum into quartiles Kraman identified a method which would reflect changes in the distribution of sound energy. Methods which rely on detecting peak frequencies or changes in the amplitude of fixed frequency bands show these changes less well. No averaging of segments of transformed data was carried out so the variance of the estimated amplitude values was 100%. This may account for some of the large variation of quartile frequencies.

In both of these studies low frequency sounds were attenuated by filtering as these sounds were presumed to be muscle sounds or heart sounds. Kraman (1983) attempted to distinguish lung sounds from sounds produced by muscular contraction [15]. Four normal subjects were studied. Sound was initially recorded onto magnetic tape. On replaying, this signal was sampled at 2000Hz. Sound spectra were calculated by the FFT method with a frequency resolution of 8Hz. The power in four frequency bands centered at 50Hz,100Hz,200Hz and 300Hz was calculated throughout the breath cycle. The bands centered at 200Hz and 300Hz showed a marked increase in power during the inspiratory period. This increase was less apparent in the 100Hz band, and absent in the 50Hz band. These results suggest that much of the sound energy below 100Hz may be generated by muscular contraction but no quantified measurements were presented.

Studies of the frequency spectrum of lung sound are inconclusive as to the factors which determine the shape of the sound spectrum heard at the chest wall. The spectrum appears to be independent of airflow and lung volume. It seems probable that low frequency components are of muscular origin.

References

1. Campbell E.J.M. Dickenson C.J. Slater J.D.H. Clinical Physiology. Blackwell, London 1974 :105–7

2. Crofton J. Douglas A. Respiratory Disease. Blackwell Scientific Publication, London 1975

3. Murphy R.L.H. Holford S.K. Lung Sounds. American Thoracic Society publication 1980

4. Forgacs P. Lung Sounds. Bailliere Tindall, London 1978

5. Leblanc P. Macklem P.T. Ross W.R.D. Breath Sounds and Regional Ventilation. Am Rev Respir Dis 1970;102:10–16

6. Ploysongsang Y. Martin R.R. Breath Sounds And Regional Ventilation. Am Rev Respir Dis 1977; 116:187–199

7. Kraman S.S. Austrheim O. Comparison Of Lung Sound And Transmitted Sound Amplitude In Normal Men. Am Rev Respir Dis 1983; 128:451–454

8. Kraman S.S. Determination Of The Site Of Production Of Respiratory Sounds By Subtraction Phonopneumography. Am Rev Respir Dis 1980;122:303–309

9. Ploysongsang Y. The Lung Sounds Phase Angle Test For Detection Of Small Airways Disease. Resipiration Physiology 1983; 53:203–214

10. Forgacs P. Nathoo A.R. Breath Sounds.Thorax 1971;26:288–295.

11. Shykoff B.E. Ploysongsang Y. Chang H.K. Airflow and Normal Lung Sounds. Am Rev Respir Dis 1988; 137:872–876

12. Gavriely N. Palti Y. Alroy G. Spectral Characteristics Of Normal Breath Sounds. Physiological Society 1981 :307–314

13. Rice D.A. Transmission Of Lung Sounds. Seminars In Respiratory Medicine Vol 6 No 3 January 1985, Thieme–Stratton, New York

14. Kraman S.S. Effects Of Lung Volume And Airflow On The Frequency Spectrum Of Vesicular Lung Sounds. Respiration Physiology 1986; 66:1–9

15. Kraman S.S. Does The Vesicular Lung Sound Come Only From The Lungs ? Am Rev Respir Dis 1983;126:622:626

2.3 Logging And Spectral Analysis Of Lung Sound

2.3.1. Introduction.

The study of lung sound at Glasgow University was begun under the supervision of Dr J.E.S. Macleod by Mr J. McGee, Dr R.B. Urquhart and was continued by Dr A. Luk. Both Luk and Urquhart recorded lung sound from the chest wall and used fast Fourier transform methods to calculate the frequency spectrum of the sound. Sound was detected using a hand held microphone designed by McGhee and stored on magnetic tape [1,2].

With the advent of cheaper, more powerful personal computers it was possible for the author to design and build a more reliable logging system based on an Amstrad PC1512 computer. This computer was used for both data logging and analysis. The logging system and spectral analysis technique are now described.

2.3.2. Signal Logging.

For the sound signal to be recorded correctly the signal amplitude must be within the ±5V range of the 12 bit A/D converter and the signal bandwidth must be limited to avoid aliasing effects. From results obtained by Luk it was concluded that there is little information in the lung sound spectrum above 3kHz. On this assumption the system bandwidth was defined as 0Hz to 3kHz. To retain compatibility with Luk's work a sampling frequency of 9.6kHz was used. The Nyquist frequency was 4.8kHz. A low pass filter is required to attenuate signals above the Nyquist frequency in order to prevent aliasing. An attenuation of 20dB at the Nyquist frequency relative to the pass band (cuttoff frequency 3kHz) was judged to be adequate. This defines the specifications of the low pass filter.

These specifications can be met by a 4th order Chebychev transfer function (ripple = 0.5dB). This prototype was chosen because it has a faster cutoff rate than a Butterworth filter of the same order. The predicted attenuation at 4.8kHz is 25.2dB. The pole positions were found from standard tables [3]. A fourth order low pass filter can be realised from two second order low pass sections in series. The circuit diagram is shown in Figure 2.3.1. The measured attenuation at the Nyquist

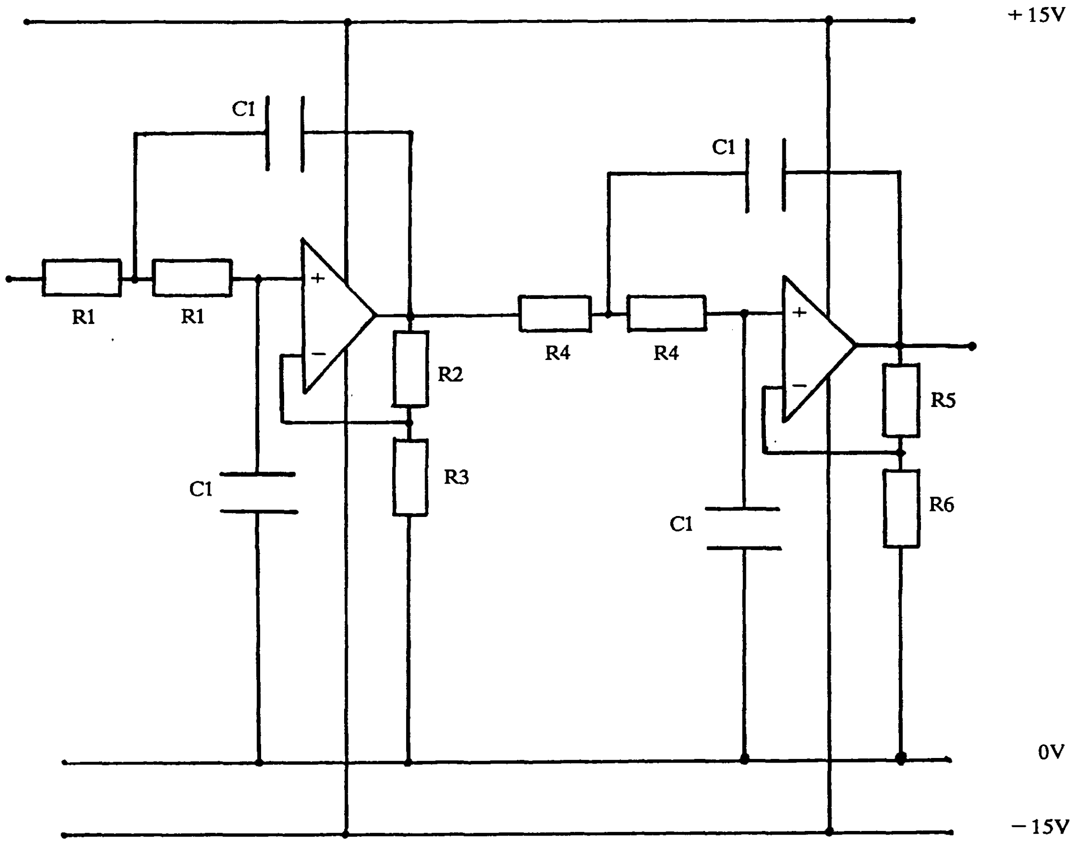


Figure 2.3.1. Circuit diagram of 4th order filter.

Component values: R1= 51.44KΩ R2= 10KΩ

R3= 16.6KΩ R4= 86.86KΩ R5= 10KΩ R6= 5.82KΩ

C1= 1nF

frequency was 22dB.

The signal conditioning unit houses the filter and three amplifier circuits as is shown by the block diagram in Figure 2.3.2. The input amplifier is a.c. coupled and has a high input impedance. The output amplifier has a preset gain of one, alternatively a potentiometer may be switched into the feedback loop of this amplifier to allow the gain to be increased or reduced to allow for loud or quiet lung sounds. An amplifier is also required to drive a pair of headphones to allow the operator to listen to the sounds.

In addition to the signal conditioning unit the logging system includes a voltage converter unit which provides a logic "1" TTL level output signal when the hand held switch is pressed. This signal is monitored by the computer, and the sound signal can be monitored on the oscilloscope. The microphone is the same hand held model used by previous researchers. The logging system is shown in Figure 2.3.3.

A logging program was written in Turbo Pascal to monitor the TTL channel and to take 16K samples of the sound signal when this channel goes to the high state. Data is sampled at 9.6kHz over 1.7 seconds. A more detailed description of the operation of the A/D board can be found in an internal report by the author [4].

2.3.3. Spectral Analysis.

The methods used to calculate the power spectrum were developed from the work of Luk. One window function was selected from the four types investigated by Luk and used for all data analysis. In Luk's method the segments of data were overlapped to reduce the variance of the estimate of spectral power, this was discontinued to reduce the number of Fourier transform calculations required.

The fast Fourier transform (FFT) algorithm of Cooley and Tukey requires $N\log_2 N$ multiplications, a great improvement on the N^2 multiplications which the calculation was previously believed to require (where N is the length of the transform). Their algorithm relies on the fact that a discrete Fourier transform of length N may be rewritten as the sum of two transforms of length N/2. This

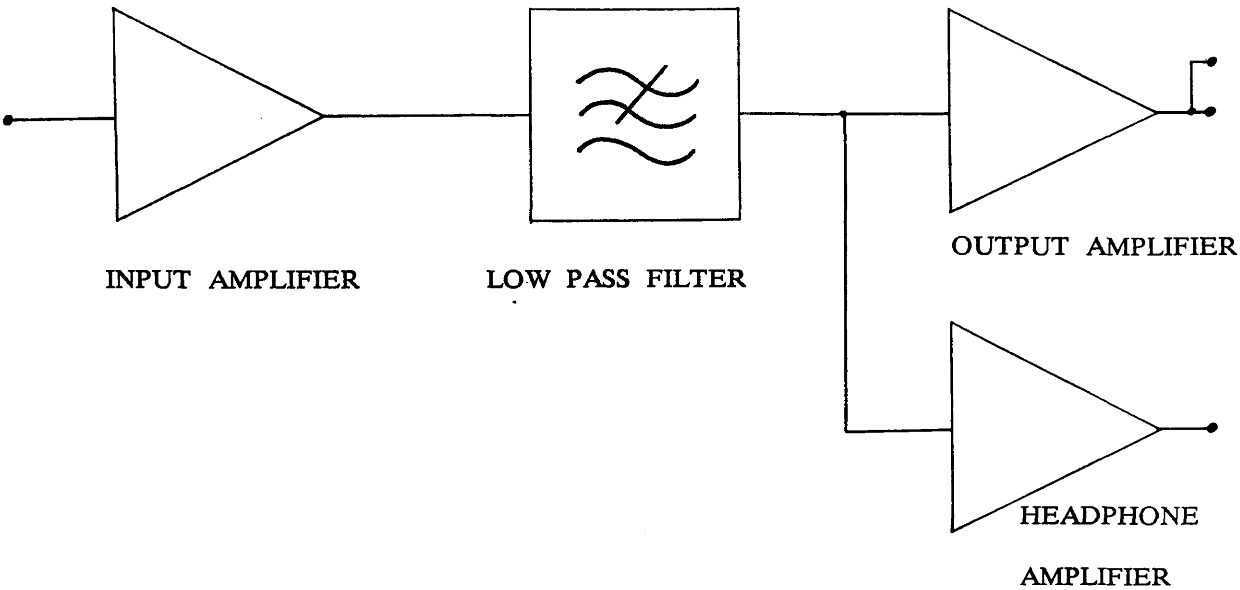


Figure 2.3.2. Block diagram of the signal conditioning unit.

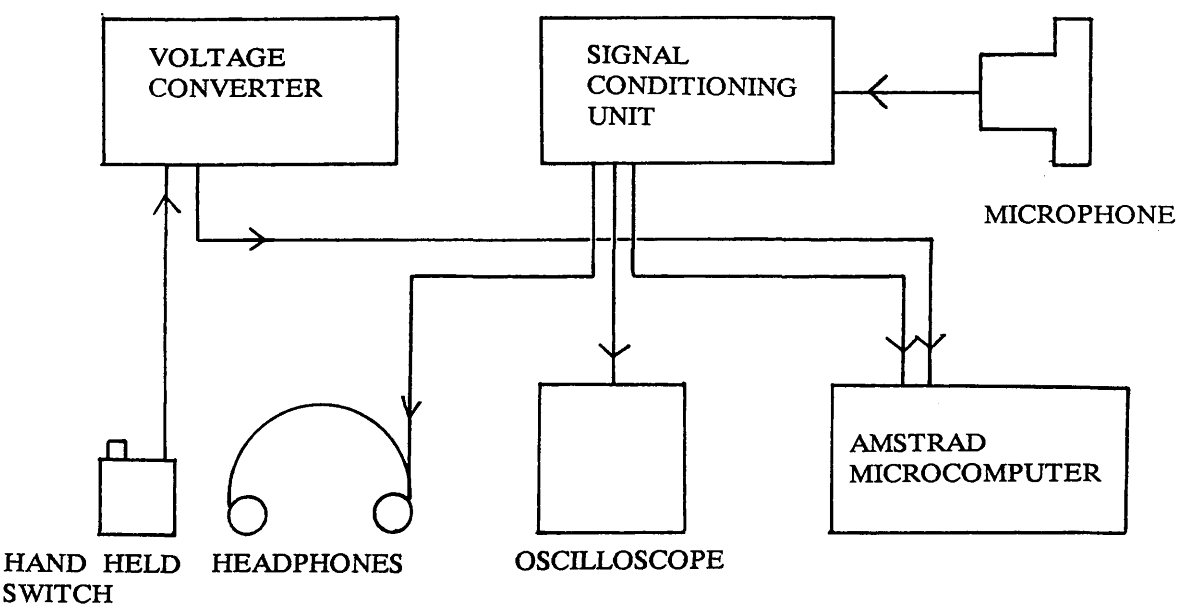


Figure 2.3.3. The logging system.

procedure is applied recursively until the transform is of length 1. The FFT routine employed by the spectral analysis program uses this method.

The power spectral density was defined for 0 and discrete positive frequencies as follows. Letting T denote the time interval between samples and N the number of samples the Fourier transform is estimated for the discrete frequencies

$$f_n = n/NT, \quad n = -N/2, \ldots, N/2$$

If $H(f_n)$ is the discrete transform, then the power spectral density P_H is defined by the equation:

$$P_H(f_n) = 1/N^2 \left(|H(f_n)|^2 + |H(-f_n)|^2 \right)$$

A more detailed explanation of the discrete FFT can be found in the references [4,5,6].

Each estimate of power at a discrete frequency represents the power over a range of continuous frequencies. It is then an average of continuous power over a narrow window centered at that frequency. It can be shown that this window is wide enough for leakage to occur between discrete frequency intervals. This effect can be reduced by a technique known as windowing.

The discrete time signal transformed can be regarded as the result of multiplying an infinite number of samples by a window function in time. In the techniques first presented, this function is a 'rectangular window', having the value 1 during the sampling period and 0 elsewhere. The spectrum which results is the convolution of the transform of the data and the transform of the window function. Leakage can then be shown to be the result of the rapid change in the rectangular sampling window from 0 to 1. Leakage can be reduced by using a window which changes more gradually. Many such windows have been defined, one being the 4 term Blackman Harris window chosen for this application.

The spectral analysis of lung sound begins by Blackman–Harris windowing a sequence of 2048 samples. The power spectrum is then calculated by squaring the amplitude values output by the FFT routine. This is repeated for four consecutive sequences of samples. The power at each discrete frequency is summed to give an estimate of the power spectrum over the whole time interval. This reduces the variance of the estimate, by a factor of four if the signal is considered to be stationary. The last step is the use of the Daniel window to smooth out any spurious peaks in the spectrum. This is achieved by weighting the estimate of spectral power at one frequency by its immediate neighbours. The frequency resolution is 4.69Hz.

The FFT algorithm was obtained from a library of numerical routines [5]. The windowing algorithms were translated into Pascal from Fortran programs written by Luk. Graphics routines were written to display plots of the spectra.

The microphone design incorporates a Tufnol diaphragm to improve the acoustic match between the microphone element and the chest wall. This has the effect of increasing the gain of the microphone at frequencies above 250Hz [4]. Consequently the power spectrum of the microphone output signal is that of the sound source, modified by the microphone's characteristics. The energy in lung sound is small and it is an advantage to be able to detect sound above 250Hz. The sound spectra were characterised by their distribution of energy which must reflect the distribution of energy of the sound source, changes in this distribution from breath to breath and from subject to subject can then be compared.

The programs written by the author enable the Amstrad to function as a lung sound spectrum analyser, data being logged in real time and the power spectrum calculated after logging is complete. Data is stored in its original form as a time series, and the power spectrum is also stored and is available for further analysis.

References

1. Urquhart R.B. Some New Techniques For Pattern Recognition And Lung Sound Analysis. PhD Thesis Glasgow University 1983

2. Luk A. Some New Results In Nearest Neighbour Classification. PhD Thesis Glasgow University 1987

3. Kuo F.F. Network Analysis And Synthesis. John Wiley, New York 1962

4. Aitken J.S. The Logging And Analysis Of Lung Sound. Internal Report 1988

5. Press W.H. Numerical Recipes. Cambridge University Press 1985

6. Proakis J.G. Manolakis D.G. Introduction To Digital Signal Processing. Macmillan, New York 1988

2.4. Histamine Challenge Study

2.4.1. Summary

Inspiratory breath sounds were recorded from the chest wall during histamine challenge in 5 subjects with mild asthma (baseline FEV1 > 60% of predicted normal). The median frequency of the power spectrum of the breath sounds was found to correlate with the percentage change in FEV1 induced by histamine and with FEF50. The analysis suggests that for a decrease in FEV1 of 20% the median frequency of breath sound would increase by 80Hz. Variation in airway calibre produced a consistent alteration in the distribution of energy in inspiratory breath sound, in the absence of wheeze. Spectral analysis of breath sound may be a useful addition to conventional spirometry in identifying changes in airway diameter.

Note: This section is based on paper titled 'Variation Of Breath Sound And Airway Calibre' accepted for publication in the American Review Of Respiratory Disease. The co-authors were K.Anderson R.Carter F.Moran (Glasgow Royal Infirmary) and J.E.S.Macleod (Glasgow University).

2.4.2. Introduction

Some of the first investigations into breath sounds used a measure of loudness (average amplitude or power) of the sound signal to quantify the sound spectrum. Later work has been concerned with the analysis of the power spectrum of the breath sound [Section 2.2]. Gavriely found that the breath sound amplitude(dB)/frequency curve was linear for normal subjects [1]. However this relation did not hold for asthmatic subjects [2]. Kraman characterised the frequency spectrum by dividing it into quartiles [3]. Using this technique, a change in the distribution of energy within the spectrum will be reflected by a change in the quartile frequencies. The assumption that there is a linear relationship between amplitude(dB) and frequency is not made and consequently this method of characterisation is applicable to normal and abnormal cases alike.

The present study assesses whether the median frequency can be related to changes induced in the FEV1 and flow volume curve in mild asthmatic subjects during a histamine challenge, and hence to physical changes in the lung. Breath sound was recorded during inspiration because of the unpredictable and premature airway closure which may occur during expiration [4], hence inspiratory recording can produce measurements which are more repeatable.

2.4.3. Method

The patient was seated throughout the study period in a position between the breath sound recording equipment and the lung function measurement equipment. Four recordings were made of breath sound during inspiration from the lower right posterior chest wall 3–5 cm below the inferior angle of the scapula. The patient was then asked to exhale into a spirometer (Compact Spirometer, Vitalograph, Kansas 66215 USA) and an expiratory flow volume curve and FEV1 was obtained. These provided baseline measurements before the first dose of histamine was given, this dose was followed by measurement of lung function and the recording of two inspiratory breath sounds. This procedure was continued in a stepwise fashion throughout the histamine challenge. Collecting data in this way enabled comparison of breath sound and lung function data throughout each histamine challenge.

The histamine challenge was performed using a modification [5] of Cockrofts method [6]. Patients were instructed to breath slowly and deeply, each dose of histamine consisted of 10 deliveries by the dosimeter (volume in each dose = 0.095 (SD 0.002)ml). The pulmonary function measurements were recorded 30 and 90 seconds after each dose of histamine. The initial concentration of histamine used was 0.03g/l after saline, which served as the control, and this concentration was doubled every three minutes to a maximum of 16g/l or until there was a fall in FEV1 of more than 20%. Data were obtained from 5 mildly asthmatic subjects. Each lung function value is given with an estimate of the predicted normal appropriate for the method of measurement [7–9]. All subjects were non–smokers.

The equipment for the logging and analysis of lung sound has been described previously [Section 2.3]. The power spectrum of each breath sound was calculated using the fast Fourier transform and was defined at 1024 discrete frequency points. The frequency resolution was 4.69Hz. A method was required to reduce the number of points characterising a spectrum down to one or two. This was developed by first calculating the total power in the range 100Hz to 1500Hz. The median power frequency was then obtained by finding the frequency below which 50% of the total energy lies, within the chosen range. This defines F50 for a breath sound spectrum. F85 was defined in a similar way (85% of the total energy lies below F85). These numbers then characterise the shape of each spectrum in terms of the distribution of energy, within the frequency range corresponding to that typical of breath sounds.

2.4.4. Results

Lung function at rest is shown in Table 2.4.1. versus age and gender. The differences in the power spectrum between baseline and last dose are shown in Figure 2.4.1. for two patients. This indicates that there is an increased amount of energy at frequencies above about 400Hz in the breath sound after the final dose. No wheeze was heard on auscultation or seen in the power spectrum of any patient at any stage during the histamine challenge.

Patient	Age yrs	Gender	FVC l	FEV1 l/s	FEV1 /FVC %	FEF50 l/s	KCO mmol.min.$^{-1}$ KPA.$^{-1}$l.$^{-1}$
1	22	M	4.63(79)	2.75(63)	60	1.95(35)	1.7(81)
2	31	M	4.97(89)	3.56(81)	72	3.49(63)	1.9(92)
3	43	M	4.32(85)	3.35(86)	76	3.23(63)	2.0(99)
4	24	F	3.16(86)	2.49(68)	78	2.66(56)	1.7(79)
5	37	M	6.02(104)	3.86(94)	64	3.09(70)	1.9(75)

Table 2.4.1. Baseline pulmonary function measurements for each subject, with age and gender (M = male, F = female). The values in parenthesis are a percentage of the predicted normal for each measurement.

Parameters y,	x	Correlation r	v:P< v
F50,	FEV1	−0.01	1
F50,	FEF50	−0.59	0.01
FEF50,	FEV1	0.49	0.05
F50,	%ΔFEV1	−0.70	0.01
F50,	%ΔFEF50	−0.70	0.01

Table 2.4.2. Correlation of parameters.
y is the dependent variable.

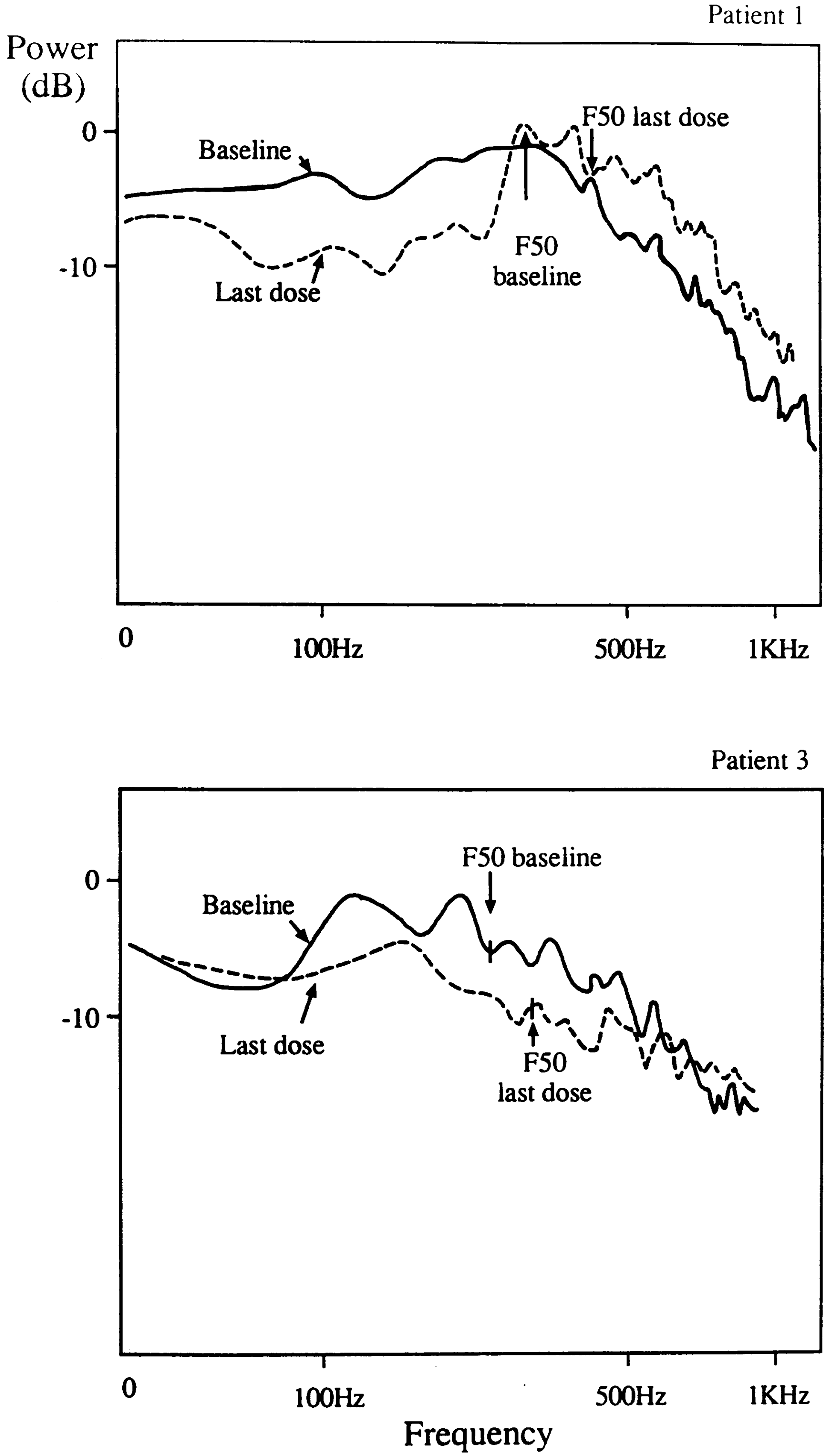


Figure 2.4.1. Baseline and last dose lung sound spectra.

The changes in FEV1 and F50 for Patient 2 during the challenge are shown in Figure 2.4.2. The results show that as the dose of histamine is increased FEV1 falls, as would be expected, and F50 increases. The relationship between F50 and the forced expiratory flowrate at 50% volume (FEF50) is shown by Figure 2.4.3. where data obtained throughout each of the five trials is plotted. The linear regression line is also drawn in this figure ($r = -0.59$, $p < 0.01$). There is also a significant relationship between FEF50 and FEV1 throughout each trial. The correlation between F50 and FEV1 is not significant. When F50 is correlated with the percentage changes in FEV1 and FEF50 from their baseline values, both relationships are significant. These results are summarised in Table 2.4.2. The changes in F50 from baseline measurement to last dose are statistically significant as is proved by a t–test where the null hypothesis is that the means of the values of F50 before and after the trial are the same. This hypothesis is rejected in all cases, at a 5% significance level.

Data obtained from two non–asthmatic subjects showed that while FEV1 remained constant during these trials there was no systematic change in F50.

It is possible to compare the energy distribution of several power spectra by plotting two features extracted from each as points in two dimensions. This could be done by plotting F85 against F50, but the two measures are clearly related in that as there can be no points such that F85 < F50, half of the x,y plane will be unused. To overcome this problem R85 was defined as the ratio of the difference between F85 and F50 to the total remaining frequency range above F50.

$$R85 = (F85 - F50)*100/(1500 - F50)$$

Points defined by F50,R85 may lie anywhere in the first quadrant of a two dimensional plane.

An example of this type of plot is given in Figure 2.4.4. The distribution of energy in the power spectrum is seen to change as the challenge proceeds. It has been shown in all cases that the increase in F50 is significant, however this is not true of F85. Examining the overall changes in F85 and F50 for all cases reveals that for Patient 1 the breath sound spectrum appears to have simply shifted up in

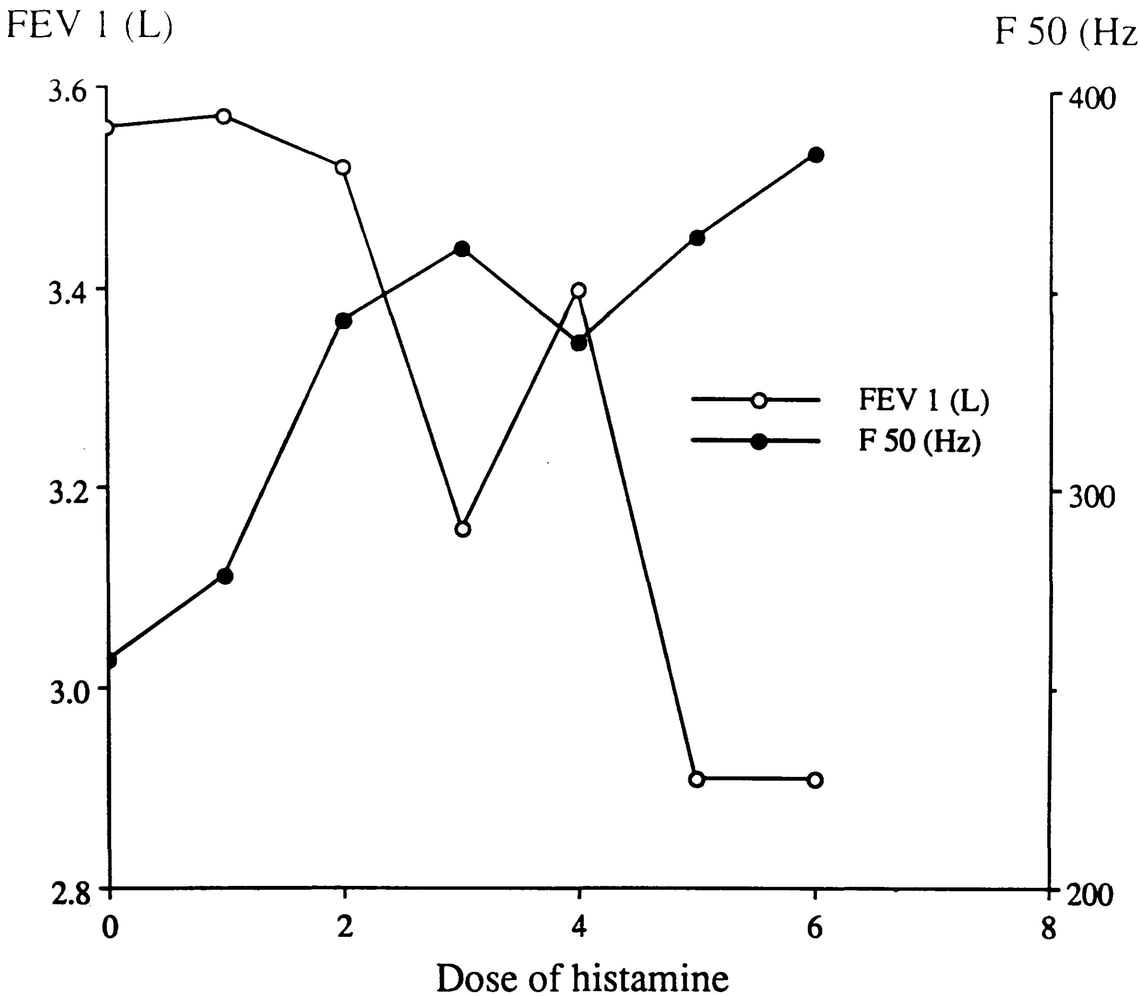


Figure 2.4.2. F50 and FEV1 against dose of histamine.

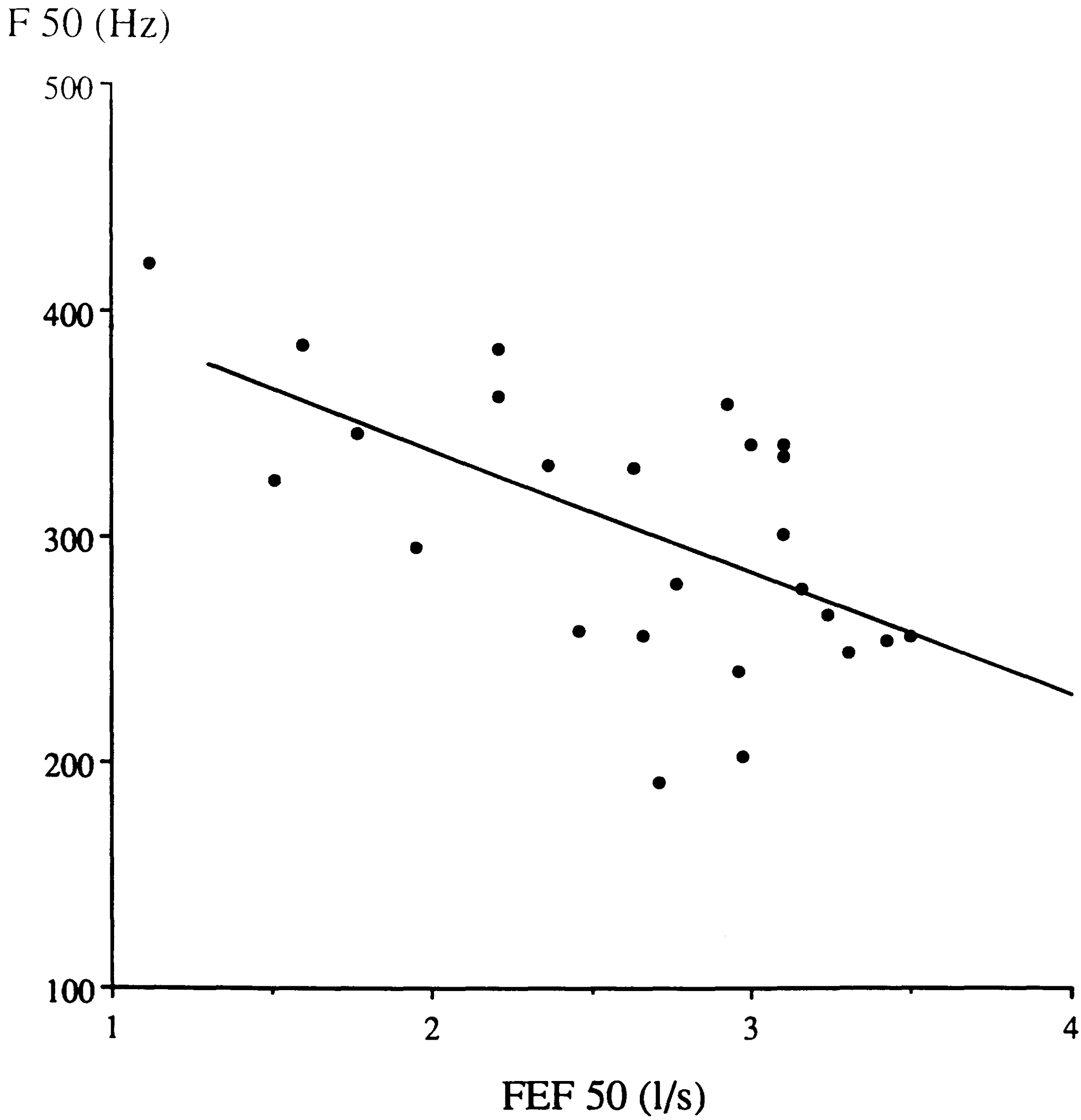


Figure 2.4.3. F50 against FEF50 throughout all trials.

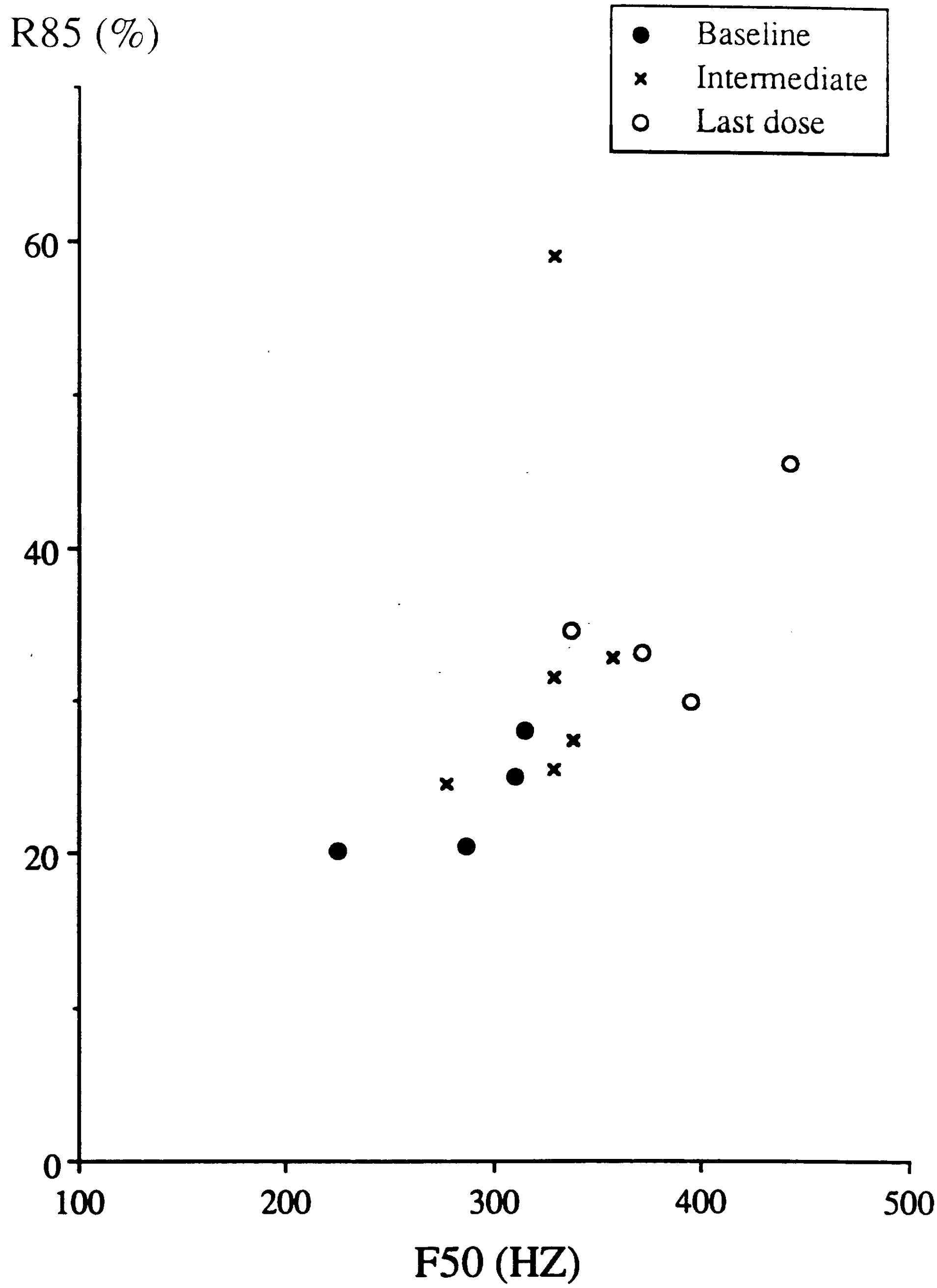


Figure 2.4.4. R85 against F50 for Patient 5.

Baseline ● **Intermediate ×** **Last Dose ○**

frequency while for Patient 3 the spectrum appears to have changed shape, less low frequency energy being evident. This is shown in data-reduced form by Figure 2.4.5. The spectra of Figure 2.4.1. show the changes in sound spectrum for two individual breaths, one baseline recording and one last dose recording for both Patients 1 and 3. Both these spectra, and the data reduced plot of Figure 2.4.5., show that different degrees of change are produced in the breath sounds of these subjects as a result of inhaling histamine.

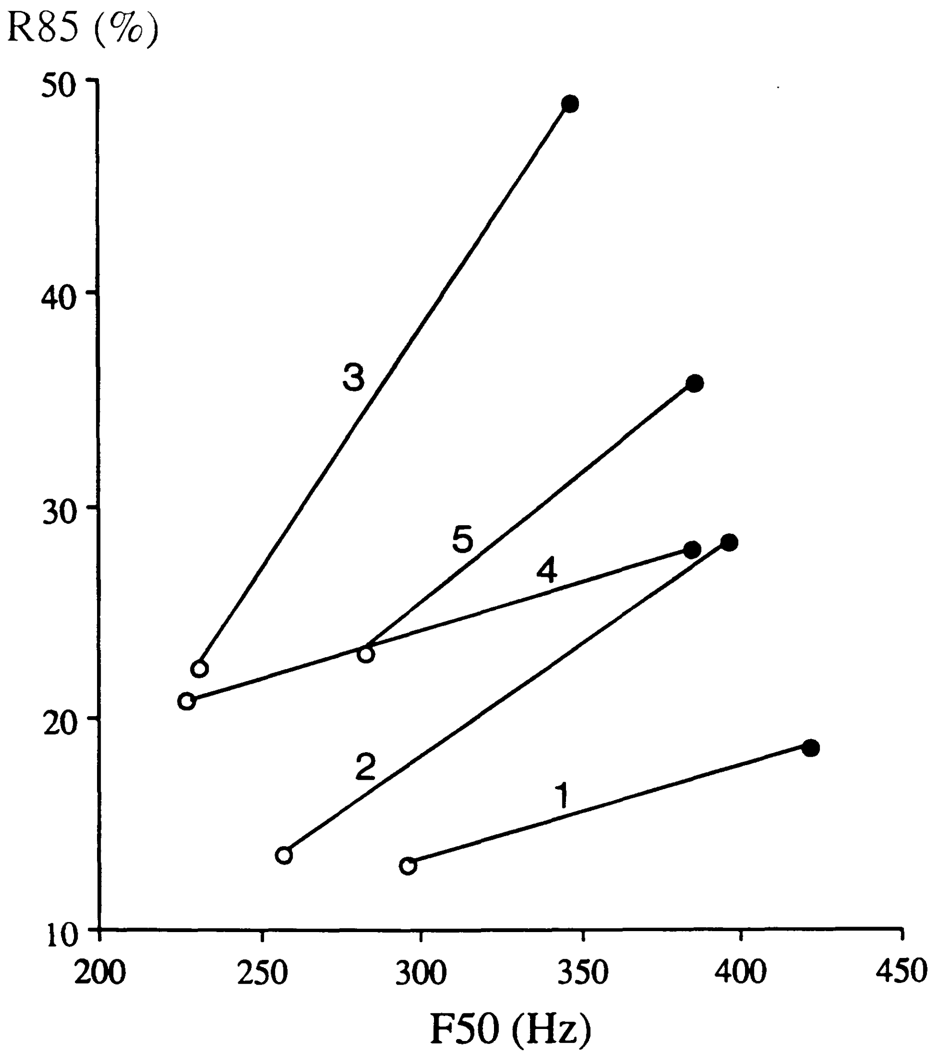


Figure 2.4.5. R85 against F50 for all patients.

Baseline ○ **Last Dose** ●

2.4.5. Discussion

A relationship between F50 and FEF50 and change in FEV1 has been demonstrated. By the use of histamine to cause the bronchi to constrict, it has been possible to measure the changes in breath sound spectrum which accompany a decrease in diameter of the bronchi. These changes have been measured at the base of the lung confirming that during inspiration breath sound detected at this position is not generated peripherally but is conditioned by the generations of bronchi affected by histamine.

From the analysis of the power spectra it appears that different changes of spectrum with histamine, between patients, may be the result of differences in the generations of bronchi which are affected or the degree to which they constrict. For example, it may be possible that breath sounds can be used to distinguish the constriction of the upper bronchi from a general narrowing of many generations of bronchi, in response to a bronchoconstrictor. The altered spectra which we observed occurred in the absence of wheeze suggesting that the airway calibre was not reduced enough to induce the so-called reed-like mechanism suggested as the underlying cause of asthmatic wheeze [10].

In an investigation into the effects of airflow and lung volume on the frequency spectrum of lung sound Kraman [3] concluded that these factors were only weakly related to the median frequency of the sound spectrum. The variation in airflow was achieved by the subjects varying effort during breathing. Kraman's experiments would not change the diameter of the bronchi to the same degree that histamine challenge does, and consequently changes in spectrum of sound are correspondingly smaller. In our investigation the measures of airflow and percentage change in FEV1 were a direct result of narrowing of the airways by the action of histamine.

The conclusion we draw from our study and from Kraman's is that the spectrum of breath sound during inspiration is a function of airway diameter. The diameter and geometery of the upper bronchi determine the pattern of turbulence that is set up during inspiration, and hence the spectrum of acoustic vibrations.

It is generally believed that the spectrum of sound produced in the lung during inspiration may be dependent on airflow, airway geometery and lung volume, and consequently that a change in these parameters should result in a change in the sound spectrum produced. This study suggests that airway diameter is a major factor determining changes in breath sound spectra induced by histamine challenge. The consistent relationship between breath sound and airway calibre is of practical use clinically [11] but can be interrupted by factors which interrupt sound transmission [12].

Any analysis of breath sound where recordings are compared between groups of individuals must also account for the filtering effect of the chest wall on the sound detected. In this study the changes in breath sound were induced in the same patient with, we presume, similar chest wall characteristics throughout the study.

A further development of this method using breath sound analysis as an indicator of changes in airway calibre might be of interest during challenge with provocation techniques such as isocapnic hyperventilation, exercise, cold air or antigen challenge where the site of airway constriction is uncertain.

References

1. Gavriely N. Palti Y. Alroy G. Spectral Characteristics Of Normal Breath Sounds. Physiological Society 1981 :307–314

2. Gavriely N. Palti Y. Alroy G. Frequency Characteristics Of Asthmatic Breath Sounds. Proc. XII Int. Conf.(1979) Medical And Biological Eng. Jerusalem August 19–24

3. Kraman S.S. Effects Of Lung Volume And Airflow On The Frequency Spectrum Of Vesicular Lung Sounds. Respiration Physiology 1986; 66:1–9

4. Campbell E.J.M. Dickenson C.J. Slater J.D.H. Clinical Physiology. Blackwell, London 1974 :105–7

5. Carter R. Mills R.J. The Nebicheck:A Dosimeter For Bronchoprovocation Testing. Breath 1986;29:8.

6. Cockroft D.W. Killian D.N. Bronchial Reactivity To Inhaled Histamine: A Method And Clinical Survey.Clin Allergy 1977;7:235–243.

7. Berglund E. Birath G. Bjure J. Grimby G. Kjellmer I. Sandqvist L. Soderholm B. Spirometric Studies in Normal Subjects between 7 and 70. Acta Medica Scandinavica 1963;173:185–92

8. Knudson R.J. Lebowitz M.D. Holberg C.J. Burrows B. Changes in the Normal Maximal Expiratory Flow Volume Curve with Growth and Ageing. Am. Rev. Resp. Dis. 1983;127:185–92

9. Cotes J.E. Hall A.M. The Transfer Factor For The Lung: Normal Values In Adults. In Normal Values For Respiratory Function In Man. Arcangeli P. (ed) Pasiminerva Medica, Torino 1970;127:725–734

10. Forgacs P. Lung Sounds. Bailliere Tindall, London 1978

11. Spiteri M.A. Cook D.G. Clarke S.W. Reliability Of Eliciting Physical Signs In Examination Of The Chest. Lancet 1988 i:873–5

12. Anderson K. Aitken S. Macleod J.E.S. Moran F. Lung–Sound Transmission And Reliability Of Chest Signs. Lancet 1988 ii:228

2.5. Lung Sound And Respiratory Disorder

2.5.1 Introduction.

This study was conducted on 26 subjects with respiratory disorders resulting from various diseases. The aim was to assess the correlation between the spectrum of lung sound, the condition of the airways and the state of the lung tissue. The extent to which sound is filtered by the lung structure is difficult to characterise in normal subjects and the changes in transmission produced by pathological changes in the lung is unknown [1,2]. Factors determining the generation of lung sound, primarily airway geometery, and one component of the transmission of lung sound, the alveolar structure, are examined in this study. The influence of the pleural membrane and chest wall on transmission is unknown. In a previous study of a single subject it was shown that the transmission of lung sound is greatly altred by pleural effusion [Appendix 1]. No subject in this study had such a condition.

Abnormalities in lung structure can be seen by radiological examination but diffuse changes cannot be easily be quantified. For this reason x–ray examinations were not used in this study. A reduction in diffusing capacity is a feature of fibrotic lung diseases (rheumatoid and asbestosis groups). Therefore diffusing capacity (DLCO) was used as an estimate of the extent to which the alveolar structure was damaged.

The increase in the median frequency of lung sound during inspiration with a reduction in FEV1 was demonstrated by the histamine challenge study [Section 2.4.]. The study now described confirms this result and seeks to account for any changes in sound spectrum resulting from changes in alveolar structure. Data was collected for present study and for the histamine challenge study over the same period of time. The results of the latter study were not known when the present study was being designed.

2.5.2 Pathology

Several lung conditions are associated with rheumatoid disease. These include pleural thickening, pleural effusion and rheumatoid fibrosing alveolitis, a diffuse interstitial disease of the lung. The latter is considered to be a variant of diffuse fibrosing alveolitis which is characterised by diffuse inflammitory processes in the lung, beyond the terminal bronchioles. The alveolar walls thicken and show a tendency towards fibrosis, and abnormal cells are found within the alveolar spaces. The degree of fibrosis may progress from changes in alveolar walls only, through a blurring of the alveolar architecture to severe distortion of the normal lung structure. The changes in lung function which accompany the pathological changes are the restrictive defect, where vital capacity is reduced but the FEV1/FVC ratio is normal, and a reduction in diffusing capacity. Dyspnoea and crackles are common features of rheumatoid fibrosing aveolitis [3].

The cause of fibrosis in asbestosis is the inhalation of asbestos dust. Needle shaped asbestos particles, typically $50\mu m$ in length and $0.5\mu m$ in diameter, are responsible for the harmful effects. These particles are too long to be distributed throughout the lung and they tend to follow the axial bronchi into the lower lobes. The small diameter of the particles increases their chances of penetrating far into the lung. When asbestos needles reach the alveoli they are coated by fibrous tissue. The aveoli are eventually obliterated by fibrosis. The first functional abnormality to occur is a reduction in diffusing capacity which often occurs before x-ray evidence can be identified. A progressive restrictive defect may develop. The early reduction in diffusing capacity is presumed to be caused by alveolar lesions before these have progressed to diffuse fibrosis. Crackles are common [3].

Sarcoidosis may also cause a restrictive defect, when diffusing capacity may be reduced. Pulmonary fibrosis may occur in the later stages of this disease [4].

2.5.3 Method.

Each subject underwent the routine dynamic ventilatory tests where FVC and FEV1 were measured. Diffusing capacity was measured at the same time. Lung sounds were recorded on the same day or within one week of pulmonary function testing.

Lung sound was recorded during early inspiration at three sites on the posterior chest wall. These were the left lower (2cm below the inferior angle of the left scapula), right lower (2cm below the inferior angle of the right scapula) and right upper (at the level of the superior border of the right scapula) locations. The microphone, logging and spectral analysis were performed using the equipment described previously (Section 2.3.). A minimum of three recordings was made at each site, the quality of the sound and the presence of crackles or other extraneous sounds was noted for each recorded breath. The power spectrum of each recording was characterised by calculating the median frequency, F50, and F85 over the frequency range 100Hz to 1500Hz. The intensity (dB) of each breath sound recording was calculated over the range 100Hz to 500Hz as most normal inspiratory breath sound lies within this band. A figure for the noise floor of the microphone and recording system (including hand holding noise) was found. Power spectra whose intensity was within 3dB of this figure were rejected. Recordings where crackles or other extraneous sound occurred were also rejected. A single value of F50 was obtained for each recording site by averaging the F50 values of each breath sound recording made at that site. A single value of F50 for each subject was obtained by averaging all F50 values from all recording sites. The same method was used to find the average values of F85. In the following analysis all values of F50 and F85 are average values.

2.5.4. Results

The pulmonary function measurements of each patient are shown in Table 2.5.1. Three subjects had severe restrictive defects, three had severe obstructive defects and in ten cases the diffusing capacity was below the predicted normal range. The age range of the group as a whole is 20 to 75 years. This is partly due to the length of time the conditions being studied take to develop which determines the availability of suitable subjects. Eighteen of the twenty six subjects were in one of the following groups: asbestosis, rheumatiod, sarcoidosis, the remainder were having treatment for a range of pulmonary disorders. No single group is large enough for a statistical analysis of that group to be valid so the data set was analysed as a whole. This is

Subject Code	Age yrs	Gender	FVC l	FEV1 l/s	DLCO mmol min^{-1} kPa^{-1}	Diagnosis
QA	68	M	2.22	0.67(24)	2.6(45)	RHEUMATOID
PD	68	M	3.04	0.79(26)	7.2(114)	RHEUMATOID
PF	64	F	2.10	1.05(56)	2.6(45)	RHEUMATOID
RC	62	M	3.55	2.03(68)	6.5(96)	RHEUMATOID
RA	27	F	2.60	2.30(71)	5.7(80)	RHEUMATOID
QG	50	M	4.1	2.60(69)	8.1(99)	RHEUMATOID
RB	70	M	2.40	0.70(25)	4.8(78)	ASBESTOSIS
OD	63	M	2.55	1.90(58)	6.1(79)	ASBESTOSIS
NB	67	M	2.83	2.19(88)	6.3(101)	ASBESTOSIS
ND	57	M	3.05	2.25(68)	5.3(70)	ASBESTOSIS
NE	56	M	3.45	2.45(76)	8.1(104)	ASBESTOSIS
QE	60	M	3.55	2.60(75)	6.8(84)	ASBESTOSIS
RD	58	M	4.70	3.00(73)	9.9(112)	ASBESTOSIS
MF	39	F	2.40	1.95(66)	6.3(77)	SARCOIDOSIS
PA	52	M	2.25	2.05(59)	5.6(69)	SARCOIDOSIS
PB	40	M	2.60	2.15(56)	5.4(57)	SARCOIDOSIS
SE	48	M	3.91	3.29(85)	7.4(82)	SARCOIDOSIS
RG	42	M	5.45	3.85(92)	11.1(116)	SARCOIDOSIS
PG	57	F	1.47	0.64(29)	5.8(98)	BRONCHIOL.
SG	52	M	2.50	1.05(28)	———	ASTHMA
ME	62	M	3.19	1.42(44)	4.2(68)	EMPHYSEMA
OA	20	F	2.00	1.74(48)	3.6(39)	P.B.D.
QF	48	M	3.85	3.19(77)	7.5(76)	P.B.D.
QB	62	F	2.75	1.90(88)	4.8(73)	P.B.C
NF	37	M	3.91	3.15(73)	7.0(64)	PULMN.
MD	75	M	3.80	2.05(67)	7.8(99)	P.T.

Table 2.5.1. Pulmonary function data for each subject with age and gender (M = male, F = female). The values in parenthesis are a percentage of the predicted normal for each measurement.

BRONCHIOL. = bronchiolitis
P.B.D. = pigeon breeders disease
PULMN. = pulmonary fibrosis
P.T. = pleural thickening
P.B.C. = primary biliary cirrhosis

justified by the common pathological changes expected.

The occurrance of crackles during inspiration can have a significant effect on the distribution of sound energy. Two spectra obtained from the same subject are shown in Figure 2.5.1., the value of F50 is greatly increased where crackles are present.

There is a correlation between FVC and FEV1 in this group of subjects. Figure 2.5.2. shows that FVC reduces with a reduction in FEV1 ($r=0.80$, $p<0.01$). There is also a reduction in diffusing capacity accompanying a reduction in FEV1, this relationship is shown in Figure 3.5.3. ($r=0.72$, $p<0.01$). A reduction in FEV1 is often but not necessarily accompanied by reductions in DLCO and FVC.

The relationship between F50, F85 and FEV1 at each recording site is shown in Table 2.5.2. where in 11 of 16 calculations there is a significant correlation between lung sound and FEV1 and this indicates that the average F50 or F85 value calculated over all sites correlates best with FEV1. F50 is plotted against FEV1 in Figure 2.5.4., F85 is plotted against FEV1 in Figure 2.5.5. and the linear regression parameters are given in Table 3.5.3.

The correlation between F85 and FEV1,FVC and DLCO are examined in Table 2.5.4. The strongest relationship is with FEV1. The correlation of F85 with FVC and DLCO is through their relationship with FEV1. This view is reinforced by comparing the change in F50 with FEV1 in this experiment with the change observed in the histamine challenge study. The values are -64.5Hz/l and -51.3Hz/l respectively (the figure for the histamine challenge study was obtained by considering four subjects whose baseline FEV1 measurements were similar, resulting in a significant relationship between F50 and FEV1 ($r=-0.47$, $p<0.05$)).

PARAMETERS	LOCATION							
	Right Lower		Left Lower		Right Upper		All Sites	
	r	p	r	p	r	p	r	p
F50,FEV1	−0.59	0.05	−0.08	—	−0.43	—	−0.48	0.05
F50,FEV1%	−0.71	0.01	0.09	—	−0.48	0.05	−0.55	0.01
F85,FEV1	−0.43	—	−0.63	0.05	−0.68	0.01	−0.67	0.01
F85,FEV1%	−0.52	—	−0.62	0.05	−0.55	0.05	−0.68	0.01

Table 2.5.2. Linear regression coefficients.

PARAMETERS y, x	a	b	x_0	r	p
F50,FEV1	282.7	−64.5	2.04	−0.48	0.05
F85,FEV1	626.4	−181.0	2.04	−0.67	0.01

Table 2.5.3. Linear regression parameters.

$$y = a + b(x - x_0)$$

PARAMETERS	r	p
F85,FEV1	−0.67	0.01
F85,FVC	−0.51	0.01
F85,DLCO	−0.52	0.05

Table 2.5.4. Linear regression coefficients.

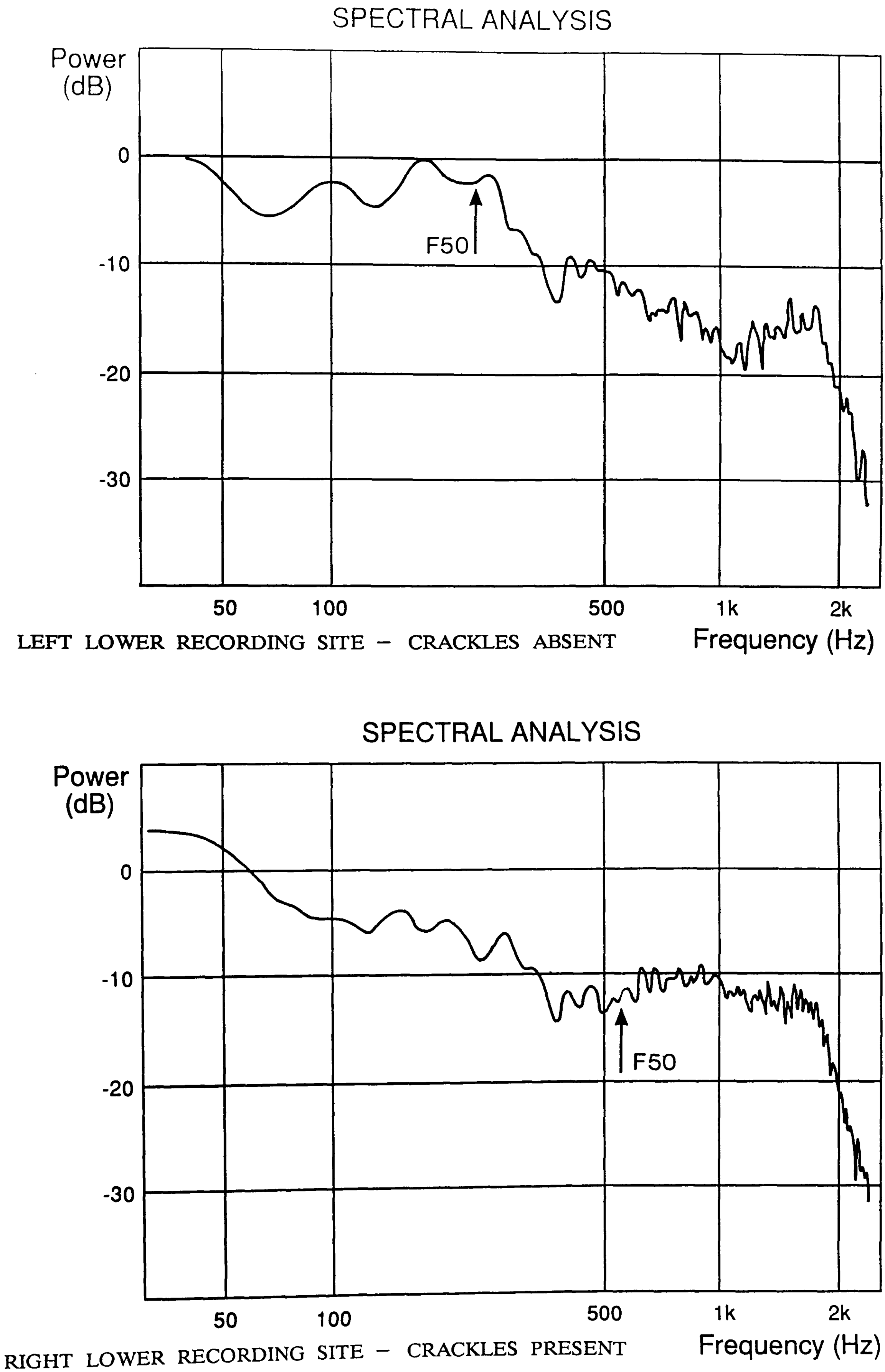


Figure 2.5.1. Lung sound spectra.

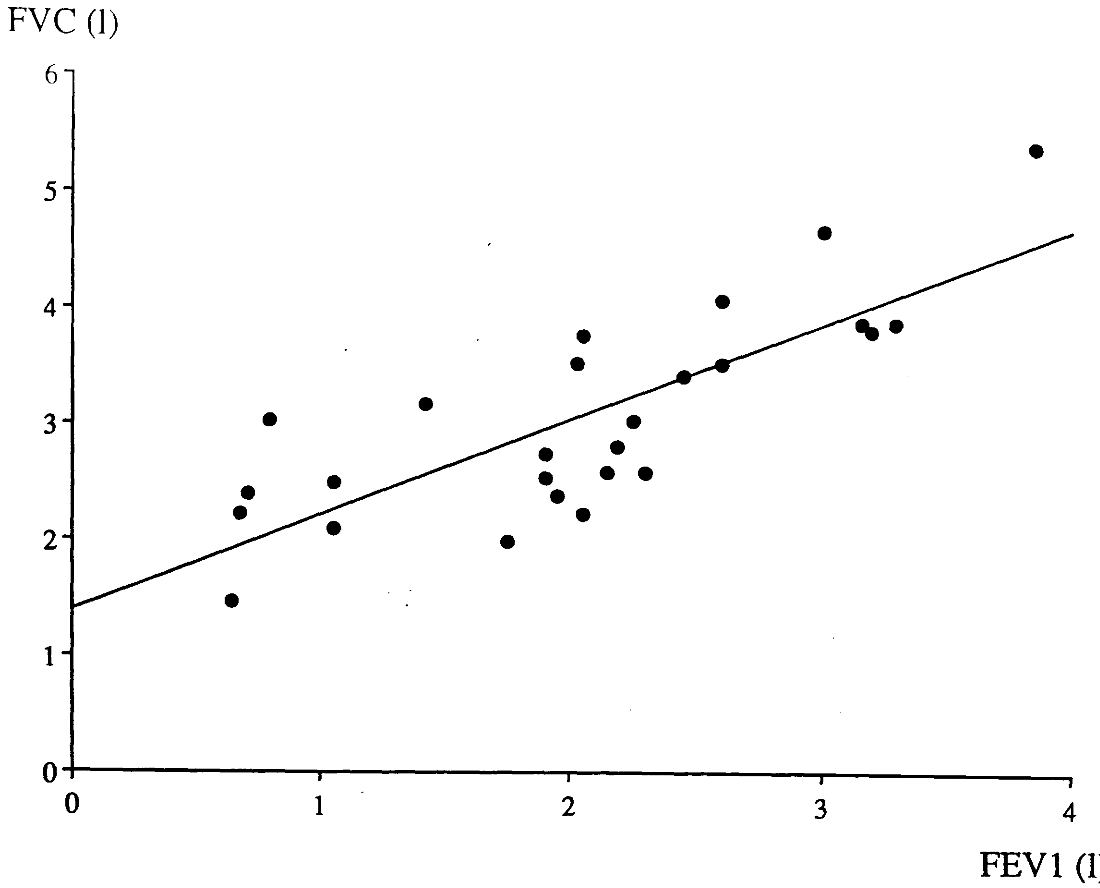


Figure 2.5.2. FVC against FEV1.

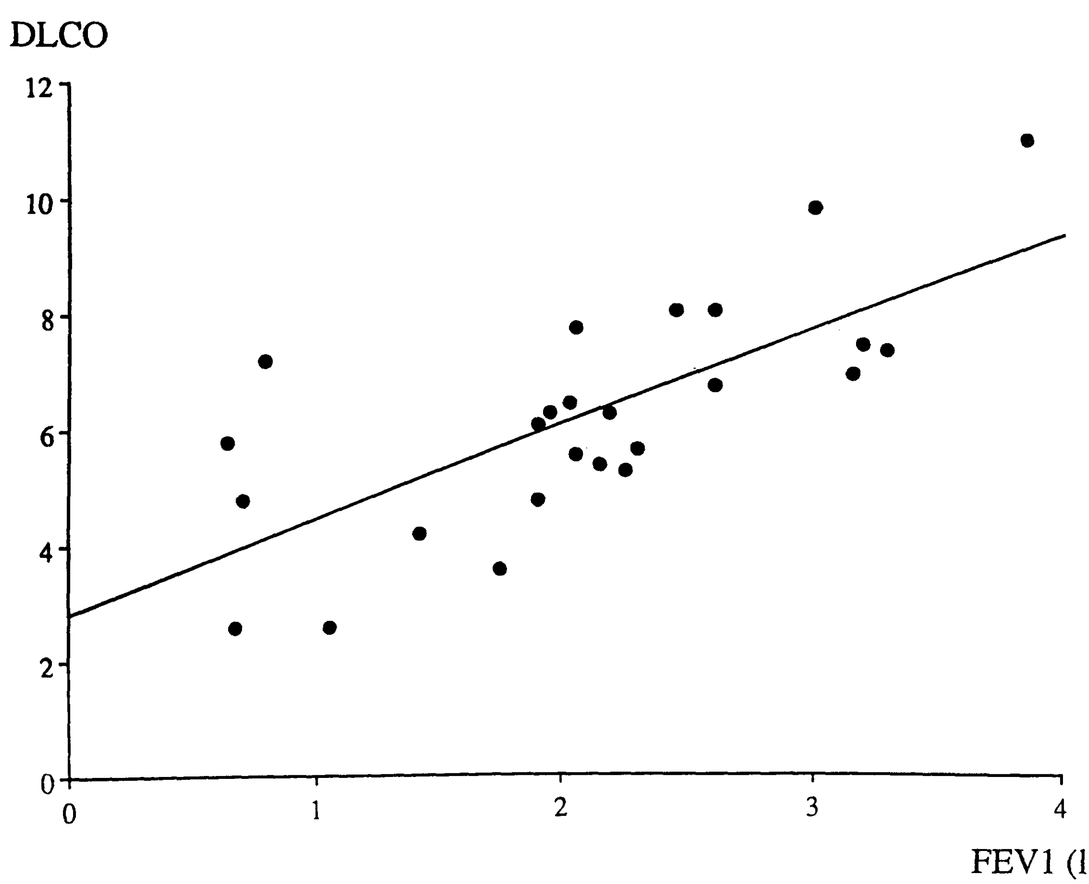


Figure 2.5.3. DLCO against FEV1.

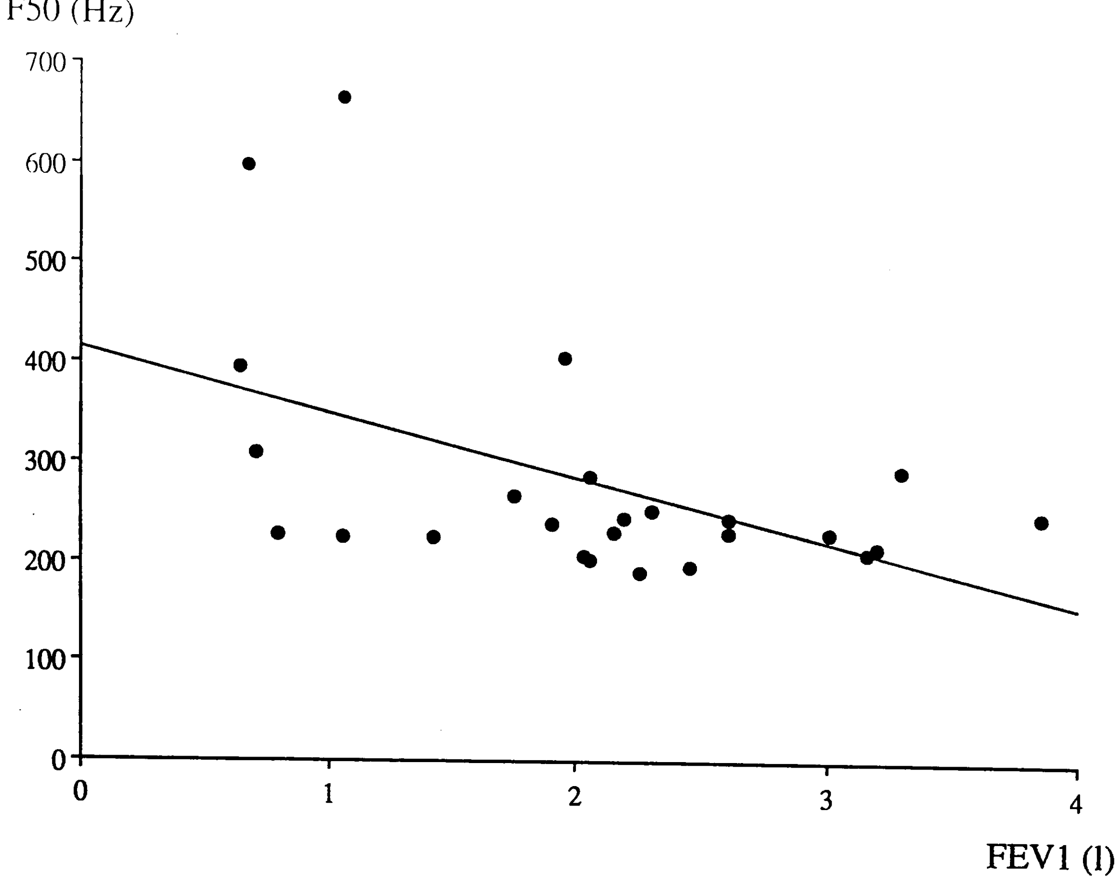


Figure 2.5.4. F50 against FEV1.

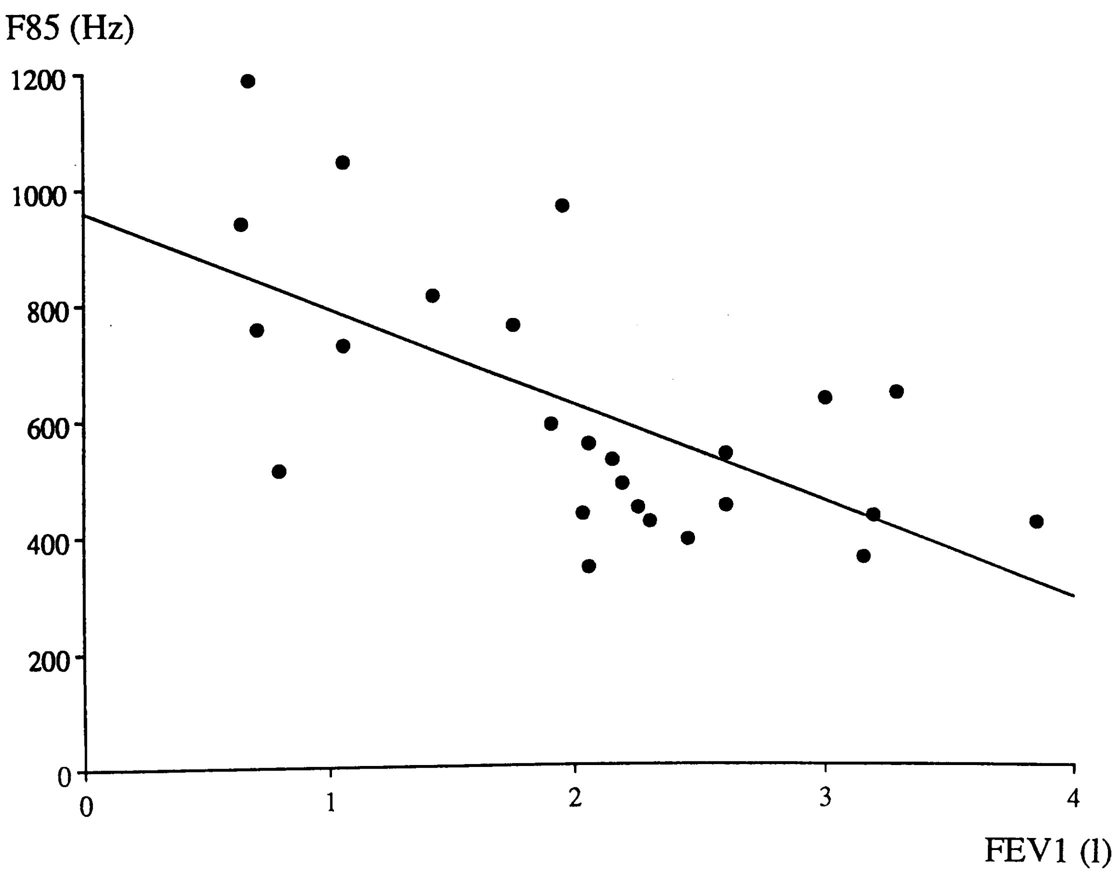


Figure 2.5.5. F85 against FEV1.

2.5.5 Discussion

The results show a significant relationship between the distribution of energy of inspiratory lung sound and FEV1. This confirms the result of the histamine challenge study. Damage to the alveolar structure or small airways does not appear to be the major factor determining the spectrum of lung sound.

If lung sound was generated in the peripheral airways, as is believed by some researchers, then a significant alteration in sound spectrum would accompany a distortion of the structure of the small airways. The results of this study contradict this theory.

It is likely that sound is conducted through the airway walls of the periphery of the lung and this transmission is not significantly dependent on the structure of the tissue but on its density and elastic properties. In this study transmission and generation effects are not separable. Airway calibre is the main factor determining the sound spectrum and it is difficult to estimate the contribution of secondary effects due to transmission. Future work should measure the transmission properties and the elastic properties of the lung to determine the relationship between them and their influence on sound spectrum.

One consequence of proving a general relationship between airway calibre, through FEV1, and F50 is the possibility of assessing regional airway calibre by comparing the spectrum of lung sound recorded at one region to that recorded at another. Transmission factors would also have to be estimated. This would be a non-invasive technique for the assessment of regional airway diameter.

References.

1. Loudon R. Murphy L.H. Lung Sounds. Am Rev Respir Dis 1984; 130:663-673

2. Kraman S.S. New Tools In Lung Sound Research. Seminars In Respiratory Medicine. Vol 6 No 3 Thieme-Stratton, New York 1985

3. Crofton J. Douglas A. Respiratory Disease. Blackwell Scientific Publications, London 1978

4. Flenley D.C. Respiratory Medicine. Balliere Tindall, London 1978

2.6. Conclusions

The major conclusion of this chapter is that the frequency spectrum of inspiratory lung sound is dependent on airway calibre. A decrease in airway calibre, brought about by histamine or through disease, is associated with an increase in the median frequency of the inspiratory lung sound. These results have not previously been reported.

The method of analysing lung sound by examining the inspiratory part of the breath cycle, calculating the Fourier transform and then the median frequency, is shown to be valid by the statistical significance of the results obtained. The author believes that the method could be improved by the use of a microphone with a linear amplitude/frequency response in any future work.

There is no generally accepted method of assessing the transmission characteristics of the chest wall. Differences in sound transmission between various sites on the chest wall of an individual can be estimated, however the results suggest the method to be unreliable. The results of Section 2.5. suggest that differences in chest wall thickness are not a major factor in determining the frequency spectrum of lung sound. It is probable that if the inspiratory sound spectrum could be compensated by a reliable estimate of the transmission characteristic then the relationship between median frequency and airway calibre could be determined with greater accuracy.

Spectral analysis has been used to quantify the inspiratory lung sound detectable at the chest wall. Experimental results indicate that the distribution of sound energy is related to airway calibre. Consequently differences in the distribution of sound energy between different sites on the chest wall may indicate differences in the underlying lung structure, primarily, differences in airway calibre. Conventional methods of flow volume analysis cannot assess such differences.

Chapter 3 Logic, Resolution And Belief:

Literature Review And Theoretical Background

3.1. Introduction

Chapter 3 presents the principles and techniques of automated theorem proving that are employed in the expert system developed by the author. The expert system is presented in Chapter 4. First order logic is the logic system most commonly used to express knowledge and is the system for which automated deduction techniques are most advanced. This chapter begins by developing the syntax (structure) and semantics (interpretation or meaning) of propositional and first order logic. An efficient theorem proving technique for first order logic, known as the resolution principle, is briefly described.

A major theme of this thesis is that while many problems can be expressed in first order logic, this logic cannot clearly express many important concepts such as belief or time. Both of these notions can be analysed in terms of worlds models, where for example, a sentence may be true in a world representing the beliefs of an agent but false in a world representing the real world. The semantics of modal logic can be explained in terms of worlds models and logics of knowledge and belief are often based on one of the modal systems. Much of the analysis of the problem of combining the quantifiers of first order logic and modal operators also applies to logics of belief. Modal logic provides sound systems on which logics of belief can be based, consequently a short study of modal logic precedes the study of belief logics in this chapter.

The deduction model of belief logic developed by Kurt Konolige is the system which is presented in most detail as much of the authors work is based on that system.[1] The authors work (presented in Chapter 4) is concerned with the derivation of new resolution methods for the deduction model of belief. The use of belief logic to express knowledge in an expert system is explored: the practical implementation of this system relies on the new resolution methods.

The use of belief logic allows an agent to reason about the beliefs of other agents and to reason about its own beliefs. Through introspective reasoning an agent can discover what it does and does not know and can deduce facts which are possibly true when there is not enough information to make a sound deduction. This chapter concludes with a review of logic systems which allow deduction to be made when the available information is incomplete. Several of these logics use modal operators but the systems do not correspond to classical modal systems. A new method of reasoning with incomplete information is presented in Chapter 4.

References

1. Konolige K. A Deduction Model Of Belief. Pitman London 1988

3.2. Logic and Resolution

3.2.1. Introduction

This section presents the definition of the logical language that will be used throughout this thesis. The theory of the resolution principle as a rule of inference is introduced. The resolution principle is designed to be suitable for automatic computation. Resolution has only one inference rule, in contrast to the ten rules of derivation which are available in propositional logic [1]. Consequently the proof algorithm need only apply this one rule rather than select from several [2]. The role of logical and Herbrand models is emphasised in this work. Resolution is a purely syntactic method for proving a set of logical sentences to be inconsistent. The method can be made more efficient if the semantic nature of the sentences to be analysed is taken into account.

3.2.2. Polish Notation

The Polish notation for propositional logic differs from the standard notation in that the truth functional operators are written before the formulae on which they operate and not between them [3]. Propositional sentences in Polish notation contain no brackets as the syntax of the notation ensures that there is no ambiguity in deciding which operators apply to which formulae (brackets are often used inconsistently in standard form). These features make the notation useful where the logical sentences are to be manipulated by the computer language Prolog.

The truth functional operators are:

A	Alternation	(or)	Example:	A P Q	$P \vee Q$
K	Conjunction	(and)	Example:	K R S	$R \wedge S$
C	Conditional	(if then)	Example:	C Q R	$Q \supset R$
N	Negation	(not)	Example:	N R	$\neg R$

Definition 3.2.1. The propositional language I is composed of the following symbols.

1. A set of statement letters eg. P,Q,R,S...

2. A set of truth functional operators A,K,C,N.

A statement letter is an atomic formula, by definition.

Definition 3.2.2. A well formed formula (wff) of I is defined by the formation rules.

FR1. An atomic formula is well formed.

FR2. If α and β are well formed formulae, then

A α β; K α β; C α β; N α are well formed.

FR3. A formula is well formed only if it can be constructed by the above rules.

The following formulae are well formed:

A P Q :P or Q

C K P Q R :if P and Q then R

A A R S T :R or S or T (equivalently ((R or S)or T) : (R or (S or T)) etc)

The statement letters are propositions which have truth values true or false. Propositions may be combined and the truth value of the resulting well formed formula calculated from truth tables. The truth tables for the four truth functional operators A,K,C and N are given in Figure 3.2.1. A formula may be tested for validity by the truth table method by calculating the truth value of the whole formula from those of its components for all possible truth value assignments of the components [4].

		α	
A		0	1
β	0	0	1
	1	1	1

		α	
K		0	1
β	0	0	0
	1	0	1

		α	
C		0	1
β	0	1	0
	1	1	1

N		
α	0	1
	1	0

Figure 3.2.1. Truth Tables for the formulae

$A\alpha\beta$, $K\alpha\beta$, $C\alpha\beta$ and $N\alpha$

Formulae may also be proved to be valid by using inference rules, for example the rule MP (Modus Ponens)[1] is the rule of derivation which allows Q to be deduced from P and C P Q. A further type of proof method is known as analytic tableau [5]. These are mechanical methods useful for demonstrating proofs but not suitable for automation. The cancellation technique developed by Snyder is an example of a tableau method where only one inference rule can be selected at each step in the proof [3]. However the proof may split and if it does so n times then 2^n sequences of atomic formulae are generated, each of which must be tested against certain criteria. The number of sequences becomes unacceptably large for long or complex formulae.

In propositional logic the internal structure of the proposition or statement cannot be analysed. This structure may be examined by expressing the proposition as the relationship of a property to an object [6]. For example 'snow is white' may be considered as a whole to be a proposition or to be the property 'white' defined on an object 'snow', written 'white (snow)'. Objects (or individuals) may be constants or variables. The sentence 'everything has property P' is written $\forall x$ P(x) where $\forall$ is the universal quantifier. For this sentence to be meaningful some individuals (or objects) must exist and all must have the property P for the sentence to be true. It is clear that universally quantified formulae must be true generalisations of actual predicate–constant relations. The universal quantifier cannot be interpreted as 'most' or 'some' which are more common notions in natural language.

Defininion 3.2.3. The predicate language IP is composed from the following symbols.

1. A set of predicates of degree n written P^n.
2. A set of individual variables x,y,...
3. A set of individual constants a,b,...
4. A set of functions of degree n written f^n
5. A set of truth functional operators A,K,C,N.
6. The quantifier symbols $\forall$, $\exists$.
7. The symbols (and).

Before defining the construction of well formed formulae we define terms and atomic formula. A predicate is of degree n if it has n argument terms, $P(x_1,...,x_n)$ where $x_1 ... x_n$ are terms. If a variable, x, occurs in a formula where it is not defined to be universally or existentially quantified, then x is said to be a free variable.

Definition 3.2.4. Terms are defined as follows.

1. A constant is a term.

2. A variable is a term.

3. If f is a function of degree n and $t_1 \ldots t_n$ are terms then $f(t_1 \ldots t_n)$ is a term.

Definition 3.2.5. Atomic formulae (atoms) are defined as follows.

If P is an n place predicate symbol and $t_1 \ldots t_n$ are terms then $P\ (t_1 \ldots t_n)$ is an atom.

Definition 3.2.6. A well formed formula of IP is defined by the formation rules.

FR1. An atom is a well formed formula.

FR2. If α and β are wff then $A\ \alpha\ \beta$; $K\ \alpha\ \beta$; $C\ \alpha\ \beta$; $N\ \alpha$ are well formed.

FR3. If α is a formula and x is free in α then $\forall x\ \alpha$ and $\exists x\ \alpha$ are well formed.

FR4. A formula is well formed only if it can be constructed by the above rules.

The following formulae are well formed.

$\forall x\ C\ P(x)\ Q(x)$ For all x, if x has P then x has Q.

$C\ \forall x\ P(x)\ \ \exists y\ P(y)$ If all x have P then there exists a y such that y has P.

There is no ambiguity over the scope of the quantified variables in Polish notation. The existential and universal quantifiers are interdefinable:

$\exists x\ \alpha = N\ \forall x\ N\ \alpha$ [4]

Semantics

Propositional formulae may be tested for validity by assigning truth values to the atomic formulae, however, in predicate logic individual variables cannot be assigned truth values. To test the validity of sentences in IP a model must exist. An IP model consists of a non-empty set D of individuals $\{u_1 \ldots ui \ldots\}$ together with a value assignment V and is written $<D,V>$. V assigns members of D to each individual variable, that V assigns u to x is written $V(x)=u$. For predicates of degree n V assigns sets of n members of D (n-tuples). For each predicate Φ, $V(\Phi)$ is some set of ordered n-tuples:

$\{<u_{i_1},\ldots,u_{in}>,<u_{j_1},\ldots,u_{jn}>,\ldots\}$ of members of D.

Φ is said to be an n-adic variable.

Given $V(x_i)$ and $V(\Phi)$ for each individual variable x_i and each predicate Φ, all formulae are evaluated as follows.

Definition 3.2.7. Validity in IP.

1. For an atomic formula, that is, one consisting of an n-adic variable Φ followed by n individual variables $x_1 \ldots xn$ $(n>0)$
$V(\Phi(x_1,\ldots,xn))=1$ iff $<V(x_1),\ldots,V(xn)> \in V(\Phi)$
otherwise $V(\Phi(x_1,\ldots,xn)=0$. The formula has the value 0 or 1 according to the n-tuples of those assigned by V to Φ.

2. For the truth functional operators the truth values 0 and 1 are assigned according to the truth table definitions as in the logic I.
eg. $V(A\ \alpha\ \beta)=1$ if $V(\alpha)=1$ or $V(\beta)=1$, else $V(A\ \alpha\ \beta)=0$.

3. If α is a wff and x is an invidual variable then
$V(\forall x\ \alpha)=1$ if for every IP assignment V' which gives to all variables except x the same values as V gives to them $V'(\alpha)=1$ otherwise $V'(\alpha)=0$. This expresses the idea that if α is true irrespective of the value assigned to x, then $V(\forall x\ \alpha)=1$.

In this logical model a set of individuals is defined and these individuals (as n-tuples of individuals) are assigned to each predicate. Formulae are tested for validity by verifying that the value assignment for the individual variables of a predicate correspond to the values assigned by the set of n-tuples for that predicate. If a variable x_i is universally quantified the model must define n-tuples such that each member of D appears at the position of x_i in the n-tuple, in at least one n− tuple, for the formula to be true (otherwise there would be an individual which may be assigned to x_i for which there would be no n− tuple).

3.2.3. Normal Forms

A sentence or formula in a logical language is equivalent to infinitely many other sentences. It is useful to be able to express all sentences in an equivalent standard form. Prenex normal form is one such form. It has the following syntactic structure:

$$Q_1x_1Q_2x_2\ldots Q_nx_n \; M$$

Where $Q_1\ldots Q_n$ are quantifier symbols, $x_1\ldots x_n$ are variables and M is called the matrix and contains no quantifiers. All sentences in IP can be expressed in this format. The matrix is a combination of atomic formulae which may be standardised further by converting it to conjunctive normal form (c.n.f.) or disjunctive normal form (d.n.f.). In c.n.f. the matrix is expressed as a conjunction of disjunctions of atoms. Any formula may be converted into conjunctive normal form by the standard logical equivalences, and will then have the structure:

$$\begin{array}{ll} Q_1x_1\ldots Q_nx_n & K\ A\ \alpha_1\ A\ \alpha_2 \ldots A\ \alpha_{n-1}\ \alpha_n \\ & K\ A\ \beta_1\ A\ \beta_2 \ldots A\ \beta_{n-1}\ \beta_n \\ & \ldots \\ & \quad A\ \gamma_1\ A\ \gamma_2 \ldots A\ \gamma_{n-1}\ \gamma_n \end{array} \qquad (1)$$

The remaining step is to eliminate the existential quantifiers from the prefix and to replace the corresponding variables by skolem functions. This can be done by considering variables such as y in the sentence $\forall x \exists y\ P(x,y)$ to be dependent on the choice of x, that is, y can be considered to be a function of x. The skolem transform of $\forall x \exists y\ P(x,y)$ is $\forall x\ P(x,f(x))$ where f(x) is a skolem function which is new to the entire set of sentences under consideration. An important property of the skolem transform is that if a set of sentences is unsatisfiable then the skolem transform of this set will also be unsatisfiable [7,8].

A sentence is in clausal form if all variables in the prenex form are universally quantified and the matrix is in conjunctive normal form. The resolution principle makes use of the clausal form of sentences in attempting to prove a theorem. As sentences in clausal form have a standard structure the A and K operators need not be written and all variables may be assumed to be universally quantified. Thus equation (1) may be rewritten:

$$(\{\alpha_1 \ldots \alpha n\},\{\beta_1 \ldots \beta n\},\ldots\{\gamma_1 \ldots \gamma n\})$$

where each clause is a disjunction of atoms enclosed by brackets {} and the set () is the conjunction of the member clauses.

3.2.4. The Herbrand Model and Herbrand's Theorem.

Earlier the validity of logical formulae was shown by referring to a logical model (which could be constructed in theory at least). In a similar way Herbrand models are used in studying the semantics of a set of clauses. The Herbrand universe of a set of clauses S is defined as follows.

Definition 3.2.8. The Herbrand universe. Let H_0 be the set of constants appearing in S. If no constant appears in S then H_0 is to consist of a single constant, $H_0=\{a\}$. For i=0,1,2... let H_{i+1} be the union of H_i and the set of all terms of the form $f^n(t_1 \ldots tn)$ for all n place functions f^n occuring in S, where t_j, j=1,2...n are members of the set H_i. Each H_i is called the i level constant set of S and H ($H=H_\infty$) is called the Herbrand universe of S.

For example: Let S = ({P(a)},{NP(x),P(f(x))})

$H_0=\{a\}$

$H_1=\{a,f(a)\}$

$H_2=\{a,f(a),f(f(a))\}$

...

$H_\infty=\{a,f(a),f(f(a)),\ldots\}$

The Herbrand universe is the set of ground terms (variable free terms) of the atomic formulae occuring in a set of clauses. The Herbrand universe may be infinite as in the example above [8].

The Herbrand model of a set of clauses is the set of all atomic formulae obtained by replacing the variables in each clause by members of the Herbrand universe. The Herbrand model is a set of ground atomic formulae (or ground literals).

For example: The set of clauses S = ({P(x)},{Q(a),Q(b),Q(c)})

has the Herbrand universe H = {a,b,c}

and the Herbrand model M = {P(a),P(b),P(c),Q(a),Q(b),Q(c)}

The truth of a sequence of formulae in clausal form (clausal sequent) can be established by showing that the corresponding set of clauses is unsatisfiable. The assignment of truth values to the ground literals in a model M determines whether M satisfies a given clausal sequent. This assignment is known as the Herbrand map. This is analagous to the method of proving the validity of predicate formulae by defining a logical model and verifying that the n–tuples of constants assigned by the model to each predicate correspond to those constants which may be assigned to variables in a formula (according to the definition of the quantifier of each variable).

Truth values can be assigned to to the members of M and a tree diagram constructed (map tree) to cover all possible truth assignments. If a node of the map tree assigns a truth value to a literal which contradicts that assigned by the sequent then the tree need not be expanded from that node. Otherwise the tree grows by selecting a ground literal from M and drawing two branches from the original node, assigning the values true and false for this literal to the left and right branches respectively.

For example: The sequent H(a),∀x C H(x) M(x) => M(a)

is true when S = ({H(a)},{NH(x),M(x)},{NM(a)}) is unsatisfiable.

In this case M = {H(a),M(a)} and the Herbrand map tree for S is shown in Figure 3.2.2. In this diagram each node which contradicts a clause in S is labelled with that clause. For the sequent under consideration the map tree is finite and the tip of each branch is labelled with a member of S (the tree is said to be closed). Therefore the tree cannot be expanded further and there is no truth assignment which satisfies S. This procedure shows that NM(a) is not consistent with H(a) and

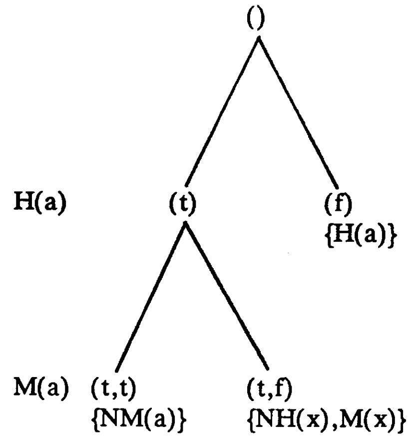


Figure 3.2.2. A Herbrand map tree.

$\forall x$ C H(x) M(x) and therefore that the original sequent is true [7].

In this method the truth of a sequent is established by the fact that no Herbrand model satisfies the set of clauses S derived from that sequent. This proof procedure makes use of Herbrands theorem which can be stated in two forms.

Herbrands Theorem

1. A set S of clauses is unsatisfiable if and only if there is a finite unsatisfiable set S' of ground instances of clauses of S.

2. A set S of clauses is unsatisfiable if and only if there is a finite closed Herbrand map tree for S.

The Herbrand map tree will be finite for a true sequent whether the Herbrand base is finite or infinite, this is known as the compactness phenomena (when the Herbrand base is finite the Herbrand map tree is finite for a false sequent).

3.2.5. Resolution

The resolution principle is a method for combining two clauses from a set of clauses S and deriving a resultant clause R. It can be shown that if S is unsatisfiable then $S \cup R$ is also unsatisfiable, consequently resolution is a sound inference principle. The resolution principle is a generalisation of ground resolution which can be defined in the following way.

Definition 3.2.9. Ground resolution. If C and D are two clauses and $L \subseteq C$, $M \subseteq D$ are two ground literals which form a complementary pair then the ground clause $\{C-L\} \cup \{D-M\}$ is called a ground resolvant of C and D. The set consisting of S together with all ground resolvents of all pairs of members of S is known as the ground resolution of S. If S contains the empty clause then S is unsatisfiable [2].

The connection between resolution and the map tree proof procedure (based on Herbrands theorem) is illustrated by the following example.

Consider the sequent H(a),C H(a)M(a) => M(a), which is true when

$$S_0 = (\{H(a)\},\{NH(a),M(a)\},\{NM(a)\})$$

a set of ground clauses, is unsatisfiable. The Herbrand map tree for S_0 is shown in Figure 3.2.3. The ground resolvant of the second and third clauses in S_0 is {NH(a)} (M(a), NM(a) are the complementary pair equivalent to L and M in Definition 3.2.9). A new set of clauses $S_1 = (\{H(a)\},\{NH(a),M(a)\},\{NM(a)\},\{NH(a)\})$ may be constructed by adding {NH(a)} to S_0, the map tree for this set is shown in Figure 3.2.4. The map tree for S_1 is smaller than that for S_0 and this is typical where an unsatisfiable set of clauses is expanded by adding a resolvant clause [7]. The first and last clauses of S_1 may be resolved to give the empty clause $\square$ and the set $S_2 = (\{H(a)\},\{NH(a),M(a)\},\{NM(a)\},\{M(a)\},\square)$ may be constructed with the map tree shown in Figure 3.2.5. The empty clause has been derived by two steps of ground resolution. By the resolution principle the sets S_2, S_1 and S_0 are unsatisfiable and therefore the original sequent is true.

The above example shows how resolution proceeds when ground literals which form a complementary pair can be identified. Where the literals in a clause contain variables a method for correctly substituting terms for variables must be found. Two complementary literals must have the same predicates but opposite truth values and the terms on which the predicates are defined must be made equivalent under some substitution of terms (which may be a combination of constants, variables or functions). The following example shows how this can be achieved.

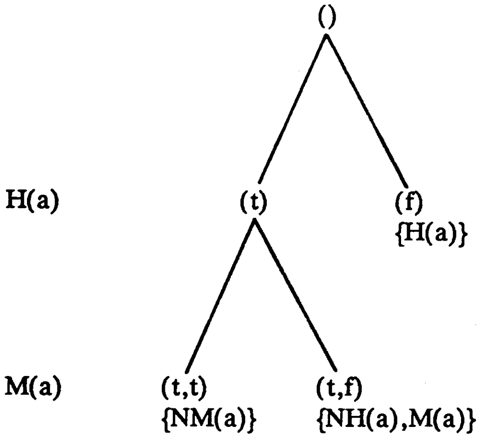


Figure 3.2.3 Herbrand map tree for S_0.

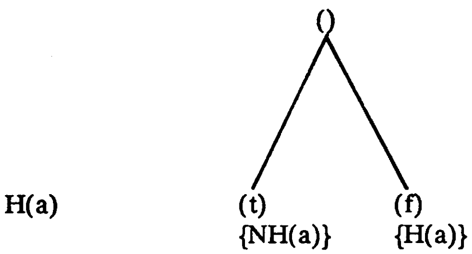


Figure 3.2.4. Herbrand map tree for S_1.

Figure 3.2.5. Herbrand map tree for S_2.

Consider two clauses $A = \{P(x), Q(x)\}$

$$B = \{NP(f(x)), R(x)\}$$

There is no literal (atomic formula) in A which is immediately complementary to one in B, however if f(x) is substituted for x in A then

$$A' = A\sigma = \{P(f(x)), Q(f(x))\}$$

where $\sigma = \{f(x)/x\}$ is the substitution that is applied to A. The resolvant of A' and B is $C = \{Q(f(x)), R(x)\}$. The most general substitution which unifies two literals can be found by the unification algorithm (described in detail in Robinson [7]). In this example σ is the most general unifier of A and B, $A\sigma$ is called a factor of A.

The resolution principle can now be stated for first order predicate logic.

Definition 3.2.10. The resolution principle. Let C and D be two clauses with no variables in common and $L \subseteq C$, $M \subseteq D$ (L and M are two literals in C and D). If L and M have a most general unifier σ then the clause

$$\{C\sigma - L\sigma\} \cup \{D\sigma - M\sigma\}$$

is called a binary resolvant of C and D. The literals L and M are called the literals resolved upon. This rule is known as the binary resolution rule.

The completeness of the resolution principle (i.e. the fact that it can prove all sets of unsatisfactory sets of clauses to be so) can be shown by a theorem based on the Herbrand map tree diagrams [7].

Hyperresolution.

The binary resolution rule shows how two clauses may be combined to derive a third, where if C and D are the parent clauses with m and n literals respectively then the new clause will contain m+n−2 literals. The binary resolution rule may be restated as follows.

$$\begin{array}{l} \{C_1\sigma \ldots Ci\sigma \ldots Cm\sigma\} \\ \underline{\{D_1\sigma \ldots Dj\sigma \ldots Dn\sigma\}\qquad} \\ \{C_1\sigma \ldots C_{i-1}\sigma, C_{i+1}\sigma \ldots Cm\sigma, D_1\sigma \ldots D_{j-1}\sigma, D_{j+1}\sigma \ldots Dn\sigma\} \end{array}$$

where σ unifies Ci and Dj

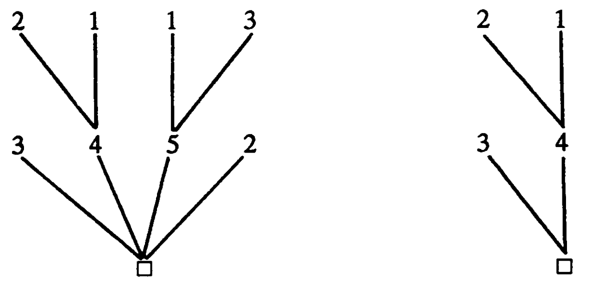
The rule can be used directly to generate new clauses by attempting to match each literal in each clause with every literal in every other clause. Where the literals are found to be complementary and the variables unifiable the new clause is added to the clause set. This method will be referred to as scheme I. In this procedure the number of matching operations is proportional to the square of the total number of literals. This search procedure is inefficient in two ways. Firstly, many clauses will be generated from which the empty clause cannot be derived, consuming computing time and resources. Secondly the meaning (or semantic content) of the original logical formulae is lost by regarding each clause simply as a list of literals, although from the syntactic viewpoint developed earlier the method of derivation is valid.

One method for reducing the search space of binary resolution is to order the set of predicates and to always begin by evaluating the predicate of greatest weight. A similar method is to evaluate the literals in a clause in a standard order, for example left to right. The advantages of ordering the literals is demonstrated in the next example.

Consider the set of clauses $S=(\{L_1,L_2\},\{NL_1\},\{NL_2\})$. In scheme I the binary resolution rule can be applied four times:

$$
\begin{array}{llll}
1.\{L_1,L_2\} & 2.\{L_1,L_2\} & 3.\{L_1\} & 4.\{L_2\} \\
\quad \underline{\{NL_1\}} & \quad \underline{\{NL_2\}} & \quad \underline{\{NL_1\}} & \quad \underline{\{NL_2\}} \\
\quad \{L_2\} & \quad \{L_1\} & \quad \Box & \quad \Box
\end{array}
$$

If the order requirement is in force then the resolution rule may only be applied twice, steps 1. and 4. above. This method will be referred to as scheme II. The inference map for scheme II is considerably smaller than that for scheme I as is shown in Figure 3.2.6. The order requirement reduces the duplication of effort which occurs in scheme I while the resolution method remains complete [1].



Scheme I Scheme II

Figure 3.2.6. The inference maps for resolution schemes I and II.

Where $1=\{L_1,L_2\}$ $2=\{NL_1\}$ $3=\{NL_2\}$ $4=\{L_2\}$ $5=\{L_1\}$

The search space may be reduced further by restricting one of the parent clauses to the set of unit clauses (clauses containing a single literal) and making the resultant clause a unit clause also. This gives the hyperresolution rule:

Definition 3.2.11. Hyperresolution.

$$
\begin{array}{l}
\{L_1\sigma \ldots Lm\sigma, Ln\sigma\} \\
\{NL_1\sigma\} \\
\ldots \\
\underline{\{NLm\sigma\}\qquad} \\
\{Ln\sigma\}
\end{array}
$$

where $Li\sigma$ and $NLi\sigma$ for $i=1$ to m are complementary literals under the most general unifier σ. Intermediate steps of the form

$$\begin{array}{l} \{Li\gamma \ldots Lm\gamma, Ln\gamma\} \\ \underline{\{NLi\gamma\}\qquad\quad} \\ \{L_{i+1}\gamma \ldots Lm\gamma, Ln\gamma\} \end{array}$$

are buried in order to see the operation as a whole. The inference map generated by this rule for the clauses in the previous example is shown in Figure 3.2.7. The empty clause is deduced in one step of hyperresolution.

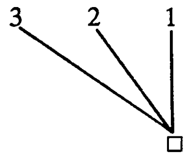


Figure 3.2.7. The inference map for hyperresolution.

The literals $L_1 \ldots Lm$ in the rule clause are evaluated from left to right as in scheme II. If the logical rules are restricted to having a single atomic formula α as the conclusion then the literal corresponding to α will always be the right most literal and the only candidate for the resultant unit clause Ln. Such clauses are known as Horn clauses. In this way a semantic restriction (one based on the meaning of the logical rule) is written into the resolution method in such a way that the search space is reduced.

In hyperresolution the conditions of the rule are examined, in order, and if all can be satisfied the conclusion can be derived. The rules are viewed as deduction rules. From a syntactic viewpoint the order of the literals is of no consequence for the validity of binary or hyperresolution methods. In Definition 3.2.11 L_1 could be swapped for Ln, L_1 becoming the resultant clause and definition would remain true, however the search space would be increased if each literal could be considered as a candidate for Ln (removing the semantic restriction increases the search space).

The importance of considering only unit clauses for resolution with literals in the rule clause is shown by contrasting hyperresolution, a forward chaining strategy with the backward chaining search strategy.

3.2.6. Forward and Backward Chaining.

A backward chaining strategy utilising Horn clauses attempts to match the left most literal in the goal clause with the right most literal in a rule clause or a unit clause. The resolvant becomes the new goal clause. If the empty clause can be deduced then the set of clauses is inconsistent, usually as the result of introducing the original goal clause [9]. It is often necessary to begin a search from a given goal. This backwards chaining method adds a semantic restriction to binary resolution and can be stated formally.

Definition 3.2.12. Backwards resolution.

$$\begin{array}{ll} \text{Clause 1} & \{G_1\sigma \ldots Gi\sigma\} \\ \text{Clause 2} & \underline{\{L_1\sigma \ldots Lm\sigma, Ln\sigma\}\quad\quad} \\ & \{L_1\sigma \ldots Lm\sigma, G_2\sigma \ldots Gi\sigma\} \end{array}$$

where $G_1\sigma$ and $Ln\sigma$ are complementary literals

The differences between hyperresolution and backwards chaining are explored in the following discussion.

As proof depends on the derivation of the empty clause, this requires rule clauses (or goal clauses) to be resolved with unit clauses otherwise the number of clauses in the resolvant will not decrease. The concrete search space can be defined as the original set of clauses linked to the unit clauses which may be derived from them. The proof of a theorem will be found within this search space but this space does not correspond to an optimum or minimal search space. The search space of hyperresolution is exactly the concrete search space.

During backward chaining the set of clauses that is generated is very different from those generated by hyperresolution. This is a problem when attempting to compare the size of the search spaces. The problem may be overcome by representing an application of the backwards resolution rule by a point in a 'transformed' search space, labelled by the right most literal in Clause 2 (Definition 3.2.12.). Taking the clauses in Definition 3.2.12. as an example, $\{Ln\sigma\}$ would be added to the search space. If one of the literals $L_1\sigma \ldots Lm\sigma$ cannot be resolved away the backwards chaining algorithm will backtrack and attempt to find another clause to resolve with Clause 1. In this case $\{Ln\sigma\}$ remains part of the search space but as the conditions on which it depends cannot be satisfied this clause is not in the concrete search space. This shows that the search space of this backwards chaining method may be greater than that of hyperresolution, however in practice the characteristics of the logical rules in question (i.e. whether they are recursive or not) is an important consideration in selecting the most appropriate search strategy.

3.2.7. Theory resolution.

Theory resolution enables the axioms of a theory to be built directly into the resolution procedure. The axioms themselves need not be directly resolved upon which can reduce the length of proof and size of the search space. The theory is a method for determining whether a set of literals is unsatisfiable, often using a non-syntactic method [10]. The ground case of a version of theory resolution can be defined:

Definition 3.2.13. Total narrow theory resolution. Let $C_1 \ldots C_m$ be a set of non-empty clauses. Let each C_i be decomposed as $L_i \cup K_i$ where K_i is a unit clause. Let $R_1 \ldots R_n$ be unit clauses.

$$\frac{\begin{array}{l} L_1 \cup K_1 \\ \ldots \\ L_n \cup K_m \end{array}}{L_1 \cup L_2 \cup \ldots L_m}$$

where $K_1 \ldots K_m, R_1 \ldots R_n$ is unsatisfiable

The theory, defined by $R_1 \ldots R_n$, determines whether the set $K_1 \ldots K_m$ is unsatisfiable. For example theory resolution may be used to show the literal less−than(c,a) is unsatisfiable under the theory T_1, where T_1 defines the ordering of constants by their position in the alphabet. By total narrow theory resolution:

$$\frac{\{\text{less−than}(c,a), \text{greater−than}(f,e)\}}{\{\text{greater−than}(f,e)\}}$$

where less−than(c,a) is unsatisfiable in T_1
(K_1= less−than(c,a),R_1= N less−than(c,a) in Definition 3.2.13.)

In theory resolution the resolvant of a set of clauses is not found by making literals syntactically identical but rather the unsatisfiability of a set of literals is determined by a theory and consequently there are semantic conditions on the resolution process.

3.2.8. Conclusions

The logical language IP and the logical model which determines the validity of sentences in this logic have been developed. From a discussion of normal forms and Herbrands theorem a mechanical theorem proving method is derived which can be shown to be complete. The parallels between the role of n−tuples of individuals or atomic formula in logical or Herbrand models respectively, are evident. In the implementation of resolution on a computer reducing the size of the search space is important in finding solutions to problems. Hyperresolution, where inferences are made from atomic formula, is shown to have a smaller search space than backward chaining methods where the same semantic restrictions are used. The role of semantics in reducing the search space of resolution and of theory resolution is important.

Both hyperresolution and total narrow theory resolution are used in the inference program which forms part of expert system described in the following chapter.

References

1. Lemmon E.J. Beginning Logic. Van Nostrad Reinhold, Wokingham 1965

2. Robinson J.A. A Machine Orientated Logic Based On The Resolution Priniple. J.A.C.M. Vol 12 No 1 (Jan 1965) :23–41

3. Snyder P.D. Modal Logic And Its Applications. Van Nostrad Reinhold, New York 1971

4. Hughes G.E. Cresswell M.J. Introduction To Modal Logic. Methuen, London 1968

5. Konolige K. A Deduction Model Of Belief. Pitman, London 1988

6. Allwood J. Anderson L. Dahl O. Logic In Linguistics. Cambridge University Press 1977

7. Robinson J.A. Logic: Form And Function. Edinburgh University Press 1979

8. Chang C.L. Lee R.C. Symbolic Logic And Mechanical Theorem Proving. Academic Press, New York 1973

9. Kowalski R. Logic For Problem Solving. North Holland, New York 1979

10. Stickel M.E. Automated Deduction By Theory Resolution. Proc. I.J.C.A.I. Los Angeles 1985

3.3. Modal Logic

3.3.1. Introduction

This section presents a sketch of the semantics of modal logic and of the axioms which specify the various systems of modal logic. Of necessity, much detail is omitted and no proofs are presented. All of this material can be found in the references [1–7]. Modal logic allows the distinction between what is necessarily true and what is possibly true to be made. Modal notions, including knowledge and belief and concepts related to time, can be given a formal analysis using modal operators which correspond to the necessity and possibility operators of modal logic.

3.3.2. Semantics

The formal definition of the semantics of modal logic is given in terms of a set of worlds U and an accessibility relation R. For each world, R defines which worlds are accessible, wiRwj states that wj is accessible from wi. Initially no conditions are imposed on R.

A world system is written $W = \langle U,R \rangle$. To define truth and validity for sentences in modal logic the truth value of each atomic sentence at each world $u \in U$ must be stated. For the case of propositional systems the set of worlds $Si \subseteq U$ at which the proposition Pi is true is specified. A modal structure π gives the truth values of atomic sentences in all worlds: $\pi = \langle U,R,S_0 \ldots Sn \ldots \rangle$. A world system may have many corresponding modal structures. π is said to be a modal structure on W or conversely W is the world system corresponding to π. To state that P is true at a world u in a modal structure π we write $\models_u^\pi P$.

The symbols L and M represent the necessity and possibility operators respectively. The truth definitions are given in terms of validity ($\models$).

1) $\models_u^\pi LP$ iff $\forall t(uRt \supset \models_t^\pi P)$

LP is true at u iff P is true at all accessible worlds

2) $\models_u^\pi MP$ iff $\exists t(uRt \land \models_t^\pi P)$

MP is true at u iff P is true at one or more accessible worlds [1,2]

The modal operators may be combined, in particular, sentences such as L...LP may be written as L^nP. For L^nP to be true at a world u P must be true at all worlds n– accessible from u. The n– accessible worlds are those which may be reached by n steps of the accessibility relation. A more detailed account of validity may be found in Lemmon [2].

Having defined the truth of modal sentences in terms of modal structures we now characterise the set of valid sentences syntactically. The propositional modal system K is specified by the axioms of propositional calculus A1, the rule of inference MP, modal axiom A2 and the rule of necessitation RN.

A1	a)	C P CQP	[$P\supset(Q\supset P)$]
	b)	C C P CQR C CPQ CPR	[$(P\supset(Q\supset R))\supset((P\supset Q)\supset(P\supset R))$]
	c)	C C CP⊥ ⊥ P	[$((P\supset\bot)\supset\bot)\supset P$]
MP		P, CPQ derive Q	[$P, P\supset Q$ derive Q]
A2		C LC P Q C LP LQ	[$L(P\supset Q)\supset(LP\supset LQ)$]
RN		from P derive LP	where P is an axiom

(sentences in Polish form are written to emphasise their structure, the equivalent sentences in standard form are enclosed by [])

The completeness theorem can be proven for K.

Theorem 3.3.1. Completeness of K. $\vdash_K P$ if and only if $\models P$. The set of sentences which may be derived by the axioms ($\vdash_K P$) is exactly the set of valid sentences ($\models P$).

By imposing conditions on the accessibility relation we derive different logical systems based on K, these are known as extensions of K. This is achieved by adding one or more of the following axioms to A1 and A2.

T:	C LP P	[$LP\supset P$]
4:	C LP LLP	[$LP\supset LLP$]
B:	C MLP P	[$MLP\supset P$]
D:	C LP MP	[$LP\supset MP$]
5:	C MP LMP	[$MP\supset LMP$]
G:	C MLP LMP	[$MLP\supset LMP$]

T states that R is reflexive i.e. that all worlds are accessible from themselves w_iRw_i. 4 states that R is transitive, if w_1Rw_2 and w_2Rw_3 then w_1Rw_3. In addition R may be a serial or symmetric relation or have a combination of these properties [3,6].

Quantification.

By associating a domain of individuals with each world a quantificational modal structure can be defined [4]. For each $u \in U$ a function ψ defines the set of individuals existing at u, $\psi(u)$. Symbols for predicates and individual variables are added to the logical language. In each world each n-adic predicate symbol has a set of n-tuples assigned to it which defines its extension in that world. To specify that in world h the predicate P(x) is true of some individuals in $\psi(h)$ and false of others we write $\varphi(P(x),h) = T$ or F respectively. The set of individuals of which P is true is called the extension of P in h. By convention P(x) is false if $x \notin \psi(h)$. With these semantics neither the Barcan formula (C∀xLP(x) L∀xP(x)) nor its converse need be theorems of the logic.

The quantified modal logic M is specified by adding the axioms A3 to A2 and A1.

A3	a)	CP ∀xP	where x is not free in P	[$P \supset \forall xP$]
	b)	C ∀xCPQ C ∀xP ∀xQ		[$\forall x(P \supset Q) \supset (\forall xP \supset \forall xQ)$]
	c)	∀yC ∀xP(x) P(y)		[$\forall y(\forall xP(x) \supset P(y))$]

Axioms may be added to obtain extensions of M as was done for K.

3.3.3 Mechanical techniques for modal theorem proving.

Human logicians often use tableau methods to prove theorems in modal logic [8,9]. These methods were not designed to have efficient computer implementations but efficient proof methods can be constructed using similar principles, for example the matrix proof method [9,10]. This method is capable of proving theorems in many modal systems. A formula is written as a tree structure and each subformula is labelled according to the world it exists in. Only formulae which exist in the same world can be compared. This method does not require formulae to be translated into a clausal form. A similar non-clausal proof method has been proposed [12,13]. In this system rules determine which pairs of formulae can be resolved. These rules are derived from the accessibility relation of the logic system under consideration.

Proof methods based on the resolution principle have been implemented by several researchers [14–18]. In the most recent work the syntax of a modal formula is modified by the addition of an index which identifies the world a formula exists in [17,18]. The correspondence between the quantifiers $\forall$, $\exists$ and the modal operators L, M can be exploited. A formula prefixed by the necessity operator is true in all worlds, a formula prefixed by $\forall x$ is true for all values of x. The modal operators can be eliminated in favour of a variable or a skolem function as is done for ordinary variables. The resolution procedure must unify the variables/skolem functions which indicate which world a literal exists in, in addition to matching the literals themselves [18].

References

1. Hughes G.E. Cresswell M.J. Introduction To Modal Logic. Methuen, London 1968

2. Lemmon E.J. An Introduction To Modal Logic. Basil Blackwell, Oxford 1977

3. Hughes G.E. Cresswell M.J. A Companion To Modal Logic. Methuen, London 1984

4. Kripke S.A. Semantical Considerations On Modal Logic. in Linsky L. (ed) Reference And Modality. Oxford University Press 1971 :63–72

5. Zeman J.J. Modal Logic. Oxford University Press 1973

6. Frost R.A. Introduction To Knowledge Base Systems. Collins, London 1986

7. Turner R. Logics For Artificial Intelligence. Ellis Horwood, Chichester 1984

8. Snyder P.D. Modal Logic And Its Applications. Van Nostrad Reinhold New York 1971

9. Fitting M. First Order Modal Tableaux. Journal Of Automated Reasoning 4, 1988 :191–213

10. Wallen L.A. Wilson G.V. Automated Deduction In S5 Modal Logic: Theory And Implementation. D.A.I.† Research Paper No. 315

11. Wallen L.A. Matrix Proof Methods For Modal Logics. D.A.I.† Research Paper No. 322

12. Jackson P. Reichgelt H. A General Proof Method For First Order Modal Logic. D.A.I.† Research Paper No. 339

13. Jackson P. Reichgelt H. A General Proof Method For Modal Predicate Logic. D.A.I.† Research Paper No. 370

14. Abadi M. Manna Z. Modal Theorem Proving. Proc. 8th International Conference On Automated Deduction 1986 :172–188 (Lecture Notes In Computer Science No.230, Springer Verlag)

15. Konolige K. Resolution And Quantified Epistemic Logics. Proc. 8th International Conference On Automated Deduction 1986 :199–208 (Lecture Notes In Computer Science No.230, Springer Verlag)

16. Farinas Del Cerro L. Resolution Modal Logic. Logique et Analyse No. 28 1985:153–172

17. Farinas Del Cerro L. Herzig A. Linear Modal Deductions. Proc. 9th International Conference On Automated Deduction 1988 :199–208 (Lecture Notes In Computer Science No.310, Springer Verlag)

† Department Of Artificial Intelligence, University Of Edinburgh

18. Ohlbach H.J. A Resolution Calculus For Modal Logics.
Proc. 9th International Conference On Automated Deduction
1988 :500– 516
(Lecture Notes In Computer Science No.310, Springer Verlag)

3.4 Belief Logic

3.4.1 The Semantics of Knowledge and Belief.

The formal system characterising knowledge and belief proposed by J.Hintikka [1] has been influential in much of the theoretical and computational work on belief systems. Hintikka provided a basic semantic definition of knowledge in terms of a modal operator and distinguished it from the concept of belief. He also gave an account of nested belief, introspection and the problem of combining quantifiers and belief operators. Subsequent study of the latter problems has shown certain weaknesses in his formulations. Hintikka's work provides a valuable starting point for a formal analysis of belief.

To establish criteria to judge the consistency of statements such as 'a knows that p', four epistemic operators were defined:

Ka 'a knows that '

Ba 'a believes that'

Pa 'it is possible, for all a knows that'

Ca 'it is compatible with everything a believes that'

The truth functional operators were defined in the usual way. The distinction between knowledge and belief is that a known sentence is true of the world, whereas this need not be the case for beliefs. Formally, a consistent set of sentences λ where $KaP \in \lambda$, remains consistent when P is added to λ. The operators Ka and Pa are dual: definitions H1 and H2 give the relationship.

H1) If λ is consistent and if $\neg KaP \in \lambda$ then $\lambda + \{Pa\neg P\}$ is consistent.

H2) If λ is consistent and if $\neg PaP \in \lambda$ then $\lambda + \{Ka\neg P\}$ is consistent.

The sentence PaP is true in a state of affairs μ if there is an alternative state of affairs μ^* where P is true (μ^* is an alternative to μ with respect to a). The consistency of Ka and Pa is determined by examining the alternative states of affairs, H3 and H4 give the definitions.

H3) If $KaP \in \mu$, and μ^* is an alternative to μ (with respect to a) then $P \in \mu^*$.

H4) If PaP ϵ μ, then there is at least one alternative to μ (with respect to a), μ^* such that P ϵ μ^*

Definition H1 connects negative belief with the truth value of a sentence in an alternative world. A method of proof by contradiction can be developed. If Pa¬P and KaP are true in μ then by H4 there is a state μ^* such that ¬P ϵ μ^* but as μ^* is an alternative to μ, P ϵ μ^* by applying H3 to KaP. The inconsistency of μ^* proves the inconsistency of Pa¬P and KaP in μ. This form of proof can be used to show KaKbP $\supset$ KaP i.e. knowledge is transmissable and that the corresponding formula for belief is not true. Hintikka accepts the following theorem:

H5) If KaP ϵ μ then KaP ϵ μ^*, where μ^* is an alternative to μ with respect to a.

In this theorem the alternativeness of the states of affairs is not well defined. The sentence Ba(P $\wedge$ ¬BaP) is shown to be false by making the implicit assumption that BaP implies BaBaP, which is a statement about intropsection. By accepting H5, BaP and BaBaP become equivalent, the introspective step in the derivation process is not made explicit. The first four definitions remain correct when Hintikkas' alternative states of affairs are replaced by the worlds or possible worlds models of modal logic.

3.4.2. Formal Models of Belief.

Many of the models of belief that have been proposed in recent years have been based on the worlds model [2–6]. An alternative approach is to add a predicate such as KNOWS() to first order logic. World models may be translated into an equivalent first order system, however the semantics for the worlds models are clearer and the syntax less complex.

The correspondence between a set of axioms for a modal logic of knowledge and the worlds model (based on Kripe semantics, see Section 3.3) has been demonstrated [2]. By imposing conditions on which worlds are accessible from which, belief may be distinguished from knowledge and types of introspection may be defined (sentences in a world representing known sentences must be true of the real world as before). This can be achieved by adding axioms to the logical system, the systems remaining sound and complete. The axioms include:

1. All tautologies of propositional calculus.

2. $(KiP \wedge Ki(P \supset Q)) \supset KiQ$ where i refers to the knower or agent

Axiom 3 states that what is known is true:

3. $KiP \supset P$

Axiom 4 defines positive introspection:

4. $KiP \supset KiKiP$

Axiom 5 defines negative introspection:

5. $\neg KiP \supset Ki\neg KiP$

Axiom 3–5 may be combined with 1 and 2 to give different logical systems corresponding to different models of knowledge, and are equivalent to the axioms of modal logic. It is implicit in these systems that all consequences of the agents beliefs are believed: the agent is said to be omniscient. This does not model human knowledge and more importantly it is not practical to compute all the consequences of a set of sentences.

One solution to this problem is to distinguish between implicit and explicit belief. All logical consequences of the explicit beliefs are included in the set of implicit beliefs. A logic of awareness can be derived from this approach where a sentence is believed, $Bi\psi$, if it can be derived from sentences which the agent is aware of, $Ai\alpha$ [3]. In contrast implicit belief, $Li\varphi$, includes all consequences of believed sentences with no restrictions. Omniscience is prevented by defining which sentences the agent is aware of at a given time. The authors do not show how this method could be implemented on a computer.

A first order logic for knowledge representation based on relevance logic has been proposed [5]. This logic has the property of being decidable. The logic does not include the implication operator and is weaker than standard first order logic (which is semi–decidable). The advantage of this logic is that the conclusions which can be drawn from a database can be made by fast computational methods and omniscience is avoided.

A worlds model was used by Moore [6] to formulate a theory combining knowledge and action. Knowledge and action interact, events change what is known, and knowledge is required before action is taken or planned. The worlds that are compatible with what an agent knows change after the (known) occurrence of an event. An event or action may produce new information, and an agent may require to reason about such situations. Moore based the semantics of his theory on a worlds model and then translated it into first order logic.

In summary, models of belief derived from modal logics provide the clearest semantics. A belief logic which is equivalent to a modal logic has the property of omniscience. Quantification into the context of belief is a problem which will now be examined.

3.4.3. The Quantifying-In Problem.

It has been argued that in certain contexts names referring to individuals do not occur referentially, that is a different name for the same individual cannot be substituted and the sentence remain true. These contexts, which include belief, are known as opaque. The conclusion is drawn that it is never correct to substitute or quantify into opaque contexts [7]. Two examples are often quoted:

1a) Philip believes that Tegucigalpa is in Nicaragua.

1b) Tegucigalpa = the capital of Honduras

2) Ralph believes that someone is a spy.

In 1a) if Tegucigalpa is replaced by 'the capital of Honduras' we get 'Philip believes that the capital of Honduras is in Nicaragua'. This is an example of a non-referential context. The derived sentence is true if Philip believes 1b), otherwise the substitution cannot be made. The problem is solved if the names given to objects that an agent believes to exist are defined as part of the agents beliefs. The names need not correspond to those of the real world.

The second example may be re-stated formally in two ways.

Let Br = Ralph believes, S be the property of being a spy

3a) $\exists x$ Br S(x) There is someone whom Ralph believes to be a spy

3b) Br $\exists x$ S(x) Ralph believes there are spies.

The interpretation of these sentences is quite different. When belief operators are combined with quantifers the order in which they appear determines the interpretation of the sentence. If 3a) is true then 3b) is also true, but not conversly.

Hintikka proposed that $\exists x$ Br S(x) may be derived from Br S(b) and $\exists x$ Br (b=x) but not from Br S(b) alone. The sentence $\exists x$ Br (b=x) means that Ralph knows who b is.

The following sentences are consistent,

4) Watson knows that Mr Hyde is a murderer.

5) Watson does not know that Dr Jekyll is a murderer.

when in fact Mr Hyde = Dr Jekyll but Watson does not know this. Hintikka rejects the Barcan formula ($\forall x$ Ka p(x) $\supset$ Ka $\forall x$ p(x)) but accepts its converse as being necessarily true (agents can only 'know' about actually existing individuals). Another solution to the quantifying-in problem is to define the relationship between the names of objects in the real world and in an agent's beliefs by a naming function, this solution is adopted in the deduction model.

3.4.4. The Deduction Model

The model of belief adopted in the design of the expert system (Sections 4.2, 4.3) is based on the deduction model of belief. This model, its logic and proof theory have been developed by Konolige in a series of papers [8-11]. The model is described in detail in this section. The tableaux and resolution proof procedures are presented to illustrate how deductions in a belief system can be carried out and automated. The correspondence between the deduction model and the systems of modal logic is discussed in section 4.2.

Typically a belief system will be composed of a base set of sentences, a set of inference rules and a control strategy. In order to answer a query deductions are made by applying the inference rules to the base set. How this is done in practice is determined by the control strategy. The relationship between the three components is shown in Figure 3.4.1.

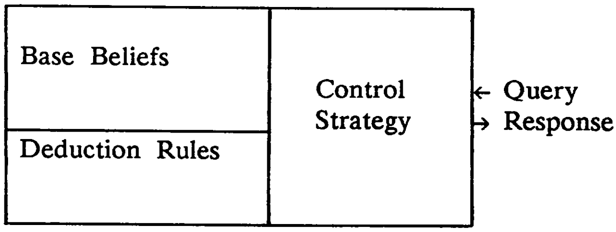


Figure 3.4.1. The belief system.

In this model belief is interpreted as follows:

$\varphi \in$ bel(Ralph)	'Ralph believes φ'
$\neg\varphi \in$ bel(Ralph)	'Ralph believes $\neg\varphi$'
$\varphi \notin$ bel(Ralph)	'Ralph does not believe φ'

The language in which Ralph's beliefs are expressed in is called his internal language L. φ is an expression in L. A belief system may have partial information about the world, for example if neither γ nor $\neg\gamma$ is a member. The belief set is contradictory if both γ and $\neg\gamma$ are members. To avoid having to formalise a complex inference process in order to predict its behaviour it is assumed that the inference rules are deduction rules and the control strategy applies these rules to the base set to generate all consequences. No distinction is drawn between a base sentence and a derived sentence: both may participate in further deductions. The belief set is said to be closed under deduction. Which inferences are actually made is determined by the deduction rules, therefore not all logically true sentences are inferred (the deduction rules will usually be logically incomplete). As a result the deduction model does not have the property of omniscience. The deduction rules must have fixed and finite premises which are effectively computable. That the control strategy applies the deduction rules exhaustively can be stated formally as the deductive closure property:

B is the base set of beliefs of an agent

If $B \vdash_R \varphi$ and $B, \varphi \vdash_R \psi$ then $B \vdash_R \psi$

where $B \vdash_R \varphi$ means that there is a proof of φ from the premises B using rules R.

The deduction structure is the formal mathematical model of the belief system, written $<B,R>$. The set B is a set of sentences, the base set of beliefs, in a logical language L, the language of belief. R is the set of deduction rules.

$$bel(<B,R>) = \text{ by definition } \{\varphi \mid B \vdash_R \varphi\}$$

The Language LB

To maintain a consistent notation throughout this thesis Konolige's language LB will be expressed in Polish notation. A modal construct $[S]\varphi$ is introduced to represent belief. The sentence φ is expressed in the internal language and S denotes the agent.

Definition 3.4.1. The formal definition of LB. Let $\{S_0,S_1...\}$ be a countable set of agents and L the internal language of the agent under consideration. A sentence of LB, based on L, is defined by the rules:

1) LB includes the sentences and formation rules of first order logic.

2) If φ is a sentence of L then $[Si]\varphi$ is a sentence of LB.

If L is a first order language $[Si]\exists x\ P(x)$ is a sentence of LB. To allow beliefs to be nested the internal language L must be replaced by LB. Both the internal language and the deduction rules are fixed and become parameters in the interpretation of LB. The interpretation is called a $B(L,\rho)$-model where $\rho(i)$ defines the deduction rules for the agent Si. The base set of sentences for Si and $\rho(i)$ make up the deduction structure $di=<B,\rho(i)>$.

To connect derivations in the internal language with the truth value of sentences in the external language the attachment lemma is required. Definition 3.4.2. states the lemma, a proof can be found in Konolige [8].

Definition 3.4.2. The attachment lemma. The denumerable set $\{[Si]\Gamma,\ N[Si]\Delta\}$ is $B(L,\rho)$ unsatisfiable if and only if for some $\sigma \in \Delta$, $\Gamma \vdash_{\rho(i)} \sigma$.

For example let $\rho(i)=\{C\ P(a)\ Q(a)\}$ if $G_0=\{N[Si]Q(a),\ [Si]P(a)\}$ then $Q(a) \in \Delta$ and $\{P(a)\} \vdash_{\rho(i)} Q(a)$ so Q(a) is not believed but may be derived so G_0 is unsatisfiable.

Quantifying– in and the deduction model.

The model of the first order language IP was defined by a set of individuals D and a value assignment V, written $<D,V>$. To allow for quantification in the context of belief this model structure is modified. The formation rules of definitions 3.2.4– 3.2.6 (Section 3.2.2) are retained. The components of the new model are defined:

Let U be a universe of individuals

φ is a mapping from constants to individuals in U

ν_0 is an assignment of truth values to all ground atoms containing only elements of U.

Definition 3.4.3. Let $m=<\varphi,\nu_0,U>$ be a model structure. By $\models_m P$ we mean that m assigns the sentence P the value true. By P^{φ} we mean the sentence P with all of its constants replaced by the corresponding elements of U. $\models$ satisfies the rules:

1. If P is a ground atom $\models_m P$ iff $\nu_0(P^{\varphi})=$ true
2. $\models_m K\ P\ Q$ iff $\models_m P$ and $\models_m Q$

   $\models_m A\ P\ Q$ iff $\models_m P$ or $\models_m Q$

   $\models_m C\ P\ Q$ iff $\models_m P$ or $\models_m Q$

   $\models_m NP$ iff $\not\models_m P$
3. $\models_m \exists x\ P$ iff for some $k \in U$, $\models_m P_k^x$

   (every x in P is replaced by k)
4. $\models_m \forall x\ P$ iff for all $k \in U$, $\models_m P_k^x$.

This definition of a quantified logic does not assume that names and individuals are equivalent. Before examining the quantifying– in problem the functions φ and η are set up.

Letting φ be the denotation function, η the naming function, they have the following relationship:

φ(mayor(NYC)) = David Dinkins :the individual

η(David Dinkins) = David–Dinkins :the name

$\neq$ mayor(NYC)

we have $\varphi(\eta i(k)) = k$ the denotation of a name (in the belief system of i) is the individual

but $\eta i(\varphi(a)) \neq a$

For $\exists x$ [S] P(x) to be equivalent to [S] P(c) (a sentence where the existential quantifier has been eliminated) the constant c must refer to the same individual that x does. Such constants are called id constants. If the quantifed sentence refers to k ϵ U then [S] P(k) is true and [S] P(c) is true when c = $\eta_S(k)$, that is, when c is the name in the belief system of [S] of the individual k.

The quantified language of belief LBQ can now be formally specified by Definition 3.4.1. with the addition of the bullet operator $\bullet$ to the language: $\bullet x$ is a term if x is a term. This operator is applied to constants substituted into the context of belief.

$\exists x$ [S] P(x) has the skolem analog [S] P($\bullet$a)

[S]$\exists x$ P(x) has the skolem analog [S] P(a)

$\bullet$a is equivalent to an id constant c,

$c = \eta_S(\varphi(a)) = \eta_S(k)$ *

The naming function η must be defined for each agent. It then becomes a parameter of the model.This method allows sentences in BLQ to be skolemized correctly. The quantified B(L,ρ) model is specified by the tuple $<\varphi, \nu_0, U, D, \eta>$. The first three elements of this tuple define the first order model structure (Definition 3.4.3.), D is the set of deduction structures $D=\{d_0, d_1, \ldots\}$ where $d_i = <B, \rho(i)>$ and $\eta = \{\eta_0, \eta_1, \ldots\}$. The deduction structure (the set of base beliefs, B, deduction rules, $\rho(i)$, in language L) and the naming map, η_i, is defined for each agent.

*The naming function is used to solve the quantifying–in problem by associating a name in the beliefs of an agent with an individual in the real world.

Analytic Tableaux

Analytic tableaux are useful for demonstrating proof techniques. The following example is typical of their use. The rules which define correct inferences in the tableau will be introduced as required. A tableau proof of the sentence C [S] C P Q C [S]P [S]Q can be constructed as follows, where T and F stand for true and false. In the external language,

1)	F C[S]C P Q C[S]P [S]Q)	Negate the initial sentence.	
2)	T[S]C P Q	1)	Apply the $C\alpha\beta$ reduction rule
3)	F C [S]P [S]Q	1)	to get 2) and 3) from 1).
4)	T[S]P	3)	Apply the $C\alpha\beta$ reduction rule
5)	F[S]Q	3)	to get 4) and 5) from 3).
*	P, C P Q $\vdash_R$ Q	2),4),5)	

Q can be deduced in [S]'s beliefs, that is, [S]Q is true, this contradicts line 5) therefore the tableau closes.

As the tableau closes the negation of the initial sentence must be false, hence the initial sentence must be true. This example uses the tableau rule that the sentence F $C\alpha\beta$ may be replaced by $T\alpha$ and $F\beta$, this rule is derived from truth tables. If the sentence $C\alpha\beta$ is true then the tableau splits and $F\alpha$, $T\beta$ head the branches.

A contradiction has been derived within [S]'s beliefs, namely Q and NQ. The tableau method can also be used to represent the deduction of Q in the belief system. This is done by replacing the final line of the above proof by the following tableau which represents [S]'s view of the world.

6)	FQ				5)	A negative belief sentence.
7)	TP				4)	Add positive belief sentences.
8)	T C P Q				2)	Apply the T $C\alpha\beta$ rule, the
9)	FP	8)	10)	TQ	8)	tableau splits into two.
	*			*		Both branches close.

This tableau was constructed by adding TP when T[S]P is a formula in the main tableau. The tableau begins with a negative belief sentence as by the attachment rule when F[S]Q is added to a branch of the main tableau we know that the branch will close if Q can be shown to be a belief of [S]. The lines 6)−10) above are a proof of T[S]Q by refutation and such a tableau is called a 'view' for agent S.

The splitting which occurs in analytic tableau becomes a problem when such proofs are automated. It is now shown that this problem can be overcome by a resolution method. With the concept of attachment and the use of views for agents, a resolution rule for belief sentences can be derived.

B– Resolution

The resolution scheme proposed by Konolige adds the B–resolution rule to the binary resolution rule. Sentences in LBQ must be converted into clause form. The conversion process is the same as for first order logic with the addition that the bullet operator must precede quantified–in variables. Sentences in the scope of belief remain in Polish form. It can be proved that a set of sentences in LBQ is unsatisfiable when its skolem transform is unsatisfiable. Herbrands theorem holds for LBQ, that is, a set of sentences in LBQ in skolem normal form is unsatisfiable if and only if a finite set of its ground instances is unsatisfiable. The B–resolution rule is an instance of the total narrow theory resolution rule (Section 3.2.7).

The B–resolution rule

$$\begin{array}{ll} [S_i] & \Pi_1, A_1 \\ [S_i] & \Pi_2, A_2 \\ \ldots & \\ [S_i] & \Pi_n, A_n \\ N[S_i] & \psi, A \\ \hline A\theta, A_1\theta, \ldots A_n\theta & \text{when } \Pi_i, \ldots \Pi_n \vdash_{p(c)} \psi \end{array}$$

This rule uses the attachment lemma as it links unsatisfiability in the internal and external languages. To allow for free variables in the derivation process the bullet transform is replaced by a schematic function, $g(\tau)$ replaces $\bullet\tau$. The derivation operator may be replaced by a resolution based refutation proof as was done for the tableaux method. Both binary and B–resolution rules are used in the following example.

1)	{[R]P(a)}		
2)	{NP(b)}		
3)	{Q(x), P(x), [R]P(•x)}		
4)	{N[R]K P(a) P(•y), Q(y)}		
5)	{NQ(b)}		

the following clauses may be deduced

6)	{Q(b), [R]P(•b)}	resolve 2) and 3)

as 4) contains a negative belief literal, open a view in an attempt to resolve it

View R Remainder Q(y)

a)	{NP(a), NP(g(y))}	from 4) (see below)
b)	{P(a)}	add 1)
c)	{NP(g(y))}	resolve a) and b)
d)	{P(g(b))}	add 6) add Q(b) to the remainder list
e)	□	resolve c) and d)
7)	{Q(b)}	add the remainder of the view, under the substitution {y/b}
8)	□	resolve 5) and 7)

Views may be opened by a negative belief literal, positive beliefs may be added and resolutions may be performed in a view. The example shows the procedure for resolution in the quantified language of belief. The decision to open up a view appears at an arbitary point in the proof, if there are several negative beliefs several views may be opened. Konolige envisages that the activity of the system be interspersed among many different refutations. This problem has been termed non-effectiveness [12] and is a problem with the practical implementation of B-resolution. A solution is presented in section 4.2.

References

1. Hintikka J. Knowledge And Belief: An Introduction To The Logic Of The Two Notions. Cornell University Press, New York 1962

2. Halpern J.Y. Moses Y. A Guide To The Modal Logics Of Knowledge And Belief. Proc. I.J.C.A.I. 1985 :480–490

3. Fagin R. Halpern J.Y. Belief, Awareness And Limited Reasoning: Preliminary Report Proc. I.J.C.A.I. 1985 :491–501

4. Vardi M.Y. A Model Theoretic Analysis Of Monotonic Knowledge. Proc. I.J.C.A.I. 1985 :509–512

5. Patel–Schneider P.F. A Decidable First Order Logic For Knowledge Representation. Proc. I.J.C.A.I. 1985 :455–458

6. Moore R.C. A Formal Theory Of Knowledge And Action. in Hobbs J.R. and Moore R.C. (eds) Formal Theories Of The Commonsense World Ablex Publishing Corporation New Jersey 1985 :322–357

7. Quine V.W.O. Reference And Modality. in Linsky L. (ed) Reference And Modality. Oxford University Press 1971 :17–34

8. Konolige K. A Deduction Model Of Belief. Pitman, London 1988

9. Konolige K. Belief And Incompleteness. S.R.I. International Technical Note No. 319 Menlo Park C.A. 1984

10. Konolige K. A Computational Theory Of Belief Introspection. Proc. I.J.C.A.I. 1985 :502–508

11. Konolige K. A Resolution Method For The Quantified Modal Logics Of Knowledge And Belief. Theoretical Aspects Of Reasoning About Knowledge, Proceedings of the 1986 Conference Morgan Kaufman 1986 :309–324

12. Wallen L. Matrix Proof Methods For Modal Logics. Department Of Artificial Intelligence, University Of Edinburgh, Research Paper No. 322

3.5 Reasoning With Incomplete Information

3.5.1. Introduction.

All deductions that can be made from a set of sentences in standard first order logic using the inference rule MP are valid. The inference procedure is sound. The consequences of a set of sentences S_0 remain true when the set S_0 is expanded by the addition of new sentences. Deduced sentences may not be retracted upon learning new information. This property is known as monotonicity.

If our knowledge is incomplete then conclusions must be drawn from incomplete information else nothing can be concluded. This is not an unusual situation in human reasoning but one in which sound, monotonic deductions cannot be made. Logic systems which formalise reasoning with incomplete information do so by making assumptions to fill the gaps in the information. The assumption may be about what is usually true of the world or about what is or is not known: these patterns of reasoning are known as default and autoepistemic reasoning respectively. The approaches are interdependent and in many logic systems they are combined.

Monotonicity may be defined in the following way, where S_0 and S_1 are sets of sentences in a logical language:

$$\text{If} \quad S_0 \subseteq S_1 \text{ then } \{W \mid S_0 \vdash W\} \subseteq \{W \mid S_1 \vdash W\}$$

Logics where this condition is not met are called nonmonotonic [1]. Such logics are required in making inferences from incomplete information, as an assumed sentence Wi, and its consequences Wj,Wk may be withdrawn on learning NWi.

The most common example of nonmonotonic reasoning is the use of the closed world assumption. Prolog uses this principle to derive NP when every possible proof of P fails. Further facts can be added to a Prolog program which may enable P to be derived. The first nonmonotonic logics addressed the problem of using default information, these were Default Logic [2] Circumscription [3,4] and Non Monotonic Logic [5]. Autoepistemic logic [6–8] was an improvement of Non Monotonic Logic and several related logics have been developed [9,10]. The following problems are typical of those addressed in the literature.

Problem 1. If most birds can fly and Tweety is a bird deduce that Tweety can fly, unless Tweety is a penguin or an ostrich.

Problem 2. Consider the sentences MC $\supset$ ND, MD $\supset$ NC where M is a modal operator interpreted as consistency. There are two distinct models for these sentences, one where D is false and C is true and a second where C is false and D is true. A model may be defined by associating truth values with propositions to give what is known as a set of fixed points. In this problem there are two distinct sets of fixed points. Two problems arise, the first is to expand the set of fixed points and retain consistency, the second is how to choose between the competing models.

In the following survey of nonmonotonic logics we present an outline of several important logics and discuss the issue of automation.

3.5.2. Closed World Reasoning.

Under the closed world assumption (CWA) negative information need not be explicitly represented. The amount of negative information increases geometically with the size of the Herbrand universe [1] so it is impratical to represent all negative literals in a knowledge base. The CWA can be formally stated:

If KB $\nvdash$ P then infer NP

This is a meta rule which states that NP may be inferred when P cannot be inferred from a knowledge base KB. The meta rule is applied uniformly to all predicates. The closed world assumption is realised by determining the naive closure of KB defined as follows: first the negative extension of KB is defined:

$$\begin{aligned} \text{EKB} = \{\text{NP}(\alpha) \mid\ & \text{P is an n-ary predicate} \\ & \alpha \text{ is an n-tuple of ground terms} \\ & \text{and KB} \nvdash \text{P}(\alpha)\} \end{aligned}$$

The naive closure of KB is KB $\cup$ EKB: the knowledge base is extended by adding a set of ground literals. In practice the derivation of NP is often shown by the 'failure to derive' P strategy.

Predicate circumscription is a means of closing off the world for a particular predicate. The n-tuples which satisfy a predicate are assumed to be the only ones which do so, on the basis of the currently known facts. This method allows explicit completeness assumptions to be conjectured. McCarthy describes circumscription as a formalised rule of conjecture that can be used along with the rules of inference of first order logic [3].

McCarthy has presented a new version of his theory, named formula circumscription [4]. This method can be viewed as minimising the number of abnormal facts through performing circumscription. In problem 1 above the abnormal case would be where Tweety could not fly. For a more complex version of this problem, formula circumscription gives the following completeness result: $\forall x$ flies(x) $\equiv$ K bird(x) N ostrich(x) (everything which flies is a bird and not an ostrich). Which predicates are abnormal is explicitly defined in first order logic. The method is unintuitive and its implementation required a second order theorem proving program.

3.5.3. Non Standard Logics.

The motivation for the Default Logic of Reiter [2] is the need to make default assumptions about incompletely specified worlds. The operator M is defined as 'it is consistent to assume' but does not have the status of a modal operator. The following default rule states that birds can fly if it is consistent to assume so.

$$\frac{\mathrm{bird}(x) :\ M\ \mathrm{fly}(x)}{\mathrm{fly}(x)}$$

The closed world default rule for the predicate R can be defined:

$$\frac{:M\ \mathrm{NR}(x)}{\mathrm{NR}(x)}$$

The aim is to extend a first order theory by adding new literals, the result is a set of fixed points (or extension) which characterises the extended theory. There may be several or no sets of fixed points for a set of sentences which include default rules.

Reiter presents a resolution inference procedure which maintains a consistent set of clauses during backwards chaining. The procedure works within one extension of the theory in an attempt to prove a goal sentence. The advantages of this method are use of explicit default rules which have clear semantics and the existence of an automated inference procedure.

Default reasoning provides plausible grounds for certain beliefs, in contrast Autoepistemic Logic makes valid inferences from sentences which may refer to everything an agent believes. The operator L is added to propositional logic. The sentence LP is true if P is in the set of beliefs. Sentences such LP and NLP are context sensitive or 'indexical' because they refer to the entire set of beliefs [6–8] (in the belief logic of Konolige the belief operator does not have this interpretation). Autoepistemic reasoning is a theory of introspection and as such may be used to make default deductions (eg. the sentence C K P LQ Q). There may be several extensions of an autoepistemic theory as is the case for default theories. The extensions can be found by enumeration but this is inefficient [8]. No first order version of this logic has been presented.

3.5.4. Discussion.

The semantics of nonmonotonic logic are nonconstructive [6], that is, the conditions of a sentence may refer to the whole of what is and is not believed and are not simply a composition of conditions which once satisfied remain true. In Autoepistemic Logic, because what is provable is dependent on what is and is not provable there is a degree of circularity about autoepistemic sentences. The nonconstructive nature of nonmonotonic inference gives rise to difficulties in finding efficient theorem proving methods. The proof procedure has been shown not even to be semi decidable (first order logic is semi decidable)[11].

Nonconstructive semantics capture an important part of commonsense reasoning, where default or introspective reasoning may be used to reach tentative conclusions. When these inference patterns are defined in a formal logic the existence of multiple models and difficulties in automation follow.

The use of a belief operator is important as it is then possible to refer to what is or is not believed whereas such statements cannot be made in first order logic. Where no belief operator is explicitly defined, a theory of what is derivable or believed is often implicit in the default theory.

To enable an expert system to make use of default reasoning the rule base would have to be redesigned and expressed in Default Logic. A major consideration would be to correctly specify the default inferences. The use of Autoepistemic Logic would require a similar approach (assuming a first order version of this theory could be defined). An alternative theory of introspection which makes use of the belief logic of Konolige and may be applied to an expert system rule base is presented in Section 4.4.

References

1. Etherington D.W. Reasoning With Incomplete Information. Pitman London 1988

2. Reiter R. A Logic For Default Reasoning. A.I. 13 1980 :81–132

3. McCarthy J. Circumscription – A Form Of Non Monotonic Reasoning A.I. 13 1980 :27–39

4. McCarthy J. Applications Of Circumscription To Formalising Commonsense Knowledge. Proceedings of the Non Monotonic Reasoning Workshop, New Paltz, New York, A.A.A.I. 1984 :295–324

5. McDermott D. Doyle J. Non Monotonic Logic I. A.I. 13 :41–72

6. Moore R.C. Possible World Semantics For Autoepestemic Logic. Proceedings of the Non Monotonic Reasoning Workshop, New Paltz, New York, A.A.A.I. 1984 :344–354

7. Moore R.C. Semantical Considerations On Non Monotonic Logic. A.I. 25 1985 :75–94

8. Moore R.C. Autoepistemic Logic. in Smets P. et al (eds) Non–Standard Logics For Automated Reasoning. Academic Press, London 1988 :105–136

9. Halpern J.Y. Moses Y. Towards A Theory Of Knowledge And Ignorance. Proceedings of the Non Monotonic Reasoning Workshop, New Paltz, New York, A.A.A.I. 1984 :125–143

10. Konolige K. A Computational Theory Of Belief Introspection. Proc I.J.C.A.I. 1985 :502–508

11. Reinfrank M. (ed) Proceedings of the 2nd International Workshop On Non Monotonic Reasoning June 1988 :vii–xiv

Chapter 4 New Models For Expert System Design

4.1 Introduction

Expert systems are computer programs which apply the knowledge of a human expert to solve problems in a specific field. The knowledge they contain is usually expressed in symbols, in contrast with the more conventional programming methods designed to carry out numerical calculations.

Expert knowledge is often expressed as rules (production rules or rules in first order logic) in the format IF (condition) THEN (action). Expert rules may be used to prove a specific goal. For example the Mycin system, designed to diagnose a certain class of blood disorders, develops a line of reasoning by the backward chaining of rules [1,2]. Expert knowledge may also be used in conjunction with other programming methods. The Dendral system which was designed to identify the structure of complex molecules employed a specialised routine to generate a set of plausible candidate molecular structures [3,4]. The molecular structure which most closely fitted the mass spectrogram data was found by a set of expert rules. This technique is known as 'generate and test'. Expert knowledge was applied in both the generation and assessment of molecular structures in the Dendral system.

Rules may be used directly to derive conclusions, to guide the inference procedure or as part of a problem solving strategy. A probability or a certainty factor is often associated with a rule. Certainty factors are propagated through the inference net. These techniques allow inexact reasoning to be carried out.

It is common for expert systems to be able to explain their reasoning by demonstrating how a conclusion was deduced. A natural language interface can improve communication between the user and the computer. These features can help with the acceptance of expert systems by those unfamiliar with A.I. programming. Expert system shells can provide a range of facilities which allow the user to concentrate on the task of acquiring and formalising the relevant knowledge.

The application of expert knowledge is central to the success of expert systems. The representation and manipulation of expert knowledge is important in the design of expert systems. The use of first order logic has been proposed because of its well understood model theory and proof methods [5]. Semantic nets [6], frames and production rules have been used to express knowledge, however the semantics and proof theory of these systems relies heavily on intuition.

The improvement of expert system design requires an improved representation of the expert knowledge. First order logic has the clearest semantics of the simple representation schemes but is not rich enough for many purposes. First order logic may be extended in one of three ways: a modal operator may be defined to give a modal system [Section 3.3], predicates may be allowed to range over both predicates and variables to give a higher order logic, or a type theory may be employed [7].

This chapter investigates the use of belief logic for the representation of expert knowledge. The belief logic employed is the deduction model of belief developed by Konolige [8]. This logic corresponds to the K system of modal logic [Section 3.3.]. There are other modal systems which may be used to represent belief. The K45 system and the corresponding belief logic are also investigated.

The model theory (the theory used to evaluate the truth of a logical sentence) of belief logic is different from the model theory of f.o.l. The logical language and model theory employed by an expert system constrain the range of concepts that can be represented and manipulated. The expert systems developed in this chapter are based on belief logic which is capable of expressing modal concepts. Expert systems based on f.o.l. cannot express modal concepts.

Because of the critical role of the logical language in representing knowledge in expert systems, the author concludes that the use of new languages and model theories constitutes new models for expert system design.

References

1. Davis R. Buchanan B.G. Shortliffe E.H. Production Rules As A Representation For A Knowledge Based Consultation Program. Artificial Intelligence 8, 1977 :15–45

2. Figenbaum E.A. Themes And Case Studies Of Knowledge Engineering. in Michie D. (Ed) Expert Systems In The Microelectronic Age. Edinburgh University Press 1979 :3–25

3. Feigenbaum E.A. Buchanan B.G. Lederberg J. On Generality And Problem Solving: A Case Study Using The Dendral Program. Machine Intelligence 6. Edinburgh University Press 1971

4. Brachman J. et al What Are Expert Systems in Hayes–Roth F. Waterman D.A. Lenat D.B. Building Expert Systems. Addison Wesley, London 1983

5. Hayes P.J. The Naive Physics Manifesto. in Michie D. (Ed) Expert Systems In The Microelectronic Age. Edinburgh University Press 1979 :242–270

6. Brachman R.J. Whats In A Concept: Structural Foundations For Semantic Networks. I.J.M.M.S.(1979) 9:127–152

7. Frost R.A. Introduction To Knowlegde Based Systems. Collins, London 1987

8. Konolige K. The Deduction Model Of Belief. Pitman London 1988

4.2 Resolution Methods For The Deduction Model

4.2.1. Introduction

The deduction model of belief [1,2] is developed by deriving two effective resolution methods for it. To show how this is done we discuss the relationship between the belief logic IB and the corresponding modal system K. This gives the sound semantics for negative or false belief sentences which is needed to re-examine the way views are opened for agents. Before presenting the resolution rules the language of belief is formally defined and the procedure for deriving the skolem transform is described and analysed. Algorithms for skolem conversion and resolution are presented.

4.2.1. The language IB.

Definition 4.2.1. The propositional language IB* is composed of the following symbols.

1. A set of statement letters (or atomic formulae) P,Q...
2. A set of agents B_0, B_1, B_2 ...
3. A set of truth functional operators A,K,C,N.
4. The symbols '<' and '>'.

Definition 4.2.2. A well formed formula is defined by the formation rules.

FR1. An atomic formula is well formed.

FR2. If α and β are well formed formulae and B is an agent then

$A\alpha\beta$, $C\alpha\beta$, $K\alpha\beta$, $N\alpha$, $\langle B\rangle\alpha$ are well formed.

FR3. A formula is well formed iff it can be constructed by the above rules.

Ordinary formulae are those which do not include the belief operator $\langle B_0\rangle$ and are given the standard interpretation. Sentences of the form $\langle B_0\rangle P$ are interpreted as 'B_0 believes P' or 'P is in the belief set of B_0'. We adopt Konolige's interpretation of belief as outlined in Section 3.4.4.[1,2] Employing the deduction model of belief means that inferences in the belief set of B_0 are made using the deduction rules ρ only. We have the deductive closure property:

$$\mathrm{bel}(\langle B,\rho\rangle) = \{\varphi : B \vdash_\rho \varphi\} \qquad \text{(where B is the base set of beliefs of } B_0)$$

*The language IB is the same as Konolige's language LB except for a modification to the syntax.

4.2.3. An alternative semantics for the deduction model.

The modal system K corresponds to the logical system IB when one agent is considered and the deduction rules are complete (the deduction rules enable all logically true inferences to be made). The following pairs of sentences are equivalent under these assumptions.

logical system:	IB	K
sentence 1.	$<B_0>P$	LP
sentence 2.	$N<B_0>P$	MNP

Sentences 1. and 2. taken together are not satisfiable in K. Proof: in all worlds accessible from w_0 P is true by 1. but in at least one of those worlds P is false by 2. therefore there is no model satisfying both 1. and 2. (P cannot be both true and false in an accessible world).

In the deduction model of IB sentences 1. and 2. are not consistent either. To prove $N<B_0>P$ to be inconsistent with a set of sentences S a view is opened for B_0 and a refutation proof for NP in that view must be obtained. This proves $<B_0>NP$ is false, therefore $<B_0>P$ is true, contradicting 1. In this example sentences 1. and 2. are immediately contradictory but usually sentence 1. would be derived by the opening of a view as described.

In the model of belief as presented by Konolige negative belief eg $N<B_0>\alpha$ is not explicitly interpreted as a commitment to the truth of $N\alpha$ in some belief world. Konolige sees $<B_0>\alpha$ and $N<B_0>\alpha$ as contradictory sentences about B_0's beliefs. We extend this interpretation to state that there may be several belief worlds compatible with what B_0 believes, in one of which α is false by the latter sentence but true in all belief worlds by the former, demonstrating a contradiction. The belief worlds correspond to the worlds of modal logic but are parameterised by the same language and deduction rules as in the deduction model. In future we refer to views in the sense that they represent deduction in a belief world. The proposed new structure for belief worlds is illustrated in Figure 4.2.1.

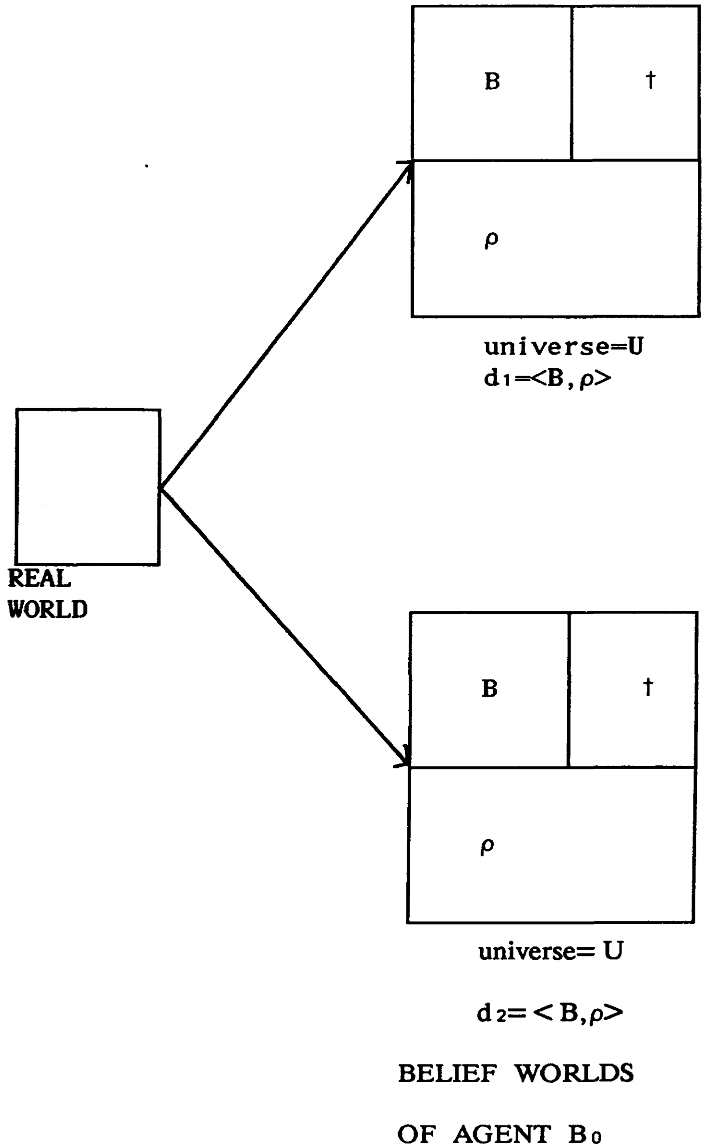


Figure 4.2.1. Belief worlds and model structure for agent B_0.

The tuple $<\varphi,\nu_0,U,D,\eta>$ specifies the model of belief where $D=\{d_1,d_2,\ldots d_n\}$ such that $d_1\ldots d_n$ are the deduction structures for agents $1\ldots n$ and φ,ν_0,U,η define the first order model theory [Section 3.4.4.]

† sentences which are not members of B

The combination of negative belief.

The question of whether two negative belief sentences, for example $N<B_0>P$ and $N<B_0>Q$, may be combined to give a single sentence $N<B_0>R$ arises. The conjunction of the former sentences in K is K MNP MNQ. It is incorrect to move the conjunction operator K to within the scope of M to derive M K NP NQ as stating that P is false in a world w_1 (w_0Rw_1) and Q is false in w_2 (w_0Rw_2) does not imply that P and Q are false in the same world. In general negative belief sentences may not be combined. Consequently in the deduction model only one negative belief sentence is allowed in a view (in both interpretations of a view).

4.2.4 Conversion to Clausal Form.

A view must always contain one sentence resulting from a negative belief sentence. From $N<B>\alpha$ the sentence $N\alpha$ is added to the view. In carrying out deduction by resolution all sentences must converted into clausal form, from $N<B>\alpha$ the clausal form of $N\alpha$ will appear in the view. This gives two additional rules for conversion to clausal form.

Rule S1 $\quad \mathrm{Sk}(<B>\alpha) = \mathit{<B>}[\mathrm{Sk}(\alpha)]$

Rule S2 $\quad \mathrm{Sk}(N<B>\alpha) = \mathit{N<B>N}[\mathrm{Sk}(N\alpha)]$

where $\mathrm{Sk}(\alpha)$ returns the clausal form of α
(clausal form is indicated by the use of italics)

The sequence *N<B>N[]* in clausal form denotes that the original sentence was a negative belief sentence (*N<B>N* may be thought of as the possibility operator). The normal form for sentences in IB can be simplified by the following rules.

Rule IB1 A sentence of the form $<B>\alpha$ is equivalent to

$<B>$ K A $P_1 \ldots P_m$ ... A $Q_1 \ldots Q_n$ (4.2.1.)

where K A $P_1 \ldots P_m$ ... A $Q_1 \ldots Q_n$ is the conjunctive normal form of α,

which is in turn equivalent to

K $<B>$ A $P_1 \ldots P_m$ ... $<B>$ A $Q_1 \ldots Q_n$ (4.2.2)

where there are only disjunctions of atomic formula within the scope of the belief operators.

Rule IB2 A sentence of the form $N\langle B\rangle\alpha$ is equivalent to

$N\langle B\rangle N\ K\ A\ P_1 \ldots P_m \ldots A\ Q_1 \ldots Q_n$ (4.2.3.)

where $K\ A\ P_1 \ldots P_m \ldots A\ Q_1 \ldots Q_n$ is the conjunctive normal form of $N\alpha$,

which in turn implies

$K\ N\langle B\rangle N\ A\ P_1 \ldots P_m \ldots N\langle B\rangle N\ A\ Q_1 \ldots Q_n$ (4.2.4)

Where there are only disjunctions of atomic formulae within the scope of the sequence $N\langle B\rangle N$, this sequence is equivalent to the possibility operator M in K. The sequence is replaced by '$\underline{N\langle B\rangle N}$' when writing sentences in clausal form.

As equation 4.2.4 is derived from but not equivalent to equation 4.2.3 the conjunctive normal form of IB, the set of conjunctions S_1, is not equivalent to the original set S_0 of sentences in IB. However if S_1 is inconsistent then so is S_0.

Proof: If S_1 is inconsistent: $S_1 \vdash P$ and $S_1 \vdash NP$ for some P

By rules IB1 and IB2 and the logical equivalences

$S_0 \vdash S_1$ therefore $S_0 \vdash P$ and $S_0 \vdash NP$ i.e. S_0 is inconsistent.

If there is only one disjunctive term in equations 4.2.3. and 4.2.4. then S_0 and S_1 are equivalent. Restricting α in rule IB2 to atomic sentences also restores the equivalence of S_0 and S_1. Alternatively if each negative belief term is indexed then one or more clauses with the same index may appear in a view and equivalence is restored.

4.2.5. Algorithms for Skolem Conversion.

Conversion of sentences into skolem normal form involves conversion to conjunctive normal form and the replacement of quantified variables by skolem functions. In addition, substitution into the context of belief must be carried out correctly. First the technique for conversion to cnf of sentences in IB is presented, followed by the definition of a quantified language of belief IBQ and the additional rules for skolem conversion.

The conversion of sentences in IB to cnf is shown in Figure 4.2.2.

Sentence in IB	Cnf	Clause form
APQ	APQ	*[{P,Q}]*
CPQ	A NP Q	*[{NP,Q}]*
KPQ	KPQ	*[{P},{Q}]*
NP	NP	*[{NP}]*
<B> P	<B> P	*[{<B>[P]}]*
NAPQ	K NP NQ	*[{NP},{NQ}]*
NCPQ	K P NQ	*[{P},{NQ}]*
NKPQ	A NP NQ	*[{NP,NQ}]*
NNP	P	*[{P}]*
N<B> P	N<B> P	*[{N<B>N NP}]*

Figure 4.2.2. Sentences in IB with their equivalent in cnf and the clausal form.

Where P and Q are atomic formula the figure gives an exhaustive conversion table. If P and Q are allowed to be any well formed formula the conversion process must continue until only atomic formulae and their negations remain. Any sentence may be converted into one beginning with A or K, any sub–sentence may be similarly converted. We use the following equivalences:

	Sentence in IB	Equivalent in IB	Clause form
1.	A AUV Q	A U A V Q	*[{U,V,Q}]*
2.	A KUV Q	K AUQ AVQ	*[{U,Q},{V,Q}]*
3.	K AUV Q	K AUV Q	*[{U,V},{Q}]*
4.	K KUV Q	K U K V Q	*[{U},{V},{Q}]*

where AUV or KUV are sentences in cnf which replace P in APQ or KPQ

With these equivalences a method for skolem conversion which uses a recursive function ***skolem*** is defined. If u is the input sentence of IB, the function ***skolem(u,v)*** returns one or more lists of disjuncts $v_1...v_n$, the clauses whose conjunction is equivalent to u. The function *skolem–not(u,v)* is equivalent to *skolem(Nu,v)*. The conversion rules are given in Figure 4.2.3. Rules 5,8 and 9 return at least two values for γ. These rules may be applied after any of the other rules during conversion. The effect is to distribute K across each of the other operators. This is justified by the equivalences 1−4 above and the rules IB1 and IB2 for the belief operator.

1. *skolem*(α, [α]) *if* α *is atomic.*

2. *skolem−not*(α, [Nα]) *if* α *is atomic.*

3. *skolem*(A$\alpha\beta$, $\alpha'\beta'$) *if skolem*(α, α')
   and skolem(β, β').

4. *skolem*(C$\alpha\beta$, $\alpha'\beta'$) *if skolem−not*(α, α')
   and skolem(β, β').

5. *skolem*(K$\alpha\beta$, γ) *if skolem*(α, α')
   and skolem(β, β')
   and $\gamma=\alpha'$ *or* $\gamma=\beta'$.

6. *skolem*(Nα, α') *if skolem−not*(α, α').

7. *skolem*(<B> α, <B>α') *if skolem*(α, α').

8. *skolem−not*(A$\alpha\beta$, γ) *if skolem−not*(α, α')
   and skolem−not(β, β')
   and $\gamma=\alpha'$ *or* $\gamma=\beta'$.

9. *skolem−not*(C$\alpha\beta$, γ) *if skolem*(α, α')
   and skolem−not(β, β')
   and $\gamma=\alpha'$ *or* $\gamma=\beta'$.

10. *skolem−not*(K$\alpha\beta$, $\alpha'\beta'$) *if skolem−not*(α, α')
    and skolem−not(β, β').

11. *skolem−not*(Nα, α') *if skolem*(α, α').

12. *skolem−not*(<B> α, N<B>Nα') *if skolem−not*(α, α').

Figure 4.2.3. Clausal form conversion rules for IB.

where $\alpha'\beta'$ is the union of lists α' and β'

The conversion rules can be immediately written as rules in Prolog and can be adapted to convert the quantified logic. One similar Prolog routine requires six sets of rules to achieve conversion to skolem normal form [3]. The method presented above is adapted from the reduction stage of the tableau proof method as described by Snyder [4].

Quantification

The definition of the quantified language of belief of IBQ is essentially the same as that defined in Konolige [1, see also Section 3.4.4]. The syntax and semantics are now given.

Definition 4.2.3. The language IBQ is composed of the following symbols:

1. A set of predicate letters of degree n written P^n.

2. A set of individual variables x,y,...

3. A set of individual constants a,b,...

4. A set of function of degree n written f^n.

5. A set of truth functional operators A,K,C,N.

6. The quantifier symbols $\forall, \exists$.

7. A set of agents B_0, B_1...

8. The symbols < ,> ,(,).

Terms and atomic formulae are defined in the usual way [Section 3.2.2.].

Definition 4.2.4. A well formed formulae (wff) of IBQ is defined by the following rules:

1. An atom is a well formed formula.

2. If α and β are wff then $A\alpha\beta$, $C\alpha\beta$, $K\alpha\beta$, $N\alpha$, $<Bi>\alpha$ are well formed.

3. If α is well formed and x is free in α then $\forall x\ \alpha$ and $\exists x\ \alpha$ are well formed.

4. A formula is well formed only if it can be constructed by the above rules.

Definition 4.2.5. The interpretation of IBQ. The interpretation is specified by a tuple $<\varphi, \nu_0, U, D, \eta>$ where:

U is a universe of individuals

φ is a mapping from constants to individuals in U

ν_0 is an assignment of truth values to all ground atoms containing only elements of U

$D = \{d_0, d_1, \ldots\}$ $di = <B, \rho(i)>$ and $bel(di) = \{\psi \mid B \vdash_{\rho(i)} \psi\}$

$\eta = \{\eta_0, \eta_1 \ldots\}$

Let $m = <\varphi, \nu_0, U, D, \eta>$ be a model structure. By $\models_m P$ we mean that m assigns

*The language IBQ is the same as Konolige's language LBQ except for a modification to the syntax.

the sentence P the value true. By P^{φ} we mean the sentence P with all of its constants replaced by the corresponding elements of U.

1. If P is a ground atom $\models_m P$ iff $\nu_0(P^{\varphi})=$ true
2. $\models_m$ K P Q iff $\models_m P$ and $\models_m Q$

   $\models_m$ A P Q iff $\models_m P$ or $\models_m Q$

   $\models_m$ C P Q iff $\not\models_m P$ or $\models_m Q$

   $\models_m$ N P iff $\not\models_m P$
3. $\models_m \exists x P$ iff for some $k \in U$, $\models_m P_k{}^x$

   (every x in P is replaced by k)
4. $\models_m \forall x P$ iff for all $k \in U$, $\models_m P_k{}^x$
5. $\models_m \langle B_i \rangle P$ iff $P \in bel(d_i)$

The additional skolemization rules required by IBQ are given in Figure 4.2.4. These rules are expressed in terms of the functions *skolem* and *skolem−not* as

13. *skolem*($\forall x\alpha$, α') *if skolem*(α, α')
    (by definition α *is within the scope of x)*

14. *skolem*($\exists x\alpha$, β) *if* $\exists x\alpha$ *lies within the scope of universally quantified variables* $x_0 \ldots x_n$
    and skolem(α, α')
    and $\beta = \alpha'.\{h(x_0 \ldots x_n)/x\}$
    where h() is a skolem function new to the entire set of sentences

15. *skolem−not*($\forall x\alpha$, β) *if* $\forall x\alpha$ *lies within the scope of universally quantified variables* $x_0 \ldots x_n$
    and skolem−not(α, α')
    and $\beta = \alpha'.\{h(x_0 \ldots x_n)/x\}$
    where h() is a skolem function new to the entire set of sentences

16. *skolem−not*($\exists x\alpha$, α') *if skolem−not*(α, α')
    (by definition α *is within the scope of x)*

Figure 4.2.4. Additional skolemization rules for IBQ.

before. Rules 14 and 15 require the list of universally quantified variables that α lies within the scope of: this list is generated by rules 13 and 16. The rules implement the procedure for skolem conversion as described by Robinson [5]. In practice the

Prolog rules maintain a list of the universally quantified variables which is a parameter of the *skolem* function.

A universally quantified variable may 'become' existentially quantified as the conversion process proceeds so the substitution of skolem functions for variables must be done during conversion. The scope of a quantified variable does not change through conversion and so a variable x, quantified outside a belief operator must be replaced by the schematic function g(x) within the scope of the belief operator, prior to conversion to clause form. This procedure is correct because if x is universally quantified g(x) must appear in the agents belief set else if x is existentially quantified g(h) must appear in the belief set (where h is a skolem function). The procedure for the substitution of terms for variables ensures that all occurrences of x are replaced by the new term (x or h) irrespective of whether x is the argument of the schematic function.

4.2.6. Resolution.

The resolution method described by Konolige requires that deductions in an agents belief set are carried out in a view which must be opened by a negative belief literal. This presents a difficulty in designing effective resolution methods. For example, backwards chaining is a common strategy for refutation proofs, the goal is to resolve away a set of literals which may include ordinary literals, negative belief literals and positive belief literals. The latter case presents a problem as backwards chaining must stop when such a literal is encountered and the algorithm must begin selecting negative belief literals and opening views in order to resolve the positive literal. There is no obvious criterion for selecting the negative belief literals.

From our interpretation of views as belief worlds and the commitment to the truth of $N\alpha$ in some view, given $N\langle B\rangle\alpha$, as described earlier, we give the following rules for performing deduction by resolution in a view for agent B_0.

Rule V1 A view for agent B_0 may be opened by a positive or a negative belief literal.

V2 When a negative belief literal is added to a view the view is said to be fixed.

V3 Positive belief literals $<B_0>[]$ may be added to a view.

V4 Negative belief literals $N<B_0>[]$ may be added to a view if it is not in the fixed state.

V5 The resultant of two clauses in a view is found by binary resolution.

V6 A view is closed when it is in the fixed state and the null clause is derived

Note: By the definition of skolem normal form each belief literal has a single clause as its argument.

These rules ensure that exactly one negative belief literal is included in a view and that views may be opened by a positive or a negative belief literal. The binary resolution rule ensures that the null clause is derived if the clauses are inconsistent and the attachment rule connects ordinary literals and belief literals. These ideas are summarised in the EB– resolution rule which we give after re–stating the rule for B– resolution:

B– resolution

$$\frac{\begin{array}{l} N<B>A_0,\ R_0 \\ <B>A_1,\ R_1 \\ \ldots \\ <B>A_k,\ R_k \end{array}}{R_0 \ldots R_k} \qquad \text{where } A_1 \ldots A_k \vdash_\rho A_0$$

We replace the deduction of A_0 by a refutation proof and require that one negative belief literal is included.

EB− resolution

$$
\begin{array}{ll}
<B>A_1,\ R_1 & \\
N<B>N\ A_0,\ R_0 & \\
\ldots & \\
\underline{<B>Ak,\ Rk\qquad} & \\
R_0\ldots Rk & \qquad \textit{where} \qquad
\begin{array}{l}
A_0 \\
\ldots \\
\underline{Ak\quad}\ \ \rho \\
[]
\end{array}
\end{array}
$$

In practice the deduction rules ρ will be sentences of the form $<B>C\alpha\beta$, if $\{<B>D_1\}\ldots\{<B>Dn\}$ is the clause form of a subset of ρ then the refutation proof can be written:

$$
\begin{array}{l}
A_0 \\
\ldots \\
Ak \\
D_1 \\
\ldots \\
\underline{Dn\quad} \\
[]
\end{array}
$$

An algorithm which implements the EB− resolution rule is given in Figure 4.2.5. The goal G is proven by translating NG into clause form and deriving the null clause. The functions *resolve* and *EB−resolve* return the value true if the null clause can be derived. The Wise Man Puzzle can be solved using this algorithm. Finding the solution by backwards chaining requires a view to be opened for a positive belief literal.

Quantification.

A Herbrand's theorem for IBQ can be derived: a set of sentences in IBQ in skolem normal form is unsatisfiable if and only if a finite set of its ground instances is unsatisfiable [1]. For refutation proofs of sentences in the quantified logic, in addition to returning the value true or false the functions *resolve* and *EB−resolve* must return the substitutions which allow the literals to be unified. The constants which correspond to the term g(c) in the belief system of the agent, where c is an individual in the real world, must be substituted as defined by the naming function. This function must be defined explicitly as part of the model of the agents beliefs.

```
resolve([]).

resolve(G) if G=[g1...gm]
              and data(D)
              and di ∈ D
              and di complements g1
              and resolve(G-g1)
              and resolve(D-di).

resolve([v<B0>G1/G2]) if EB-resolve(B0, G1)
              and the view closes
              and resolve(G2).

EB-resolve(B0, []).

EB-resolve(B0, G) if G=[g1...gm]
              and data(D)
              and D=(v<B0>D1)D2 and di ∈ D1
              and di complements g1
              and EB-resolve(B0, D1-di)
              and EB-resolve(B0, G-g1)
              and resolve(D2).

EB-resolve(B0,[v<B1>G1/G2]) if EB-resolve(B0B1, G1)
              and the view closes
              and EB-resolve(B0, G2).
```

Figure 4.2.5. Backwards chaining algorithm implementing EB-resolution.

D and G are lists of literals, d and g are literals and $D-d$ is the list D minus d, v takes the value N if $<B>G$ is a negative belief literal. The checks which ensure rules V1-6 are enforced are omitted from the algorithm for the sake of clarity.

Note that in depth first search the above backwards chaining algorithm is homogeneous whereas Konolige's system would have to switch between different views.

4.2.7. Hyperresolution.

Hyperresolution is a forward chaining resolution method. Inferences are made by selecting a rule and deriving the conclusion if the conditions can be satisfied. In the deduction model the agents' rules are a parameter of the model and inferences are made if the conditions of the rules can be satisfied by clauses in the belief set. The deduction model exactly describes the inferences that are made if hyperresolution is applied to clauses in the belief set of an agent. The model requires that deductions are made by applying rules and that derived clauses may participate in further deductions, this is exactly what is carried out by hyperresolution. The hyperresolution method we present implements the deduction model directly. One advantage of hyperresolution over backwards chaining is that the search space is sparser [Section 3.2.6].

The derived clauses represent the agents beliefs to a specified depth of search and as they are explicitly stored in the database they can be examined and introspective reasoning can be carried out. These aspects of the inference procedure are examined in the two following sections.

The hyperresolution rule uses binary resolution and a version of EB− resolution where only unit clauses may be resolved with the goal clause, called HEB− resolution.

HEB− resolution

$$\begin{array}{ll} <B>[P_1 \ldots Pm] & \\ <B>[P_1'] & \\ \underline{N<B>N[Pi']} & \\ \ldots & \\ \underline{<B>[Pm']} & \\ [\,] & \textit{where } Pi \textit{ and } Pi' \textit{ can be shown to be} \\ & \textit{complementary by binary or HEB−resolution} \end{array}$$

Now the hyperresolution* rule H1 can be defined:

Rule H1

$$\begin{array}{ll} <B>[L_1 \ldots Lm,Ln] & \\ <B>[L_1'] & \\ \ldots & \\ \underline{<B>[Lm']} & \\ <B>[Ln] & \textit{where } Li \textit{ and } Li' \textit{ can be shown to be} \\ & \textit{complementary by binary or HEB−resolution} \end{array}$$

This rule shows how positive beliefs are propagated. There is also a need to propagate negative belief as is shown by considering the following sentences:

1. $<B>CP_1P_2$

2. $N<B>NP_1$

$<B>NP_2$ is inconsistent with 1. and 2. therefore $N<B>NP_2$ can be inferred, but this sentence cannot be derived by H1. The hyperresolution rule which allows such deductions to be made is defined as follows:

*This is an incomplete version of hyperresolution for Horn clauses.

Rule H2

$<B>[L_1 \ldots Lm, Ln]$
$<B>[L_1']$
$N<B>N[Li']$
$\ldots$
$\underline{<B>[Lm']}$
$N<B>N[Ln]$

where one negative belief literal occurs and where Li and Li' can be shown to be complementary by binary or HEB–resolution

Tableau proofs of rules H1 and H2 are given in Appendix 2. Intuitively, Rule H1 propagates belief in all belief worlds whereas Rule H2 propagates belief in the belief world of the one negative belief literal (the world where *Li'* is true). Both rules are required to solve the Wise Man Puzzle by forward chaining, the solutions to this puzzle are discussed further in Appendix 3.

Each rule in the belief set has the form $<B>[L_1 \ldots Lm, Ln]$. The algorithm for implementing hyperresolution selects each rule in turn and attempts to derive the null clause from $[L_1 \ldots Lm]$, the tail of the rule, by HEB or binary resolution. If one negative belief literal is included in the derivation then $N<B>N[Ln]$ is added to the database (by Rule H2), else $<B>[Ln]$ is added (by Rule H1). This procedure is repeated n times where n is the depth of inference.

4.2.8. Conclusions

Effective forwards and backwards chaining resolution methods have been developed for the deduction model of belief. These methods may be used alone or combined to prove theorems in belief logic. The analysis of belief worlds, the procedure for deriving the skolem normal form and the resolution methods are new developments of the deduction model.

The clear semantics of Konolige's belief logic are retained as is the concept of attachment and his method of solving the quantifying–in problem. The belief logic does not have the property of omniscience as the deductions which can be made are determined by the deduction rules, an explicit parameter of the model [1, Sections 3.4.2.– 3.4.4.]. The new resolution method uses the name of the agent as a pointer to the world a literal exists in, whereas a variable or skolem function is used for this purpose by the resolution methods for the full modal logics [Section 3.3.3.]. The new resolution methods cannot prove all theorems of the analagous K modal system, however the new methods can deduce all beliefs from rules which are true in all belief worlds which is what is required in order to model belief systems.

The deduction model was designed to be a model of the beliefs of an agent. Agents are able to make deductions within their own beliefs and to reason about the beliefs of other agents. A simple system of temporal reasoning can be developed in the same framework. Facts and rules that were believed in the past may be viewed as being true in a <past> belief world. An agent could have the following beliefs:

<Joe><past> P	Joe believes that P was true in the past.
<Joe> C <past> R T	Joe believes that if in the past R was true then T is true.

Deductions can be made in the past belief world and past beliefs may be used to determine present beliefs. Other modal concepts can be treated within this framework, for example, words such as 'could', 'would' etc. in English are often used to distinguish what is desired or wished from reality. A belief operator <wish> could express these ideas (wishful thinking !).

The addition of the belief operator to first order logic gives a new logic which is more expressive than the original. The framework of belief logic can express many modal ideas. Inferences from a set of sentences in belief logic can be made using the automated deduction methods developed by the author.

References

1. Konolige K. A Deduction Model Of Belief. Pitman, London 1988

2. Konolige K. A Resolution Method For The Quantified Modal Logics Of Knowledge And Belief. Proc. Theoretical Aspects Of The Reasoning About Knowledge Conference 1986 :309–324

3. Clocksin W.F. Mellish C.S. Programming In Prolog. Springer–Verlag, Berlin 1984

4. Snyder P.D. Modal Logic And Its Applications. Van Nostrad Reinhold, New York 1971

5. Robinson J.A. Logic: Form And Function. Edinburgh University Press 1979

4.3 An Expert System For Pulmonary Function And Lung Sound Analysis

4.3.1. Introduction

In this section the building of an expert system based on a logical model of the field of expertise is described. The application is the interpretation of pulmonary function tests. We show how the model is developed and refined and its dependence on the underlying logical language. The refinement of two groups of rules in first order logic is explained as a process of defining new rules and new logical concepts. This refinement is continued when the rules are expressed in belief logic. This procedure gives an explicit logical model of the domain, in the form of a rule base. The use of belief logic in an expert system is a novel approach to expert system design. The rules for analysing lung sound data are derived from work done in Section 2.5 and are integrated with the rules for pulmonary function analysis. The integrated system is tested.

We begin by discussing the basic features of pulmonary function test interpretation and the conversion of this knowledge into rules and the rule base in first order logic. The more novel implementation in belief logic is given later in the chapter.

4.3.2. The Interpretation of Pulmonary Function Test Data.

A measurement such as vital capacity (vc) is interpreted with respect to the predicted normal value calculated from the sex, age and height of the subject [1]. Often several airflow or volume measurements are assessed in order to decide whether airways are obstructed or lung volume is reduced. First we consider the calculation of predicted normal values.

For a male adult the predicted normal value of vc is given by the equation:

$$vc_p = -0.022*Age + 5.29*Height - 3.09 \qquad (4.3.1.)$$

For a female adult the prediction equation is:

$$vc_p = -0.024*Age + 4.44*Height - 2.59 \qquad (4.3.2.)$$

Confidence in the predicted normal value is determined by the standard deviation (SD) associated with each prediction equation. The predicted normal range is defined as $vc_p \pm 2SD$, that is, the normal region is bounded by the predicted upper bound ($vc_p + 2SD$) and the predicted lower bound ($vc_p - 2SD$). The facts about the sex, age, height and weight of the subject are stored in the database. Figure 4.3.1. shows these facts expressed in first order logic. The notation used is that of the belief logic IBQ described in Section 4.2.5. (the '#' symbol marks the end of a logical sentence). Figures 4.3.1. and 4.3.2. are taken from the database and rule base, both of which are text files. The predicates have the following definitions.

code(x)	x is the subjects identification code
sex(x)	x is one of the symbols 'male', 'female'
age(x)	x is the subjects age in years
height(x)	x is the subjects height in metres
weight(x)	x is the subjects weight in kg

These predicates are used as the conditions of logical rules which express equations 4.3.1. and 4.3.2. The conclusion is the predicate 'prednormal' which has the following definition.

prednormal(x, y)	x is the measurement (vc for example) y is the predicted normal value of x

In Figure 4.3.2. rules 101,102 implement equations 4.3.1. and 4.3.2. The predicate 'equation' also appears as a condition of the rules, it performs the following calculation:

equation($v_1,v_2,w_1,w_2,x_1,x_2,y,z$)

where $v_1*v_2+w_1*w_2+x_1*x_2+y = z$

'equation' performs a combination of additions and multiplications which is commonly used in the set of rules for calculating the predicted normal values. It can be introduced into the inference procedure as an instance of theory resolution, where the theory is simply real number arithmetic. The predicates 'add' and 'multiply' can also be included.

add(x,y,z)	where $x+y=z$
multiply(x,y,z)	where $x*y=z$

To calculate the predicted upper and lower bounds the predicted normal value must be known, together with the predicate 'normallimit' which has the following definition:

```
1
code ( pf )#
2
sex ( female )#
3
age ( 67 )#
4
height ( 1.39 )#
5
weight ( 1.57 )#
```

Figure 4.3.1. A section of the database.

```
101
Vx Vy Vz C K sex ( male )
        K age ( x )
        K height ( y )
        equation ( -0.022, x, 5.29, y, 0, 0, -3.09, z )
        prednormal ( vc, z )#
102
Vx Vy Vz C K sex ( female )
        K age ( x )
        K height ( y )
        equation ( -0.024, x, 4.44, y, 0, 0, -2.59, z )
        prednormal ( vc, z )#
```

Figure 4.3.2. A section of the rule base.

```
117
Vv Vw Vx Vy Vz C K prednormal ( v, w )
        K sex ( x )
        K normallimit ( x, v, y )
        add ( w, y, z )
        predupperbound ( v, z )#
118
Vv Vw Vx Vy Vy1 Vz C K prednormal ( v, w )
        K sex ( x )
        K normallimit ( x, v, y )
        K multiply ( y, -1.00, y1 )
        add ( w, y1, z )
        predlowerbound ( v, z )#
119
normallimit ( male, vc, 1.1 )#
120
normallimit ( female, vc, 0.88 )#
```

Figure 4.3.3. Rules in first order logic.

normallimit(x,y,z,)	x is male or female y is the measurement z is 2SD for measurement y, sex x

Figure 4.3.3. gives the rules for calculating the predicted upper bound values for all measurements. A total of 34 rules is required to calculate all predicted values for all measurements.

One or more of eight pulmonary function measurements may be used to assess the state of the lungs. These measurements are referred to by the shorthand symbols given in Figure 4.3.4.

Symbol	Measurement
vc	vital capacity
frc	functional residual capacity
tlc	total lung capacity
rv	residual volume
rvtlc	the ratio rv/tlc
fev	forced expiratory volume in 1 second
fevfvc	the ratio fev/fvc
kco	transfer coefficient

Figure 4.3.4. Definition of medical symbols.

The degree of airway obstruction is assessed as mild if the following conditions are met:

$$0.8*(fevfvc_p - 2SD) \leqslant fevfvc < fevfvc_p - 2SD \qquad (4.3.3.)$$

where fevfvc = the measured value of the ratio fev/fvc

and $fevfvc_p$ = the predicted value of the ratio fev/fvc

equivalently

0.8*predicted lower bound fevfvc ⩽ measured fevfvc

< predicted lower bound fevfvc

The predicted values of each measurement can be calculated. We define two more predicates

measured(x,y)	x is the measurement y is the value of x
degree(x,y)	y is the clinical abnormality x is the degree (or seriousness) of y

The rule defining mild airways obstruction can be written:

$$
\begin{aligned}
\forall x \forall y \forall z \ \ C \ \ & K \ \text{measured(fevfvc, x)} \\
& K \ \text{prednormal(fevfvc, y)} \\
& K \ \text{multiply(y,0.8,z)} \\
& K \ \text{lessthan or equal(z,x)} \\
& \text{lessthan(x,y)} \\
& \text{degree(mild, airwaysobstruction)} \qquad (4.3.4.)
\end{aligned}
$$

The translation of each pulmonary function interpretation rule from the conceptual form of equation 4.3.3. to the logical form of 4.3.4 is carried out as above.

The knowledge was available in a form where it could be readily translated into logical rules. This situation arose because a computer based interpretation system based on these rules already exists in the Department of Respiratory Medicine at Glasgow Royal Infirmary [2]. Consequently the task of knowlege acquisition has been carried out over a number of years by Dr.R.Carter, R.McCusker and other health service staff. The advantage of the expert system approach is that the calculation of normal values and test interpretation can be defined in one logical language, at present two programs written in different languages are required (the languages are Fortran and Quark). The logical model of the area of expertise can be examined more readily when the rules are written in a formal logical notation. This logical model will be referred to as the "knowledge model". The knowledge model is expressed in the language of the underlying logical system which may be first order logic or belief logic. The refinement of the knowledge model is now described.

4.3.3 The Knowledge Model in First Order Logic.

This section presents two important sets of rules which are part of the rule base. The first set to be examined assesses the degree of airways obstruction, the second set assesses the degree of restriction. Consideration of these two groups demonstrates the method we wish to present. The complete rule base is too large for each rule to be discussed individually here.

One of two sets of rules for the assessment of airways obstruction is applied depending on the value of fev. Fev is low if:

measured fev < 0.7*predicted lower bound fev

otherwise fev is not low. Rules 149 and 168 [Appendix 4] define whether 'low(fev)' is true or 'notlow(fev)' is true. In the latter case the rules in Figure 4.3.5. are applied. Their expression in first order logic is given in Figure 4.3.6. These rules have only three conditions in their conceptual form and up to seven conditions in the logical form. The rules are not complex and the translation from conceptual to logical form is immediate. When fev is low the rules given in Figure 4.3.7. are applied, Figure 4.3.8. gives the logical form of these rules. Again the translation to logical form is immediate. The assessment of the degree of airways obstruction is based on which range, defined with respect to the predicted lower bound, the measured fev/fvc ratio falls into.

The rules for assessing the severity of a restrictive defect are more complex than those quoted so far. In order to express them clearly one rule must be broken down into several, and new predicates must be defined. These rules are defined in terms of the values of tlc and the ratio rv/tlc in addition to the ratio fev/fvc, Figure 4.3.9. gives three rules in conceptual form.

The comparison of a measured value with its predicted upper or lower bound appears at least twice in each rule. A predicate which defines whether a measurement is above or below its upper or lower bound can be defined:

range(x,y)	x is the measurement y is one of the symbols 'aboveupperbound' 'belowupperbound','abovelowerbound', 'belowlowerbound'

Four rules define which values of y are true for each measurement. They all have the form:

∀x∀y∀z C K measured(x,y)

K predupperbound(x,z)

lessthan(y,z)

range(x,belowupperbound)

The condition 'measured fev/fvc > predicted lower bound fev/fvc' appears in

mild airways obstruction if:
*0.8*predicted lower bound fevfvc ⩽ measured fevfvc*
*< predicted lower **bound** fevfvc*

moderate airways obstruction if:
*0.6*predicted lower bound fevfvc ⩽ measured fevfvc*
*< 0.8*predicted lower **bound** fevfvc*

moderately severe airways obstruction if:
*0.45*predicted lower bound fevfvc ⩽ measured fevfvc*
*< 0.6*predicted lower **bound** fevfvc*

severe airways obstruction if:
*0.45*predicted lower bound fevfvc > measured fevfvc*

Figure 4.3.5 The assessment of airways obstruction when FEV1 is not low.

```
135
Vx Vy Vy1 Vy2 C K notlow ( fev )
        K measured ( fevfvc, x )
        K predlowerbound ( fevfvc, y )
        K multiply ( y, 0.8, y1 )
        K lessthanorequal ( y1, x )
        lessthan ( x, y )
        degree ( mild, airwaysobstruction )#
136
Vx Vy Vy1 Vy2 C K notlow ( fev )
        K measured ( fevfvc, x )
        K predlowerbound ( fevfvc, y )
        K multiply ( y, 0.6, y1 )
        K multiply ( y, 0.8, y2 )
        K lessthanorequal ( y1, x )
        lessthan ( y2, x )
        degree ( moderate, airwaysobstruction )#
137
Vx Vy Vy1 Vy2 C K notlow ( fev )
        K measured ( fevfvc, x )
        K predlowerbound ( fevfvc, y )
        K multiply ( y, 0.45, y1 )
        K multiply ( y, 0.6, y2 )
        K lessthanorequal ( y1, x )
        lessthan ( y2, x )
        degree ( moderatelysevere, airwaysobstruction )#
138
Vx Vy Vy1 Vy2 C K notlow ( fev )
        K measured ( fevfvc, x )
        K predlowerbound ( fevfvc, y )
        K multiply ( y, 0.45, y1 )
        lessthan ( x, y1 )
        degree ( severe, airwaysobstruction )#
```

Figure 4.3.6. Rules in first order logic.

major airways obstruction if:
*0.65*predicted lower bound fevfvc ⩽ measured fevfvc*
< predicted lower bound fevfvc

severe airways obstruction if:
*0.65*predicted lower bound fevfvc > measured fevfvc*

Figure 4.3.7. The assesment of airways obstruction when FEV1 is low.

```
142
Vx Vy Vy1 Vy2 C K low ( fev )
        K measured ( fevfvc, x )
        K predlowerbound ( fevfvc, y )
        K multiply ( y, 0.65, y1 )
        lessthan (  x, y1 )
        degree ( severe, airwaysobstruction )#
143
Vx Vy Vy1 Vy2 C K low ( fev )
        K measured ( fevfvc, x )
        K predlowerbound ( fevfvc, y )
        K multiply ( y, 0.65, y1 )
        K lessthan ( x, y )
        lessthanorequal ( y1, x )
        degree ( major, airwaysobstruction )#
```

Figure 4.3.8 Rules in first order logic.

mild restrictive defect if:
predicted lower bound fevfvc < measured fevfvc
*and 0.9*predicted lower bound tlc ⩽ measured tlc*
< predicted lower bound tlc
and predicted upper bound rv/tlc > measured rv/tlc

moderate restrictive defect if:
predicted lower bound fevfvc < measured fevfvc
*and 0.8*predicted lower bound tlc ⩽ measured tlc*
*< 0.9*predicted lower bound tlc*
and predicted upper bound rv/tlc > measured rv/tlc

severe restrictive defect if:
predicted lower bound fevfvc < measured fevfvc
*and 0.8*predicted lower bound tlc > measured tlc*
and predicted upper bound rv/tlc > measured rv/tlc

Figure 4.3.9. The assessment of restrictive defects.

```
150
Vx Vx1 Vx2 C K noevidence ( airwaysobstruction )
        K range ( rvtlc, belowupperbound )
        K range ( tlc, belowlowerbound )
        K measured ( tlc, x )
        K predlowerbound ( tlc, x1 )
        K multiply ( x, 0.9, x2 )
        lessthanorequal ( x2, x )
        degree ( mild, restrictivedefect )#
151
Vx Vx1 Vx2 Vy C K noevidence ( airwaysobstruction )
        K range ( rvtlc, belowupperbound )
        K measured ( tlc, y )
        K predlowerbound ( tlc, x )
        K multiply ( x, 0.9, x1 )
        K multiply ( x, 0.8, x2 )
        K lessthan ( y, x1 )
        lessthanorequal ( x2, y )
        degree ( moderate, restrictivedefect )#
152
Vx Vx1 Vx2 C K noevidence ( airwaysobstruction )
        K range ( rvtlc, belowupperbound )
        K measured ( tlc, x )
        K predlowerbound ( tlc, x1 )
        K multiply ( x, 0.8, x2 )
        lessthan ( x, x2 )
        degree ( severe, restrictivedefect )#
```

Figure 4.3.10. Rules in first order logic.

each rule. The meaning of this comparison is that if the measured value of fev/fvc is greater than the predicted lower bound then there is no evidence of airways obstruction. The numerical comparison may be replaced by 'no evidence(airways obstruction)', a new predicate which is more meaningful. Rule 147 defines when this is the case:

147 C range(fevfvc, abovelowerbound)

no evidence(airwaysobstruction)

The severity of the restrictive defect is determined by which range, defined with respect to the predicted lower bound, the measured value of tlc lies in. The rules are given in logical form using the new predicates 'range' and 'no evidence' in Figure 4.3.10.

The translation of the rules for the assessment of restriction involved additional logical concepts. Replacing numerical comparison tests by a predicate such as 'range' makes the rules more compact and easier to read and understand. Such rules can be said to be based on higher level concepts. The procedure of defining rules and predicates will be referred to as building the knowledge model. The longer or more complex the rule the greater the need for higher level symbolic concepts. As each new predicate is defined by one or more rules it is possible to ensure that there is no 'overlap' between concepts (ie they are formally defined to be distinct).

4.3.4. The Knowledge Model in Belief Logic.

The belief logic IBQ is the logical system developed to implement the deduction model. The deduction model formalises all aspects of the inference process including which inferences are made by the control strategy employed. This model has the property of deductive closure which means in practice that the deduction rules are applied exhaustively to the data set. We use the deduction model to formally define the inference process of the <u>expert system</u>. One or more expert agents may be defined, associated with each is a set of deduction rules. In practice we define one agent called <expert>. The set of rules ρ constitutes the rules for the interpretation of pulmonary function data. Agents may have beliefs about other agents or about their own beliefs. Introspective reasoning is explored in Sections 4.4 and 4.5.

In this approach to expert system design not only the logical language and theory are formally defined but also the actual theorem proving method as implemented on the computer is formally specified.

The knowledge model is the logical model of the area of expertise defined by the deduction rules. It is expressed in the logical language that implements the deduction model, namely the belief logic IBQ. Our discussion of the knowledge model in first order logic made no reference to the framework within which it was set i.e. how inferences were to be made. When the knowledge model is specified in belief logic as deduction rules, these rules are specified as a component of the deduction model. An introduction to the deduction model is given in Section 3.4.4., the algorithms for forwards and backwards inference are developed in Section 4.2.

The deduction model, its language and implementation have been described and it remains to define the deduction rules. This is achieved by translating the knowledge model into belief logic. This can be done immediately by preceding each first order rule by '<expert>', thus placing the first order rule base within the agents belief system. In fact we do not adopt this approach but will use the belief language to further refine the knowledge model. The beliefs of the expert agent are restricted to symbolic concepts. All arithmetical predicates are written on the level of first order logic and all symbolic concepts appear in the belief system (IBQ includes first order logic). We define the universe of individuals which exist in each world to be the same universe U (i.e. the Barcan formula and its converse are true in the model of belief that we adopt for the present application). The deduction rules ρ are given in terms of symbols only, there is a second set of rules ρ' which have numerical terms. The expert rule base is the combination of ρ and ρ'.

By developing the knowledge model in this way we continue the trend towards defining higher level concepts described earlier. This enables a type of introspective reasoning called plausible reasoning to be carried out. The remainder of this section describes the translation of the knowledge model into belief logic. We define the universe of individuals U to be the same in the belief world as in the real world.

Examining the rules for the classification of airways obstruction (Figures 4.3.5 and 4.3.7) it can be seen that the degree of obstruction is dependent on two conditions. The first is whether fev is reduced or not, the second is the range within which the measured value of fevfvc lies. If names are given to these ranges then the logical rules may be specified in terms of symbols alone. The division of the predicted value scale into ranges is illustrated in Figure 4.3.11.

	Range	Multiplier of $fevfvc_p - 2SD$
		1.0
	a	
		0.8
fevfvc →	b	
(measured		0.65
value)	c	
		0.6
	d	
		0.45
	e	
		0.0

Figure 4.3.11. Classification of the fevfvc measurement with respect to the predicted lower bound.

This division is represented in the logical language by the predicate 'define.fevfvc.range':

define.fevfvc.range(x,y,z)

x is the larger multiplier
y is the smaller multiplier
z is the name of the range

Rules 142–146 in Figure 4.3.12 define the ranges, rule 147 compares the measured value of fevfvc to each defined range and adds the predicate 'fevfvc.in.range' to the experts' belief set when the conditions can be satisfied. These rules are applied in the same way as deduction rules and are part of the expert system rule base.

The airways obstruction interpretation rules may now be written more clearly than previously was the case. For example the conclusion of rule 135 is that the degree of airways obstruction is mild if fev is not low and fevfvc is in range a, this is expressed in IBQ as follows:

```
142
define.fevfvc.range ( 1.0, 0.8, a )#
143
define.fevfvc.range ( 0.8, 0.65, b )#
144
define.fevfvc.range ( 0.65, 0.6, c )#
145
define.fevfvc.range ( 0.6, 0.45, d )#
146
define.fevfvc.range ( 0.45, 0, e )#
147
Vx Vy Vy1 Vy2 Vz Vz1 Vz2
        C K measured ( fevfvc, x )
        K predlowerbound ( fevfvc, y )
        K define.fevfvc.range ( z, z1, z2)
        K multiply ( y, z1, y1 )
        K multiply ( y, z, y2 )
        K lessthanorequal ( y1, x )
        lessthan ( x, y2 )
        < expert > fevfvc.in.range ( z2 )#
```

Figure 4.3.12. Rules in belief logic.

```
135
< expert > C K notlow ( fev )
        fevfvc.in.range ( a )
        degree ( mild, airwaysobstruction )#
136
< expert > C K notlow ( fev )
        A fevfvc.in.range ( b )
        fevfvc.in.range ( c )
        degree ( moderate, airwaysobstruction )#
137
< expert > C K notlow ( fev )
        fevfvc.in.range ( d )
        degree ( moderatelysevere, airwaysobstruction )#
138
< expert > C K notlow ( fev )
        fevfvc.in.range ( e )
        degree ( severe, airwaysobstruction )#
139
< expert > C K low ( fev )
        A fevfvc.in.range ( a )
        fevfvc.in.range ( b )
        degree ( major, airwaysobstruction )#
140
< expert > C K low ( fev )
        A fevfvc.in.range ( c )
        A fevfvc.in.range ( d )
        fevfvc.in.range ( e )
        degree ( severe, airwaysobstruction )#
```

Figure 4.3.13. Rules in belief logic.

<expert> C K notlow(fev)

fevfvc.in.range(a)

degree(mild, airways obstruction)

(compare this with the equivalent rule in Figure 4.3.6)

The refined versions of the six rules 135–140 are given in Figure 4.3.13, these are all deduction rules. The method outlined above may be applied to all sets of first order rules. This completes the description of the knowledge model for pulmonary function test interpretation. The complete rule base in IBQ is listed in Appendix 4.

4.3.5. The Interpretation of Lung Sound Data.

Lung sound data and pulmonary function data was available for the group of subjects studied in Section 2.5. Rules for the interpretation of lung sound data are derived from the results of that study. The predicted values of median frequency (F50) and F85 can be calculated from the linear regression equations obtained by statistical analysis. Comparison of the measured values of F50 to the predicted values enables the sound spectrum recorded at a particular site to be assessed as above, within or below the predicted normal range. By examining both F50 and F85 conclusions can be drawn about the significance of the differences between the measured and predicted spectral shape.

We have no direct evidence associating spectral characteristics with physiological changes (other than through FEV1) so the rules which will be defined make no such deductions. The rules give an analysis of spectral shape in the light of clinical observations.

The normal region is defined as the predicted value ± the 90% confidence interval and is expressed in logical rules as was done for the predicted normal values of pulmonary function measurements. The predicted values of F50 and F85 are dependent on the measured value of FEV1 only. The measured value of F50 must lie within one of the sound ranges 'aboveupperbound', 'normalrange' or 'belowlowerbound'. The predicate 'sound.range' is defined below and is added to the belief set when the appropriate conditions are satisfied.

sound.range(x,y,z)	x is recording site y is the measurement z is the sound range

One of two sets of rules is applied depending on whether crackles were observed during recording or not. Where crackles were absent the rules given in Figure 4.3.14 were applied. It is always the case that the conditions of one of these rules can be satisfied.

If F50 is above the predicted upper bound and if F85 above the predicted upper bound then there is a major shift upwards in frequency (relative to the predicted spectral shape). If F50 is above the predicted upper bound then it would be expected that F85 would show a similar shift, if this is not the case we conclude that the spectrum has an abnormal shape. The rules are based on the ideas that a change in F50 should be reflected in a change in F85 and that a change in F85 alone is less significant than a change in F50. The rules for the analysis of sound spectrum where crackles were observed are based on the same considerations and are shown in Figure 4.3.15.

The conversion of rules from the conceptual form to deduction rules is straightforward, the following rule is typical:

<expert> ∀x C K decision.assess(x, crackles)

K sound.range(x, f50, aboveupperbound)

sound.range(x, f85, aboveupperbound)

conclude(severe, crackling, x)

(the variable x stands for one of the recording sites)

These rules can be specified in belief logic without going through the step of giving a first order representation. By carefully defining the 'sound.range' predicate and the interpretation rules it is ensured that all possible combinations of values of F50 and F85 are given some analysis.

sound spectrum shows a major shift to high frequency if:
f50 above the predicted upper bound
and f85 above the predicted upper bound

abnormal sound spectrum if:
f50 above the predicted upper bound
and f85 below the predicted upper bound

normal sound spectrum if:
f50 is in the normal range
and f85 is below the predicted upper bound

sound spectrum shows a moderate shift to high frequency if:
f50 is in the normal range
and f85 is above the predicted upper bound

sound spectrum shows a shift to low frequency if
f50 is below the predicted lower bound
and f85 is below the predicted upper bound

sound spectrum is abnormally broad if:
f50 is below the predicted upper bound
and f85 is above the predicted upper bound

Figure 4.3.14. Lung sound interpretation where no crackles were observed.

severe crackling if:
f50 is above the predicted upper bound
and f85 is above the predicted upper bound

mild crackling if:
f50 is in the normal range
and f85 is above the predicted upper bound

abnormal sound spectrum if:
f50 is below the predicted lower bound
and f85 is above the predicted upper bound

moderate crackling if:
f50 is above the predicted upper bound
and f85 is below the predicted upper bound

insignificant crackling if:
f50 is below the predicted upper bound
and f85 is below the predicted upper bound

Figure 4.3.15. Lung sound interpretation where crackles were observed.

4.3.6. The Integrated Expert System.

The expert system was implemented by two compiled programs. The shell program provided the user interface and was able to call the inference program when required. This division frees memory for stack use when the inference algorithm is being run. The shell provides the user interface in the form of a screen editor which is used to view and edit the rule base and database as these are text files. A command to convert the rules into clausal form is provided, any ill-formed rules are reported as being in error. Commands are provided to call the inference procedure, which need not be given a specific goal, and to to convert the derived clauses (hyperresolvents) from clausal form into the format of belief logic. The results file may be examined using the editor. A trace of the rules used in the derivation of each new fact is given in the results file. This form of output is available as a result of employing the hyperresolution strategy.

A sketch of the programs and files that make up the expert system is presented in Figure 4.3.16. The algorithms for the conversion of sentences to clausal form and to implement hyperresolution are those developed in Section 4.2. The rules for pulmonary function test interpretation and those for lung sound analysis are combined to form the expert rule base. The expert system composed of the shell programs and the pulmonary function and lung sound rule base will be referred to as the 'Inspire' system.

The pulmonary function test results for each subject were stored in a database file. The rule base was applied to each data file and the results stored. Comparison between the results of the pulmonary function interpretation system of the Royal Infirmary and those of the expert system showed agreement in 24/24 cases. In two further cases no comparison was possible as results were obtained using rules which were not implemented in the expert system. The fact that a similar interpretation system exists has eased the problems of knowledge acquisition and of verifying the rule base.

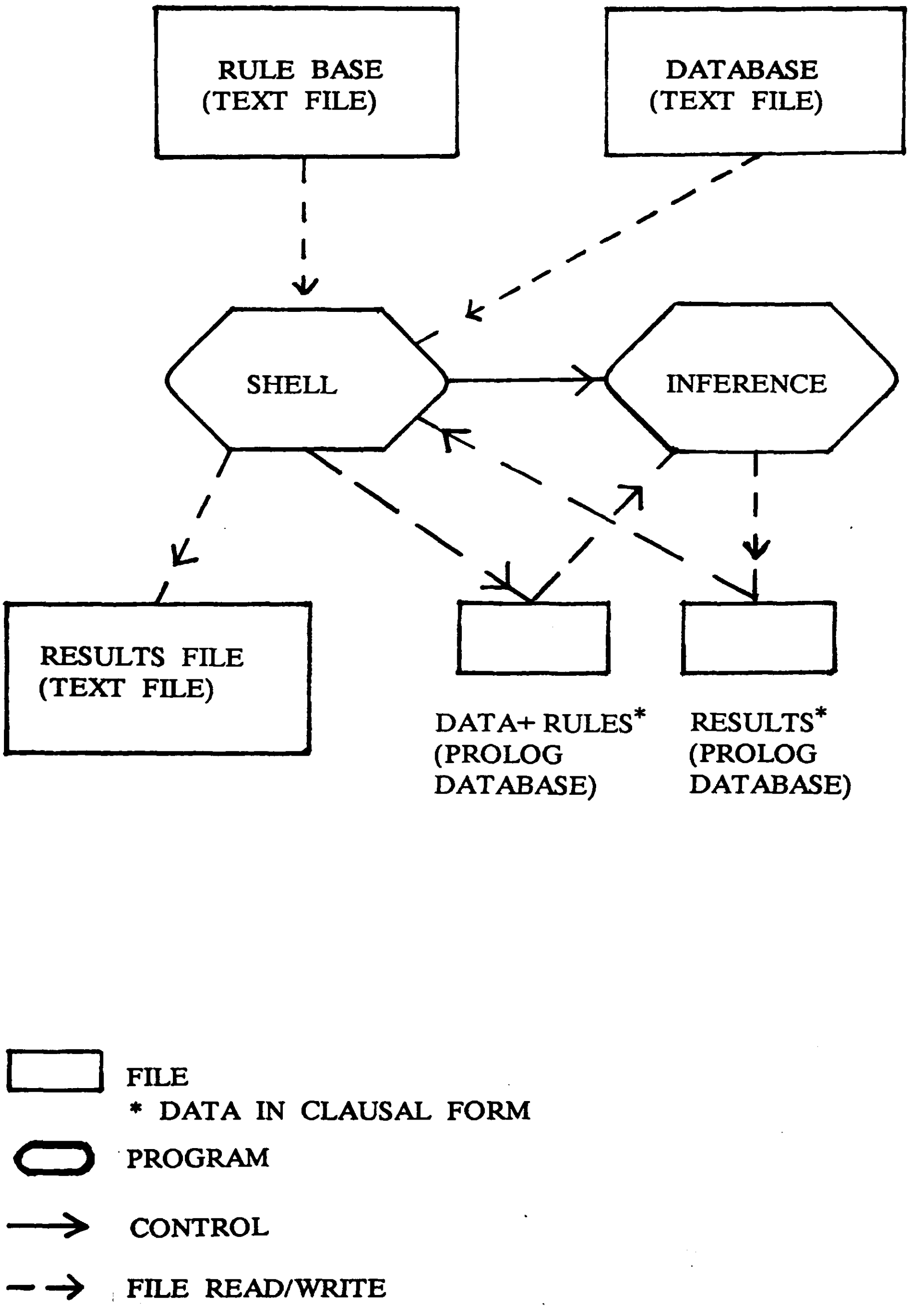


Figure 4.3.16. Programs and files of the Inspire expert system.

4.3.7. Conclusions

The Inspire system was able to integrate numeric and symbolic information in one framework. The technique of defining higher level concepts during the development of a rule base was aided by the use of belief logic. The actual theorem proving method is specified by the deduction model. This feature makes the control mechanism of the inference strategy explicit. The use of a forward chaining search strategy proved useful during the process of refining the knowledge model as incorrect inferences resulting from logical or typing errors in the rule base were identified relatively easily by examining the set of derived sentences. The use of the deduction model to represent other modal notions such as time was not required in the interpretation of pulmonary function measurements. A modal operator such as <past> could be useful in writing rules which describe the progression of disease, as past observations are often used in the assessment of the current condition of a patient. Belief logic can represent this form of reasoning.

An expert system called Puff, which was designed for the interpretation of pulmonary function data, has been implemented by Aikens et al at Stanford University [3]. Puff was built using Emycin, an expert system shell derived from the Mycin system. Knowledge was represented as production rules which were backward-chained to prove a specific goal. Two versions of Puff were developed. One version was written in Interlisp and ran on a DEC 10 computer, the second version was written in Basic and ran on a PDP 11 minicomputer. The latter version was used in a medical department where it was in daily use. Aikens reported that difficulties arose when rules were added or modified, changes in one rule affected others in unexpected ways. Another limitation was the inability of Puff to represent typical cases of lung disorder. Both problems may (in part) be due to the inability of the production rule scheme to fully represent the model of the domain held by the expert. The Puff rule for the diagnosis of moderate airways obstruction is shown in Figure 4.3.17. This rule can be compared with the equivalent rule in belief logic, Figure 4.3.13. The knowledge model is expressed by a larger number of simpler

```
RULE011
-------

 If:   1) A: The mmf/mmf-predicted ratio is between 35 and 45, and
          B: The fvc/fvc-predicted ratio is greater than 80, or
       2) A: The mmf/mmf-predicted ratio is between 25 and 35, and
          B: The fvc/fvc-predicted ratio is less than 80
 Then:  1) There is suggestive evidence (.5) that the degree of
           obstructive airways disease as indicated by the MMF
           is moderate, and
        2) It is definite (1.0) that the following is one of the
           findings about the diagnosis of obstructive airways
           disease: Reduced mid-expiratory flow indicates
           moderate airway obstruction.

PREMISE:  [$AND ($OR ($AND (BETWEEN* (VAL1 CNTXT MMF) 35 45)
                           (GREATERP* (VAL1 CNTXT FVC) 80))
                     ($AND (BETWEEN* (VAL1 CNTXT MMF) 25 35)
                           (LESSP* (VAL1 CNTXT FVC) 80]
ACTION:  (DO-ALL (CONCLUDE CNTXT DEG-MMF MODERATE TALLY 500)
                 (CONCLUDETEXT CNTXT FINDINGS-OAD
                               (TEXT $MMF/FVC2) TALLY 1000))
```

Figure 4.3.17. A Puff production rule in English and Lisp versions.

rules when expressed in belief logic than is the case for the production rule scheme. The rules which define the degree of airways obstruction were conceived of as a group when belief logic was employed. This was a result of refining the knowledge model as described previously.

The Inspire system is an advance on Puff in that the representation scheme employed is based on a more expressive logic.

Inspire was programmed in Prolog and run on IBM PS/2 microcomputers. More development of the user interface will be necessary to bring the Inspire system into clinical use. The inference procedure, including the unification algorithm, was programmed in Prolog and as a result the time taken to process a data file was of the order of minutes. A faster inference procedure could be written in a language such as C or Lisp.

References

1. Cotes J.E. Lung Function: Assessment And Application In Medicine. Blackwell Scientific Publications, Oxford 1968

2. Department Of Respiratory Medicine, lasgow Royal Infirmary, Internal Document 1989

3. Aikens J.S. Kunz J.C. Shortliffe E.H. Fallat R.J. Puff: An Expert System For Interpretation Of Pulmonary Function Data. Computing And Biomedical Research 16 (1983) :199–208

4.4 A Logic Of Plausible Belief

4.4.1 Introduction

This section presents a method of reasoning with incomplete information that is compatible with the first order rules typical of an expert system rule base. The logic of plausible belief is similar to the K45 system of classical modal logic. The set of believed sentences are those which may be derived using the deduction rules. This is an approximation to the ideal case where all logically true consequences may be enumerated (as in the K45 system).

We begin by describing the K45 system, its relationship with the deduction model and the derivation of two rules of inference which allow plausible deductions to be made.

4.4.2 The K45 Modal System.

Propositional K45 is defined by adding modal axioms 4 and 5 to the K system (defined in Section 3.3.2.). The additional axioms are:

4. C LP LLP
5. C MP LMP

When L is interpreted as the belief operator, these axioms define positive and negative introspection respectively. The structure of the resulting modal system is shown in Figure 4.4.1. Any sentence true in all possible worlds wi, $i > 0$ must be true in all worlds accessible from these (by axiom 4), a sentence true in one possible world must be true in at least one accessible world (by axiom 5). These conditions are satisfied if the accessibility relation R is both reflexive (wiRwi,$i > 0$) and an equivalence relation (wiRwj $\supset$ wjRwi, $i,j > 0$), hence all possible worlds are accessible to all other possible worlds as illustrated in Figure 4.4.1.

This system was chosen to model belief because it captures the notion that there are sentences which are believed LP and sentences which are compatible with what is believed MQ. If LP is true then all possible worlds/belief worlds assign the value true to P. P is then said to be a member of the agents' belief set. If MQ is true then Q is true in some belief world, Q is compatible with the belief set but not a

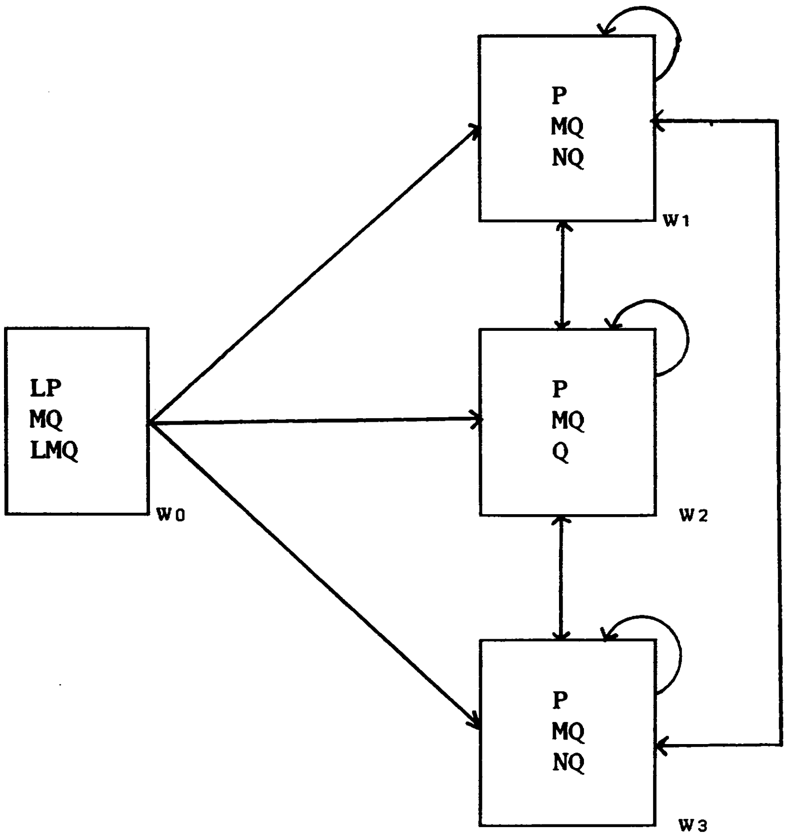


Figure 4.4.1. K45 modal structure.

member of it. Through negative introspection MQ is believed if MQ is true, that is, MQ is true in every belief world (LMQ is true) and therefore a member of the belief set, although Q may not be true in all belief worlds. The agents belief set includes true beliefs (P) and qualified beliefs (MQ). Certain and qualified beliefs are distinguished.

It is possible to identify sentences of the form M $KQ_1KQ_2...Q_n$ through introspection. Such sentences identify n formulae which form a consistent extension of the theory defined by expert rules. This study examines the case where n=1 i.e. we wish to identify sentences of the form MQ.

Autoepistemic logic has been shown to be equivalent to the modal system S5 which is the system obtained by adding the axiom C LP P to K45 [1]. S5 is a logic of knowledge whereas K45 is a logic of belief.

4.4.3. The Deduction Model and the K45 System.

A method for plausible reasoning analagous to the propositional K45 system is now presented. The propositional theory forms the basis of the first order theory but was not implemented.

In defining the K45 system it was not necessary to specify a practical inference procedure. Model theory or tableaux proofs may be used to determine whether sentences are true or false. We now show how the deduction model may be employed as the inference procedure.

The deduction model explicitly states that the only inferences which can be made are those which result from applying the deduction rules to the base set of sentences. A set of deduction rules which allow every logical consequence of the base set to be derived is said to be complete. If such a set of rules is defined then all sentences $\alpha \epsilon B$ in the deduction model correspond to a sentence $L\alpha$ in K45. A set of consistent belief worlds could be constructed where sentences analagous to $M\beta$ (where β is not a member of B) would be assigned truth values in each world. Figure 4.4.2. shows such a worlds diagram, the deduction structure is modified as follows:

$$di = \langle B, BEi, \rho \rangle \quad \text{for world wi, } i > 0$$

where the set B is the belief set in the deduction model
which corresponds to the set $\{\alpha : L\alpha \text{ is true}\}$ in K45
and BE = $\cup_i$ BEi in the deduction model corresponds
to the set $\{\beta : M\beta \text{ is true}\}$ in K45

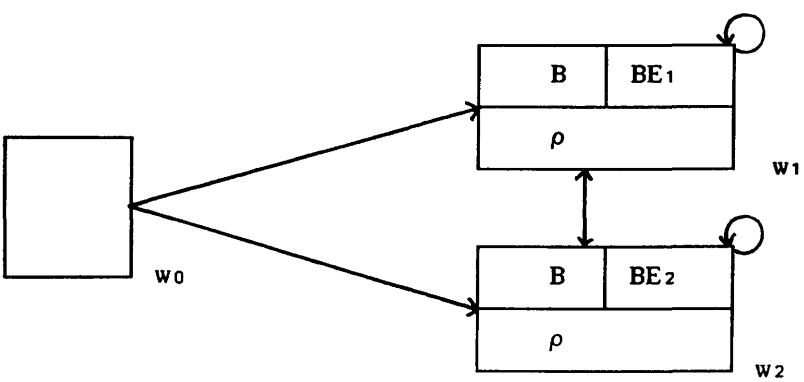


Figure 4.4.2. Belief worlds diagram of the introspective deduction model.

Each belief world must be consistent. If β_1 is true in w_1 it must not be possible to derive γ and $N\gamma$ for any formula γ as a result. If this condition is satisfied $M\gamma$ will be true in all belief worlds.

In the belief logic IB the sentence <expert> α is interpreted as 'α is a member of the belief set of the expert agent' and corresponds to $L\alpha$ in modal logic. If the expert believes 'β is possible' the sentence <expert><possible> β is true, this sentence corresponds to $LM\beta$ in modal logic. In practice the deduction rules will not be complete, therefore we write <expert><plw> β to state that β is true in a world compatible with the experts beliefs. Such worlds are called plausible worlds.

Typically deduction rules do not allow negative information (false atomic sentences) to be inferred. Such information would be specified by a complete belief set. A standard set of deduction rules must be augmented by additional rules which allow negative sentences to be derived. This can be viewed as a process of completing the knowledge model by not only specifying facts which are true but also those which are not. This requires implicit relationships between concepts to be made explicit, for example the concepts 'big' and 'small' are exclusive but this relationship would not normally be stated in an expert system rule base. Completing the knowledge model requires that implicit relationships be defined by rules (ρc). This procedure improves the approximation to the ideal case.

Deduction

The set of beliefs B may be derived by applying the deduction rules ρc n times using the hyperresolution method. Once this has been carried out the following rule of inference may be employed to derive plausible inferences:

Definition 4.4.1. Propositional rule for plausible deduction.

If $N\alpha_1 \notin B$ and $\alpha_2 \ldots \alpha n \in B$

and $CK\alpha_1 K\alpha_2 \ldots \alpha n\ \beta\ \in \rho c$

then $\beta \in BE$ and <plw> $\beta \in B$

The equivalent rule in K45 is:

If $M\alpha_1$ and $L\alpha_2 \ldots L\alpha n$ and $L\ CK\alpha_1 K\alpha_2 \ldots \alpha n\ \beta$ then $M\beta$

This rule states that if α_1 is true in some belief world wi and $\alpha_2 \ldots \alpha_n$ are true in all belief worlds and β is a consequence of $\alpha_1 \ldots \alpha_n$ then β must be true in wi.

This inference rule allows plausible atomic formulae to be added to the agents belief set using the deduction rules. There is no necessary connection between the plausible beliefs: they extend the belief set but do not form competing theories for extending the belief set. The operation of the inference rule is clear and shows the role that negative information plays as if, in applying the above definition, $N\alpha_1$ is found to be a member of B then the conclusion β cannot be derived. This method does not involve the very inefficient truth maintainance procedures required by autoepistemic logic [2].

4.4.4. A First Order Theory for Plausible Reasoning.

With the introduction of quantifiers into the language of belief a universe of individuals must be associated with each belief world. In the belief logic IBQ each belief world is associated with one universe U. In defining a modal structure it is not necessary for the same inividuals to exist in each world. In studying formal systems of belief we are concerned with the truth assignments of atomic formula (which depend upon the inividuals which exist in a particular world). Allowing the individuals which exist in each world to vary in order to satisfy a truth assignment is an unintuitive way to proceed. It is clearer to define a common universe for every world and evaluate the truth of atomic formulae on that basis. These considerations also apply to the first order deduction model whose structure is shown in Figure 4.4.3. The difference between the model structure shown in this figure and that of IBQ is that every belief world is accessible from every other. To represent plausible beliefs the operator <plw> is added to the language IBQ by defining <plw>α to be a well formed formula if α is well formed. As was the case for the propositional logic the <plw> operator is analagous to the possibility operator M in modal logic. The rule of inference is defined as follows:

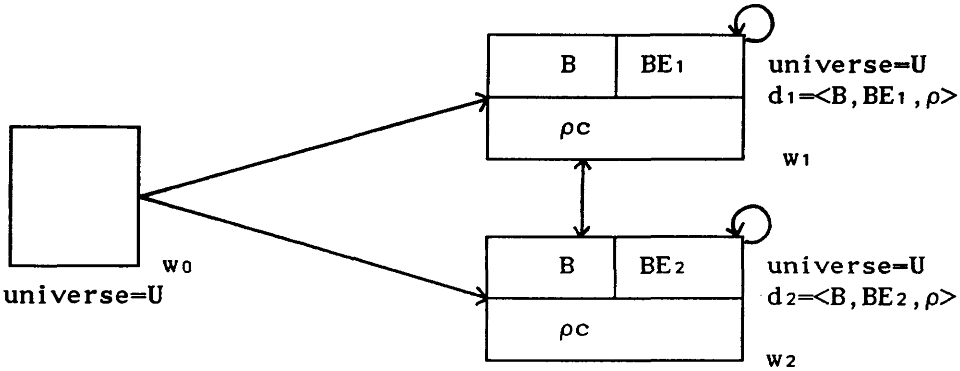


Figure 4.4.3. Belief worlds diagram of the introspective deduction model.
The tuple $\langle U,D \rangle$ specifies the model, where U is the universe of individuals and $D=\{d_1,d_2 \ldots d_n\}$ such that $d_1 \ldots d_n$ are the alternative belief worlds of the agent.

Definition 4.4.2. First order rule for plausible deduction.

If $N\gamma_1(x_1) \notin B$ and $\gamma_2(x_2)\ldots\gamma_n(x_n) \in B$

and $CK\gamma_1(x_1)K\gamma_2(x_2)\ldots\gamma_n(x_n)\ \omega(y) \in \rho c$

then $\omega(y) \in BE$ and $\langle plw \rangle\, \omega(y) \in B$

where γ_i and ω denote predicates.

Completing the logical model is a greater problem in first order logics than in propositional logic systems. If the Herbrand universe (the set of ground terms which may be substituted for a variable) is infinite then an infinite amount of negative information must be generated to complete the model. This problem is overcome if the test for plausibility $N\gamma_1(x_1) \notin B$ may only be applied where a ground substitution has been found for x_1. The restricted rule of inference is defined as follows:

Definition 4.4.3. D45 rule for plausible deduction.

If $\gamma_2(x_2)\ldots\gamma_n(x_n) \in B$

and $\{g/x_1\}\gamma_1(x_1)= \gamma_1(g)$

where g is a ground substitution for x_1

and $N\gamma_1(g) \notin B$

and $CK\gamma_1(x_1)K\gamma_2(x_2)\ldots\gamma_n(x_n)\ \omega(y) \in \rho c$

then $\omega(y) \in BE$ and $\langle plw \rangle\, \omega(y) \in B$

This rule derives a subset of the sentences which may be derived from consideration of K45 model theory. The amount of negative information required to complete the model is reduced. Consider a rule such as CK P(x) Q(x) R(x). If there is no information about which individuals have the property P then for all individuals 'a','b','c' which have the property Q it is plausible that these individuals have the property R. By definition 4.4.3. the ground terms 'a','b','c' may be substituted for x in the sentence P(x) as $Q(x) \epsilon B$ under the substitutions {a/x},{b/x},{c/x}. The same derivations would be made using the rule of Definition 4.4.2. as 'a','b','c' are the only substitutions for x in Q(x) for which $Q(x) \epsilon B$. More generally, the terms which may be substituted for x in P(x) must be obtained from ground substitutions of other predicates which form conditions of the rule. This restricts the choice of substitution to a 'typical' set of individuals. If 'e' could not reasonably be substituted for x in Q(x) then it is not necessary to define P(e) to be false as the truth of P(e) will never be tested (let Q be the property of being an integer and 'e' stand for the planet Mars). Consequently the amount of negative information which must be generated is reduced.

A more general version of the plausible rule of inference may be defined by allowing more than one condition of a deduction rule to be satisfied by the test for plausibility. A rule of plausible deduction Dp is added to the K system. The resulting modal system does not correspond to a classical system.

Definition 4.4.4. Dp rule for plausible deduction.

If for all $\gamma i(xi)$

i) $\gamma i(xi) \epsilon B$ or

ii) $\{g/xi\}\gamma i(xi) = \gamma i(g)$ and $N\gamma i(g) \not\epsilon B$

where g is a ground substitution for xi

and $CK\gamma_1(x_1)K\gamma_2(x_2)\ldots\gamma n(xn)\ \omega(y)\ \epsilon\ \rho c$

then $\omega(y) \epsilon BE$ and $<plw>\ \omega(y)\ \epsilon B$

The assumption is made that if γ_1 and γ_2 are plausible, and they are the conditions of a deduction rule ρ_1, then they may be true in the same plausible world. This is reasonable unless $N\gamma_2(x_2)$ may be inferred from $\gamma_1(x_1)$, consequently the conditions of ρ_1 could never be satisfied. The rule base must not contain rules with contradictory conditions.

4.4.5 Conclusions

The method of plausible reasoning is limited to deriving plausible atomic sentences. It is to be expected that, when the Dp inference rule is employed, the number of plausible interpretations of a set of sentences will increase as the number of known facts decreases.

Non-standard logics [Section 3.5.3] employ a consistency operator. This operator referrs to the whole of what is believed. There may be several extensions of a set of sentences expressed in a non-standard logic. The theory of plausible belief assumes that all sentences have the status of belief. If the database is incomplete there may be several consistent belief worlds. These worlds are identified through introspective reasoning. There is no consistency operator as such, however plausible sentences are derived using introspective deduction rules which make reference to the whole set of beliefs. Plausible deductions are distinguished from ordinary beliefs. Non-standard logics do not make a distinction in this way. The first order theory of plausible belief assumes that the only individuals which have a specific property are those which can be inferred to do so, on the basis of all currently known facts. Predicate circumscription takes a similar approach [Section 3.5.2.].

The theory of plausible belief allows an agent to derive conclusions from incomplete information through introspective reasoning. The theory is analagous to a classical modal system and has efficient computational properties through the use of the deduction model. Implicit relationships between sentences must be made explicit by defining new deduction rules to allow negative information to be derived.

References

1. Moore R.C. Possible World Semantics For Autoepistemic Logic. Proceedings of the Non Monotonic Workshop, New Paltz, New York, A.A.A.I. 1984 :344–354

2. Moore R.C. Autoepistemic Logic. in Smets P. et al (eds) Non–Standard Logics For Automated Reasoning. Academic Press, New York 1988 :125–143

4.5 The Use Of Plausible Reasoning In An Expert System

4.5.1. Introduction.

The theory of plausible reasoning was developed with the aim of extending first order theories in belief logic. The expert system rules defined in Section 4.3.3. are such a theory. The use of a plausible deduction rule presumes that the knowlege model is complete, in the sense defined in Section 4.4.3. This approach is not appropriate for numeric concepts as it is not practical to explicitly define all numbers to be distinct. In refining the expert rule base high level symbolic concepts replaced numerical definitions of predicted normal regions. The result was a set of deduction rules in which only genuinely symbolic concepts were defined. When the rule base is in this form it is possible to explicitly define the relationships between (ground) atomic sentences and consequently for the belief set to become complete. It is not necessary to define all negative facts as the plausible rule of inference does not allow arbitrary substitution of terms for variables.

In order to test the theory, a refined expert system rule base [Section 4.3] was completed, and the plausible inferences examined as the number of facts in the database was reduced. This method of evaluation allows the plausible facts derived from incomplete information to be compared with the true conclusions drawn from the completely specified situation.

4.5.2. Plausible Deductions from Incomplete Information.

The rule base used in this experiment consisted of 62 rules, 25 of which were deduction rules. The rules included those which calculate the predicted normal values of pulmonary function measurements and those for the analysis of lung sound.

The belief set of the expert agent was calculated to a depth of search of 6 (6 applications of hyperresolution). The inference rule (D45 or Dp) was then applied by selecting each rule in turn and generating all possible inferences. The use of forward chaining ensures that all beliefs are represented in the database, therefore determining whether an atomic formula is believed or not is decided by searching for that formula.

The rule base was completed by defining 16 additional deduction rules. The majority were concerned with defining the sound ranges to be exclusive, the following rule was typical:

<expert> ∀x∀y C sound.range(x, y, below lower bound)

N sound.range(x, y, normal range)#

This rule states that for the measurements F50 or F85 (y), made at any position on the chest wall (x), if the measurement is below the predicted lower bound then it is not in the normal range. This rule makes the fact that the sound ranges were defined to be exclusive and do not overlap explicit.

The total number of deduced facts (including plausible ones) increased from 99 before the use of Dp and the completion of ρ to 135 afterwards. The database typically consists of 32 facts about the age, height etc of each subject.

Three cases were investigated employing the rule Dp.

Case 1. No facts were deleted.

Case 2. The fact that breath sound was observed to be quiet at the right upper (ru) recording site was deleted from the database.

Case 3. The sentence defining the measurement of F85 at the ru location was deleted from the database.

The same cases were investigated employing the rule D45. For clarity they will be referred to as cases 4,5 and 6.

Results

Case 1. The sound spectrum at the upper right recording site was concluded to be of normal shape. No plausible sentences were deduced.

Case 2. Two plausible sentences were deduced:

i. <expert><plw> sound.spectrum (ru, normal)

It is plausible that the sound spectrum at the ru location is normal

ii. <expert><plw> conclude (insignificant, crackling, ru)

It is plausible that there is an insignificant degree of crackling at the ru location

As there was no information as to the presence or absence of crackles it was plausible to assume their presence and to conclude that the degree would be insignificant or to assume their absence and to conclude that the sound spectrum would be normal. In fact these assumptions are mutually exclusive.

Case 3. Four plausible beliefs were derived. Sentences i. and ii. were again deduced, by assuming that the F85 lies in the predicted normal region in addition to the assumption about the presence of crackles. The following sentence were derived:

iii. <expert><plw> sound.spectrum (ru,moderate.shift.to.high.frequency)

It is plausible that the sound spectrum at the ru location shows a moderate shift to high frequency.

iv. <expert><plw> conclude (mild, crackling, ru)

It is plausible that there is a mild degree of crackling at the ru location.

These sentences were deduced by assuming that F85 lies above the predicted normal region in addition to assuming the absence/presence of crackles.

Case 4. No plausible sentences were deduced.

Case 5. The same two plausible sentences (i and ii) were deduced as in Case 2.

Case 6. No plausible sentences were deduced. Each rule which allows sentences i.– iv. to be derived in case 3. does so by making two plausible assumptions. The inference rule D45 allows only one plausible assumption to be made to satisfy the conditions of each rule selected. Therefore when the amount of data decreases no sound inferences can made.

Analysis

The true K45 structure can be enumerated for the cases examined above. Where no sentences are removed from the database and the deduction rules are complete there is only one belief world. Where one sentence is removed there are 3 alternative belief worlds which are distinguished by the assignment of the value true to the following facts:

World 1:	breath sound was observed
	the spectrum is assessed as a breath sound spectrum
	the sound spectrum is normal
World 2:	breathing was quiet
	the spectrum is assessed as a breath sound spectrum
	the sound spectrum is normal
World 3:	crackles were observed
	the spectrum is assessed as containing crackles
	the degree of crackling is insignificant

When a second sentence is removed from the database 6 belief worlds may be distinguished:

Worlds 1 and 2:	F85 in normal range
	the sound spectrum is normal
Worlds 3 and 4:	F85 above upper bound
	sound spectrum shows a moderate shift to high frequency
World 5:	F85 in normal range
	the degree of crackling is insignificant
World 6:	F85 above upper bound
	the degree of crackling is mild

In cases 2 and 5 where one sentence was deleted, both inference rules identify two plausible conclusions. In fact one of these conclusions is included in two worlds, 1 and 2, and the other in world 3. Neither D45 nor Dp can identify the observed characteristic which distinguishes worlds 1 and 2 (i.e. whether breath sound was observed or breathing was quiet). When a second sentence is deleted rule Dp identifies the four main classes of worlds of the true K45 structure. If the set of expert rules were more complex then plausible sentences which are always true in the same worlds would be derived but the rule Dp could not show this to be the case.

4.5.3. Conclusions

The use of Dp reflects the actual modal structure better than the use of the D45 inference rule. Neither rule was designed to identify the full modal structure. At present the expert system user must find the assumptions which the inference rule has used by examining the rule base and database. With some modifications to the software, the examination could be done interactively which would enable the user to explore the alternative default assumptions and their consequences. Such an approach may avoid the need for the full K45 structure to be enumerated. The techniques of completing the rule base and restricting the test for plausibility to ground terms may aid the development of efficient methods of calculating the full K45 structure.

The theory of plausible reasoning is based on the K45 system of modal logic whereas logics which employ consistency operators (Default Logic [1] and Autoepistemic Logic [2]) are not based on a classical modal system [Section 3.5.]. Let CONS be the consistency operator; then sentences such as CONS P $\supset$ Q are interpreted as 'if it is consistent to assume P then conclude Q. The sentences CONS P and NP are contradictory. The CONS operator is not equivalent to M (possibility) as MP and NP are not contradictory, even in S5. Another strategy may be to add an axiom such as Q $\supset$ LQ to the K modal system in order to make NP and MP contradictory. This is unwise as the axiom MQ $\supset$ Q follows from Q $\supset$ LQ, that is, the distinction between LP, MP and P would be lost. The CONS operator cannot be considered to be equivalent to L (necessity) as LP and NP are only contradictory in systems which include the axiom LQ $\supset$ Q and in such systems assuming LP is equivalent to assuming P. In plausible reasoning sentences such as P $\supset$ Q are qualified by the belief operator which is equivalent to L. If it is consistent to assume

P (NP is not believed, NLNP ≡ MP) then deduce Q where both P and Q have the status of being true in at least one belief world. P and Q are qualified by the possibility operator. In the K45 modal system sentences which are believed such as P ⊃ Q are true in all belief worlds and sentences which are possibly true are possibly true in all belief worlds. In each belief world there is a set of certain beliefs and a set of possible (or plausible) beliefs.

The incorporation of plausible reasoning into an expert system brings any gaps in the database to the attention of the user. The user may then investigate the reasons for the plausible deductions. Ordinary belief logic does not provide any mechanism for such introspective reasoning. Research into the application of methods for reasoning with incomplete information to real problems is at an early stage [3].

References

1. Reiter R. A Logic For Default Reasoning. A.I. 13 1980 :81–132

2. Moore R.C. Autoepistemic Logic. in Smets P. et al (Eds) Non–Standard Logics For Automated Reasoning. Academic Press, London 1988 :75–94

3. Reinfrank M. et al (Eds) Proc. Non Monotonic Reasoning, 2nd International Workshop (published as Lecture Notes In Artificial Intelligence No.346 Springer Verlag 1989)

See Konolige K. Hierarchic Autoepistemic Theories For Non Monotonic Reasoning: Preliminary Report in Reinfrank M. et al for an interesting application.

4.6 A Note On Logic And Uncertainty

First order logic (f.o.l.) allows properties to be assigned to sets of individuals. Sentences are constructed with the truth functional operators and the universal and existential quantifiers are employed to define the status of variables. Proofs can be constructed to show that f.o.l. has the properties of completeness, consistency and to show that the system is semi-decidable. First order logic is an abstract mathematical system which may be used to express ideas and to make deductions.

Much of human reasoning is inexact or uncertain. Uncertainty arises because of the random nature of the world or because knowledge is expressed in vague terms. First order logic insists that a proposition is either true or false, contradiction is not permitted. Several theories have been developed which allow a degree of belief to be associated with the conclusion of a logical rule. The use of probability theory often contains the assumption that the conditions of a rule occur independently of each other (the joint probability is not calculated). This assumption can lead to errors [1]. Certainty theory overcomes this problem. A value between -1 (false) and $+1$ (true) is associated with all propositions. The numerical values are propagated in an intuitive but ad-hoc fashion [2].

Let the certainty of a proposition P be 0.8. The certainty factor is greater than 0 hence P is more true than false. The certainty factor is not equal to 1 therefore P is not entirely true, P is both true and false. First order logic does not allow the truth assignment of a proposition to be contradictory. The use of a numerical theory to assign certainty factors is necessary because contradiction and uncertainty are part of human reasoning but f.o.l. is unable to express contradiction.

The problem of making deductions from incomplete information is related to the problem of uncertain reasoning. Default or autoepistemic rules define what assumptions are to be made when the database is incomplete [Section 3.5]. These rules are explicitly expressed. The problem is to identify a consistent extension or theory for a set of sentences. There may be no such theory or several competing theories. An alternative approach is to express sentences in the modal system K45.

Sentences which are known to be true are true in all worlds, sentences whose truth value is unknown may be true in some world. A system of possible worlds is constructed, each containing a consistent set of sentences comprising sentences which are necessarily true and a set which are possibly true. Each world represents a competing theory consistent with the known facts and rules. If the database is complete only one theory may be constructed.

The problem of inexact reasoning may be approached from a similar perspective. If there is uncertainty about the validity of a rule or about the conditions on which a conclusion depends, then such rules can be given the status of being possibly true (as opposed to being necessarily true). A system of possible worlds can be constructed to show the impact of assuming one rule to be true and another false etc. Each world must be consistent and represents a competing theory.

In f.o.l. a consistent set of sentences can only have one model (or theory). A numerical measure of certainty is used to suppress competing theories and to select the most likely solution. This method does not make the pattern of logical inference completely explicit. Numerical values and weights can obscure the steps of logical inference.

Examination of the competing theories which result from the use of modal operators as outlined above should allow an expert to clarify the rulebase. For example a rule which is possibly true may be eliminated or strengthened when an expert has examined its role in determining the competing theories. This approach is opposite to that of Ginsberg [3] who improved the performance of an expert system by fine tuning the numerical weights in an inference network, a procedure which must further obscure the logical model of the domain which an expert can construct.

References

1. White A.P. Inference Deficiencies In Rule Based Expert Systems. in Bramer M.A. (Ed) Research And Development In Expert Systems. Cambridge University Press 1984 :39–50

2. Frost R.A. Introduction To Knowledge Base Systems. Collins, London 1986

3. Ginsberg A. Automatic Refinement Of Expert Systems Knowledge Bases. Pitman, London 1988

4.7. Conclusions

The use of belief logic in expert system design allows modal concepts to be expressed and reasoned with. The procedure of refining the logical model of the area of expertise can be aided by the use of belief logic. The framework of the deduction model can be used to implement simple theories of time in addition to reasoning about belief. The problem of reasoning with incomplete information can be tackled by performing introspective reasoning which requires an agent to reason about its own knowledge. The deduction model of belief can represent reasoning about the beliefs of another agent and the special case of introspective reasoning. The axioms which define the introspective deduction model must be implemented so as to aviod generating an infinite number of false sentences. Plausible reasoning presents an insight into this problem and a solution to it.

The simple model of belief is analagous to the K system of modal logic and the introspective model of belief is analagous to the K45 system. These modal systems have well understood model theories and proof procedures. The expert systems developed in this chapter inherit the clear semantics of modal logic. The work presented in this chapter has used modal ideas in a creative way and has not been limited to the implementation of a theorem proving program (although this was an important task).

Belief logic has many potential applications. The deduction model of belief can represent the beliefs of many agents and this capability could be used to implement an expert system involving several expert agents. Reasoning about belief is necessary if a tutoring program is required to make deductions about the knowledge of a user. Another application of the deduction model is to represent the beliefs of a robot agent which must co-operate with other robots and consequently reason about their beliefs and actions. The new resolution methods presented in this chapter may be used to implement the deduction model so that these applications and others can be explored.

Chapter 5 Directions For Future Work

Directions For Future Work

Lung Sound Research

The investigation of the changes in lung sound produced by reversible changes in the lung could be a useful field of research. Exercise and the administration of allergens can produce significant changes in the lungs and consequently may affect the spectrum of lung sound produced. The changes in the lung can be measured by the forced expiratory tests. In addition airflow at the mouth should be monitored continuously and lung sound should be recorded simultaneously from several sites on the chest wall. Such investigations could give insights into the changes in the lungs produced by the stimuli and their effects on the power spectrum of lung sound.

The study of lung sound amongst groups of subjects could also employ the measurement technique outlined above whereby airflow and sound are simultaneously recorded. In addition, a reliable method of compensating the recorded sound spectrum for the chest wall characteristics would be desirable in order to increase the accuracy of the results. The relationship between FEV1 and sound recorded over several lobes of the lung could be further explored by such experiments.

If the diameter of the airways could be measured directly, by a scanning technique for example, or if the airflow in a lobe could be measured then the relationship between these parameters and the sound spectrum could be proven conclusively.

Expert System Development

The Inspire expert system is not in clinical use at present. Future development of this system should aim to improve the user interface. This can be achieved by providing a menu to guide the user and a natural language interface capable of translating the user's queries into belief logic and of translating the results from belief logic into English. These features would allow a more natural dialogue between user and computer and is the case at present. The author believes that the data and rules files should not be translated into a natural language as the precision of the logical formula would be lost. A user who wishes to modify the rule base must be familiar with belief logic.

The time taken for the computer to make inferences from a set of rules increases as the number of rules increases. At present the inference procedure is written in Turbo Prolog. A more efficient inference procedure could be written in a version of Prolog which allows the nesting of predicates or allows database predicates to contain variables as then the unification mechanism of the Prolog system could be used (at present the unification algorithm is programmed in Turbo Prolog because this language lacks the above mentioned features). Alternatively the proof procedure could be written in C.

Further Study of Logic Systems

The use of belief logic in applications such as expert system design and robotics aids the study of formal systems for knowledge representation by showing practical applications of those systems.

Worlds models can be used to represent changes in time from one world to the next, to represent belief and to represent the impact of actions on belief. Worlds models can be a powerful way to describe changes in the real world or in the beliefs of an agent. New formal systems of logic based on worlds models could replace intuitive theories of reasoning about time and reasoning where information may be incomplete or may be updated.

Appendix 1

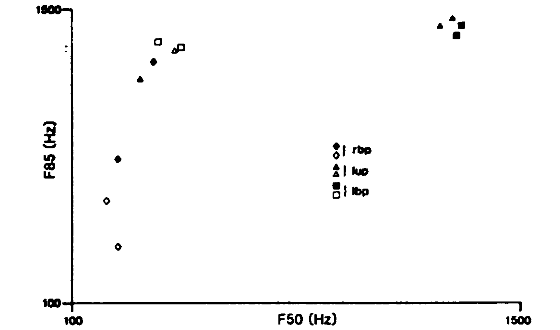
Anderson K. Aitken S. Macleod J.E.S. Moran F.

Lung– Sound Transmission And Reliability Of Chest Signs.

Lancet 1988 ii:228

LUNG-SOUND TRANSMISSION AND RELIABILITY OF CHEST SIGNS

SIR,—Dr Spiteri and colleagues (April 16, p 873) find considerable disagreement between physicians in interpretation of chest signs, with a 28% failure to achieve the correct diagnosis. Considering the factors involved in the production of chest signs, we suggest that this result is probably as good as can be expected. The development of chest X-ray was considered by some[1] to have reduced chest auscultation to a perfunctory ritual and thereby may have influenced the clinical examination of the chest when certain disorders were suspected. However, as Spiteri and colleagues conclude, abnormal breath sounds produced by airway narrowing (wheezes) correlate consistently with the forced expiratory volume in 1 s.[2] The difficulties encountered by some clinicians in assimilating clinical signs from chest examination might be explained by inexperience but consideration of the causes of these signs offers a more satisfactory explanation of what is and is not a "consistent" or "good and reliable" sign. Wheeze or stridor result from altered sound generation because of airways narrowing, which presumably does not influence transmission of these noises through the chest. However, if a structural abnormality occurs in lung tissue or in the pleural or chest walls, airway sounds may be transmitted to the chest wall to produce some of the less common clinical signs, such as bronchial breathing or whispering pectoriloquy.



Frequency spectrum/amplitude graph.

The two variables are taken from 100–1500 Hz to represent lower frequency range (F50 = median frequency of spectral energy) and higher frequency range (F85 = frequency below which 85% of spectral energy lies). Closed symbols = pre-pleural aspiration breath sounds; open symbols = post-aspiration; rbp = right base posteriorly; lup = superior border of the scapula; lbp = left base posteriorly.

We have recorded breath sounds from a thin 14-year-old boy with what was clinically considered to be bilateral pneumonic consolidation because of bronchial breathing at both lung bases. The chest X-ray suggested a normal right lung with consolidation and effusion in the left hemithorax. Ultrasound of the thorax was done to assist chest aspiration and a thin rim of pleural fluid encasing the left lung was detected. Computerised tomography, done primarily to assess the lung parenchyma, showed normal right pleura and lung, with predominantly left lower lobe consolidation and the pleural abnormality found on ultrasound. The breath sounds were recorded from both bases posteriorly before and after pleural aspiration, and a frequency spectrum produced by computer analysis and fast Fourier transformation.

The abnormal lung sounds at both lung bases were altered by removal of the fluid (figure). The left lung had a residue of high frequency sounds representing bronchial breathing which were also initially present on the right only to disappear after aspiration. The results suggest that the thin rim of pleural fluid on the left was solely responsible for producing the lung sound abnormality detected from the right lung.

Error is possible in all the components of the diagnostic process[3] and, as Spiteri et al suggest, some clinical findings in chest examination seem less reliable than others. Our case suggests that signs that arise because of pathology which influences the transmission of sound generated in the airways are most often those found to be inconsistent or misleading.

Department of Respiratory Medicine,
Glasgow Royal Infirmary,
Glasgow G31 2ES;
and Department of Electronic and
Electrical Engineering,
Glasgow University

K. ANDERSON
S. AITKEN
J. E. S. MACLEOD
F. MORAN

1. Forgacs P. Lung sounds. *Br J Dis Chest* 1969; 63: 1–12.
2. Pardee NE, Martin CF, Morgan EH. A test of the practical value of estimating breath sound intensity. *Chest* 1986; 70: 341–44.
3. Koran LM. The reliability of clinical methods, data and judgments. *N Engl J Med* 1975; 293: 695–701.

Appendix 2

Proof of resolution rule H1.

To prove that <B> Pn is a consequence of sentences 1–4 by the tableau proof method: negate the goal sentence and show that the tableau closes.

1.	T $<B>$CK P_1 K P_2 P_3 Pn	deduction rule
2.	T $<B>P_1$	assumption
3.	T $<B>P_2$	assumption
4.	T $<B>P_3$	assumption
5.	F $<B>$Pn	negation of goal

open a view for agent <B>

```
6.   F Pn                                   from 5.
7.   T P1                                   from 2.
8.   T P2                                   from 3.
9.   T P3                                   from 4.
10.  T CK P1 K P2 P3 Pn                     from 1.
11.  F K P1KP2P3             12. T Pn       tableau splits
13. F P1  14. F KP2P3             *         tableau splits
     *      15. F P2   16. F P3             tableau splits
                 *          *
```

The tableau closes. The clausal form of sentences 1–4 is as follows:

1.	$<B>[L_1,L_2,L_3,L_4]$
2.	$<B>[L_1']$
3.	$<B>[L_2']$
4.	$<B>[L_3']$
deduce	$<B>[L_4]$

where $L_1 = NP_1$, $L_2 = NP_2$, $L_3 = NP_3$, $L_4 = P_4$, and
$L_1' = P_1$, $L_2' = P_2$, $L_3' = P_3$

Proof of resolution rule H2.

To prove that $N<B>NPn$ is a consequence of sentences 1–4 by the tableau proof method: negate the goal sentence and show that the tableau closes.

1.	$T <B> CK P_1 K P_2 P_3 Pn$	deduction rule
2.	$T <B> P_1$	assumption
3.	$F <B> NP_2$	assumption
4.	$T <B> P_3$	assumption
5.	$F N<B> NPn$	negation of goal
6.	$T <B> NPn$	from 5.

open a view for agent $<B>$

7.	$F\ NP_2$	from 3.
8.	$T\ P_2$	from 7.
9.	$T\ NPn$	from 6.
10.	$F\ Pn$	from 9.
11.	$T\ P_1$	from 2.
12.	$T\ P_3$	from 4.
13.	$T\ CK P_1 K P_2 P_3 Pn$	from 1.
14. $F\ K P_1 K P_2 P_3$	15. $T\ Pn$	tableau splits
16. $F\ P_1$ 17. $F\ KP_2P_3$	*	tableau splits
* 18. $F\ P_2$ 19. $F\ P_3$		tableau splits
* *		

The tableau closes. The clausal form of sentences 1–4 is as follows:

1. $<B>[L_1, L_2, L_3, L_4]$
2. $<B>[L_1']$
3. $N<B>N[L_2']$
4. $<B>[L_3']$

deduce $N<B>N[L_4]$

where $L_1 = NP_1$, $L_2 = NP_2$, $L_3 = NP_3$, $L_4 = P_4$, and
$L_1' = P_1$, $L_2' = P_2$, $L_3' = P_3$

Appendix 3

Solutions of the Wise Man Puzzle

The Wise Man Puzzle is often used to demonstrate the ability of modal logic to represent and solve problems involving belief. The deduction model is adopted where the operators L and M are replaced by a single belief operator, for example <agentA> P states that agentA believes P to be true. More than one agent may be defined and agents may have beliefs about each other. This aspect is highlighted by the Wise Man puzzle, which can be stated as follows:

'A King wishes to determine which of his three wise men is the wisest. He arranges them in a circle so that they can see and hear each other and tells them that he will put a white or a black spot on each of their foreheads, but at least one spot will be white. In fact all three spots are white. After a while the wisest announces that his spot is white. How does he know ?'

We will consider the problem where two men are involved.

take White–A to stand for 'agentA has a white spot'

and White–B to stand for 'agentB has a white spot'

where agentA and agentB are the wise men

The following rules define the problem, they are stored in the knowledge base in clausal form.

Both agents have white spots.

1)White–A

2)White–B

Each agent believes that least one spot is white.

3)<agentA> A White–A White–B

4)<agentB> A White–A White–B

Each agent believes the other believes this too.

5)<agentA><agentB> A White–A White–B

6)<agentB><agentA> A White–A White–B

If agentA has a white spot then agentB will see it.

7)C White–A <agentB> White–A

8)C NWhite–A <agentB> NWhite–A

9)C White–B <agentA> White–B

10)C NWhite–B <agentA> NWhite–B

Each agent knows this.

11)<agentB> C White–B <agentA> White–B

12)<agentB> C NWhite–B <agentA> NWhite–B

13)<agentA> C White–A <agentB> White–A

14)<agentA> C NWhite–A <agentB> NWhite–A

Neither agent can deduce the colour of his own spot with the above information. If agentA says that he does not know the colour of his spot then we may add this fact to agentB's beliefs.

15)<agentB> N<agentA> White–A

Now it can be shown that agentB can deduce that the colour of his spot is white. The aim is to prove <agentB> White–B. The prolog programs are able to prove this, giving the following derivation by backwards chaining:

goal> <agentB> White–B

negate goal and convert to clausal form: *N<agentB>NWhite−B*

resolve with 12) new goal> *<agentB><agentA>NWhite−B*

resolve with 6) new goal> *<agentB><agentA>White−A*

resolve with 15) (*<agentB>N<agentA>NWhite−A*) to get □ (the empty clause).

The goal can be derived using hyperresolution rules H1 and H2 as follows:

16) *<agentB>N<agentA>White−B*	from 6) and 15) using H2
17) *<agentB>White−A*	from 7) and 1) using H1
18) *<agentA>White−B*	from 9) and 2) using H1
19) *<agentB>White−B*	from 12) and 16) using H1

The goal sentence 19)<agentB> White− B has been derived.

Note: the deduction rule model is not used in this example.

Appendix 4

Inspire Rule Base in Belief Logic

```
101
Vx Vy Vz C K < expert > sex ( male )
    K age ( x )
    K height ( y )
    equation ( −0.022, x, 5.29, y, 0, 0, −3.09, z )
    prednormal ( vc, z )#
102
Vx Vy Vz C K < expert > sex ( female )
    K age ( x )
    K height ( y )
    equation ( −0.024, x, 4.44, y, 0, 0, −2.59, z )
    prednormal ( vc, z )#
103
Vw Vx Vy Vz C K < expert > sex ( male )
    K age ( x )
    K height ( y )
    K weight ( z )
    equation ( 0.017, x, 5.83, y, −0.041, z, −4.28, w )
    prednormal ( frc, w )#
104
Vx Vy Vz C K < expert > sex ( female )
    K height ( x )
    K weight ( y )
    equation ( 5.64, x, −0.031, y, 0, 0, −4.95, z )
    prednormal ( frc, z )#
105
Vx Vy Vz Vz1 C K < expert > sex ( male )
    K age ( x )
    K height ( y )
    K weight ( z )
    equation ( 0.024, x, 2.18, y, −0.017, z, −1.69, z1 )
    prednormal ( rv, z1 )#
106
Vx Vy Vz C K < expert > sex ( female )
    K age ( x )
    K height ( y )
    equation ( 0.008, x, 2.95, y, 0, 0, −3.76, z )
    prednormal ( rv, z )#
107
Vx Vy Vz C K < expert > sex ( male )
    K height ( x )
    K weight ( y )
    equation ( 7.61, x, −0.019, y, 0, 0, −4.73, z )
    prednormal ( tlc, z )#
108
Vx Vy Vz C K < expert > sex ( female )
    K age ( x )
    K height ( y )
    equation ( −0.017, x, 7.37, y, 0, 0, −6.35, z )
    prednormal ( tlc, z )#
```

```
109
Vx Vy Vz C K < expert > sex ( male )
    K age ( x )
    K weight ( y )
    equation ( 0.33, x, -0.14, y, 0, 0, 23.4, z )
    prednormal ( rvtlc, z )#
110
Vx Vy Vz C K < expert > sex ( female )
    K age ( x )
    K height ( y )
    equation ( 0.28, x, 27.0, y, 0, 0, -28.0, z )
    prednormal ( rvtlc, z )#
111
Vx Vy Vz C K < expert > sex ( male )
    K age ( x )
    K height ( y )
    equation ( -0.036, x, 3.78, y, 0, 0, -1.1, z )
    prednormal ( fev, z )#
112
Vx Vy Vz C K < expert > sex ( female )
    K age ( x )
    K height ( y )
    equation ( -0.031, x, 2.94, y, 0, 0, -0.59, z )
    prednormal ( fev, z )#
113
Vy Vz C K < expert > sex ( male )
    K age ( y )
    equation ( -0.373, y, 0, 0, 0, 0, 91.8, z )
    prednormal ( fevfvc, z )#
114
Vy Vz C K < expert > sex ( female )
    K age ( y )
    equation ( -0.261, y, 0, 0, 0, 0, 92.1, z )
    prednormal ( fevfvc, z )#
115
Vy Vz C K < expert > sex ( male )
    K age ( y )
    equation ( -0.011, y, 0, 0, 0, 0, 2.43, z )
    prednormal ( kco, z )#
116
Vy Vz C K < expert > sex ( female )
    K age ( y )
    equation ( -0.004, y, 0, 0, 0, 0, 2.24, z )
    prednormal ( kco, z )#
117
Vv Vw Vx Vy Vz C K prednormal ( v, w )
    K < expert > sex ( x )
    K normallimit ( x, v, y )
    add ( w, y, z )
    predupperbound ( v, z )#
118
Vv Vw Vx Vy Vy1 Vz C K prednormal ( v, w )
    K < expert > sex ( x )
    K normallimit ( x, v, y )
    K multiply ( y, -1.00, y1 )
    add ( w, y1, z )
    predlowerbound ( v, z )#
```

```
119
normallimit ( male, vc, 1.1 )#
120
normallimit ( female, vc, 0.88 )#
121
normallimit ( male, frc, 1.23 )#
122
normallimit ( female, frc, 0.9 )#
123
normallimit ( male, rv, 0.83 )#
124
normallimit ( female, rv, 0.7 )#
125
normallimit ( male, tlc, 1.47 )#
126
normallimit ( female, tlc, 1.06 )#
127
normallimit ( male, rvtlc, 8.6 )#
128
normallimit ( female, rvtlc, 11.0 )#
129
normallimit ( male, fev, 1.1 )#
130
normallimit ( female, fev , 0.79 )#
131
normallimit ( male, fevfvc, 14.38 )#
132
normallimit ( female, fevfvc, 10.88 )#
133
normallimit ( male, kco, 0.54 )#
134
normallimit ( female, kco, 0.98 )#
135
< expert > C K notlow ( fev )
    fevfvc.in.range ( a )
    degree ( mild, airwaysobstruction )#
136
< expert > C K notlow ( fev )
    A fevfvc.in.range ( b )
    fevfvc.in.range ( c )
    degree ( moderate, airwaysobstruction )#
137
< expert > C K notlow ( fev )
    fevfvc.in.range ( d )
    degree ( moderatelysevere, airwaysobstruction )#
138
< expert > C K notlow ( fev )
    fevfvc.in.range ( e )
    degree ( severe, airwaysobstruction )#
139
< expert > C K low ( fev )
    A fevfvc.in.range ( a )
    fevfvc.in.range ( b )
    degree ( major, airwaysobstruction )#
```

```
140
< expert > C K low ( fev )
     A fevfvc.in.range ( c )
     A fevfvc.in.range ( d )
     fevfvc.in.range ( e )
     degree ( severe, airwaysobstruction )#
141
define.fevfvc.range ( 1.0, 0.8, a )#
142
define.fevfvc.range ( 0.8, 0.65, b )#
143
define.fevfvc.range ( 0.65, 0.6, c )#
144
define.fevfvc.range ( 0.6, 0.45, d )#
145
define.fevfvc.range ( 0.45, 0, e )#
146
Vx Vy Vy1 Vy2 Vz Vz1 Vz2
     C K measured ( fevfvc, x )
     K predlowerbound ( fevfvc, y )
     K define.fevfvc.range ( z, z1, z2 )
     K multiply ( y, z1, y1 )
     K multiply ( y, z, y2 )
     K lessthanorequal ( y1, x )
     lessthan ( x, y2 )
     < expert > fevfvc.in.range ( z2 )#
147
< expert > Vw C K degree ( w, airwaysobstruction )
     range ( tlc, aboveupperbound )
     with ( hyperinflation )#
148
< expert > Vw C K degree ( w, airwaysobstruction )
     K range ( rv, aboveupperbound )
     range ( tlc, belowupperbound )
     with ( airtrapping )#
149
Vx Vy Vy1 C K measured ( fev, x )
     K predlowerbound ( fev, y )
     K multiply ( y, 0.7, y1 )
     lessthan ( x, y1 )
     < expert > low ( fev )#
150
< expert > C K range ( fevfvc, belowlowerbound )
     K range ( rvtlc, leupperbound )
     A tlc.in.range ( a )
     tlc.in.range ( b )
     degree ( reduced, lungvolume )#
151
< expert > C K range ( tlc, belowlowerbound )
     K range ( fevfvc, belowlowerbound )
     range ( rvtlc, aboveupperbound )
     degree ( reduced, lungvolume )#
152
< expert > C K range ( fevfvc, belowlowerbound )
     K tlc.in.range ( c )
     range ( rvtlc, leupperbound )
     degree ( major, restrictivedefect )#
```

```
153
< expert > C range ( fevfvc, abovelowerbound )
    noevidence ( airwaysobstruction )#
154
< expert > C K noevidence ( airwaysobstruction )
    range ( tlc, aboveupperbound )
    evidence ( hyperinflation )#
155
< expert > C K noevidence ( airwaysobstruction )
    K range ( rv, aboveupperbound )
    range ( tlc, belowupperbound )
    evidence ( airtrapping )#
156
define.tlc.range ( 1.0, 0.9, a )#
157
define.tlc.range ( 0.9, 0.8, b )#
158
define.tlc.range ( 0.8, 0.0, c )#
159
Vx Vy Vy1 Vy2 Vz Vz1 Vz2
    C K measured ( tlc, x )
    K predlowerbound ( tlc, y )
    K define.tlc.range ( z, z1, z2 )
    K multiply ( y, z1, y1 )
    K multiply ( y, z, y2 )
    K lessthanorequal ( y1, x )
    lessthan ( x, y2 )
    < expert > tlc.in.range ( z2 )#
160
< expert > C K noevidence ( airwaysobstruction )
    K range ( rvtlc, belowupperbound )
    tlc.in.range ( a )
    degree ( mild, restrictivedefect )#
161
< expert > C K noevidence ( airwaysobstruction )
    K range ( rvtlc, belowupperbound )
    tlc.in.range ( b )
    degree ( moderate, restrictivedefect )#
162
< expert > C K noevidence ( airwaysobstruction )
    K range ( rvtlc, belowupperbound )
    tlc.in.range ( c )
    degree ( severe, restrictivedefect )#
163
< expert > C K noevidence ( airwaysobstruction )
    K range ( tlc, belowlowerbound )
    range ( rvtlc, geuppperbound )
    degree( reduced, lungvolume )#
164
Vx Vy Vz C K measured ( x, y )
    K predupperbound ( x, z )
    lessthan ( y, z )
    < expert > range ( x, belowupperbound )#
165
Vx Vy Vz C K measured ( x, y )
    K predupperbound ( x, z )
    lessthan ( z, y )
    < expert > range ( x, aboveupperbound )#
```

```
166
Vx Vy Vz C K measured ( x, y )
    K predlowerbound ( x, z )
    lessthan ( y, z )
    < expert > range ( x, belowlowerbound )#
167
Vx Vy Vz C K measured ( x, y )
    K predlowerbound ( x, z )
    lessthan ( z, y )
    < expert > range ( x, abovelowerbound )#
168
Vx Vy Vy1 C K measured ( fev x )
    K predlowerbound ( fev y )
    K multiply ( y, 0.7, y1 )
    lessthanorequal ( y1, x )
    < expert > notlow ( fev )#
169
< expert > C K notlow ( fev )
    K range ( fevfvc, belowlowerbound )
    K range ( fev, belowlowerbound )
    K range ( rv, belowupperbound )
    range ( rvtlc, geupperbound )
    suggests ( obstruction.or.airtrapping, underestimated )#
170
Vx Vy Vy1 C K < expert > notlow ( fev )
    K < expert > range ( fevfvc, belowlowerbound )
    K < expert > range ( fev, belowlowerbound )
    K < expert > range ( rvtlc, aboveupperbound )
    K measured ( fevfvc, x )
    K predlowerbound ( fevfvc, y )
    K multiply ( y, 0.6, y1 )
    lessthanorequal ( y1, x )
    < expert > suggests ( obstruction.or.airtrapping, underestimated )#
171
< expert > C K notlow ( fev )
    K range ( fevfvc, belowlowerbound )
    K range ( fev, belowlowerbound )
    K range ( rv, belowupperbound )
    range ( rvtlc, geupperbound )
    suggests ( get.further.info.from, body.plethsmography )#
172
< expert > C K notlow ( fev )
    K range ( fevfvc, belowlowerbound )
    range ( tlc, belowlowerbound )
    suggests ( get.further.info.from, body.plethsmography )#
173
Vx Vy Vy1 C K < expert > notlow ( fev )
    K < expert > range ( fevfvc, belowlowerbound )
    K < expert > range ( fev, belowlowerbound )
    K < expert > range ( rvtlc, aboveupperbound )
    K measured ( fevfvc, x )
    K predlowerbound ( fevfvc, y )
    K multiply ( y, 0.6, y1 )
    lessthanorequal ( y1, x )
    < expert > suggests ( get.further.info.from, body.plethsmography )#
```

```
174
Vx Vx1 Vx2 Vy Vy1 C K < expert > notlow ( fev )
    K < expert > range ( fevfvc, abovelowerbound )
    K < expert > measured ( fev, x )
    K < expert > measured ( rvtlc, y )
    K predlowerbound ( fev, x1 )
    K predupperbound ( rvtlc, y1 )
    K multiply ( x1, 1.2, x2 )
    K lessthan ( x, x2 )
    lessthanorequal ( y1, y )
    < expert > suggests ( obstruction.or.airtrapping, underestimated )#
175
< expert > C K notlow ( fev )
    K range ( fevfvc, abovelowerbound )
    K range ( rv, aboveupperbound )
    range ( rvtlc, geupperbound )
    suggests ( obstruction.or.airtrapping, underestimated )#
176
< expert > C K notlow ( fev )
    K range ( fevfvc, abovelowerbound )
    K range ( fev, belowlowerbound )
    range ( rvtlc, geupperbound )
    suggests ( get.further.info.from, body.plethsmography )#
177
< expert > C K notlow ( fev )
    K range ( fevfvc, abovelowerbound )
    K range ( tlc, belowlowerbound )
    range ( rvtlc, geupperbound )
    suggests ( get.further.info.from, body.plethsmography )#
178
Vx Vy Vy1 Vy2 C K measured ( x, y )
    K predlowerbound ( x, y1 )
    K predupperbound ( x, y2 )
    K lessthan ( y1, y )
    lessthan ( y, y2 )
    < expert > assign ( x, normalrange )#
179
< expert > C K notlow ( fev )
    K range ( fevfvc, abovelowerbound )
    K assign ( tlc, normalrange )
    range ( rvtlc, aboveupperbound )
    on.dynamic.testing ( no.airwaysobstruction )#
180
< expert > C K notlow ( fev )
    K range ( fevfvc, abovelowerbound )
    K assign ( tlc, normalrange )
    range ( vc, belowlowerbound )
    on.dynamic.testing ( no.airwaysobstruction )#
181
< expert > C K notlow ( fev )
    K range ( fevfvc, abovelowerbound )
    K assign ( tlc, normalrange )
    range ( rv, belowlowerbound )
    on.dynamic.testing ( no.airwaysobstruction )#
```

```
182
Vx Vx1 Vx2 C K < expert > low ( fev )
    K < expert > range ( rvtlc, aboveupperbound )
    K < expert > range ( rv, aboveupperbound )
    K measured ( fevfvc, x )
    K predlowerbound ( fevfvc, x1 )
    K multiply ( x1, 0.7, x2 )
    lessthan ( x2, x )
    < expert > suggests ( define.obstruction.or.trapping, by.body.plethsmography )#
183
< expert > C K low ( fev )
    K range ( rvtlc, abovelowerbound )
    range ( rv, belowupperbound )
    evidence ( airtrapping.underestimated )#
184
< expert > C K low ( fev )
    K range ( rvtlc, abovelowerbound )
    range ( rv, belowupperbound )
    suggests ( get.further.info.from, body.plethsmography )#
185
< expert > C K low ( fev )
    range ( tlc, belowlowerbound )
    suggests ( get.further.info.from, body.plethsmography )#
186
Vx Vx1 C K measured ( kco, x )
    K predlowerbound ( kco, x1 )
    lessthanorequal ( x1, x )
    < expert > degree ( normal, transfer.coefficient )#
187
< expert > C kco.in.range ( a )
    degree ( slightly.reduced, transfer.coefficient )#
188
< expert > C kco.in.range ( b )
    degree ( moderately.reduced, transfer.coefficient )#
189
< expert > C kco.in.range ( c )
    degree ( severely.reduced, transfer.coefficient )#
190
define.kco.range ( 1.0, 0.8, a )#
191
define.kco.range ( 0.8, 0.6, b )#
192
define.kco.range ( 0.6, 0.0, c )#
193
Vx Vy Vy1 Vy2 Vz Vz1 Vz2
    C K measured ( kco, x )
    K predlowerbound ( kco, y )
    K define.tlc.range ( z, z1, z2 )
    K multiply ( y, z1, y1 )
    K multiply ( y, z, y2 )
    K lessthanorequal ( y1, x )
    lessthan ( x, y2 )
    < expert > kco.in.range ( z2 )#
194
Vx Vy Vz C K measured ( x, y )
    K predupperbound ( x, z )
    lessthanorequal ( y, z )
    < expert > range ( x, leupperbound )#
```

```
195
Vx Vy Vz C K measured ( x, y )
    K predupperbound ( x, z )
    lessthanorequal ( z, y )
    < expert > range ( x, geupperbound )#
201
Vx Vy C K measured ( fev, x )
    equation ( -64.51, x, 0, 0, 0, 0, 414.3, y )
    prednormal ( f50, y )#
202
Vx Vy C K measured ( fev, x )
    equation ( -181.04, x, 0, 0, 0, 0, 995.7, y )
    prednormal ( f85, y )#
203
normallimit ( male, f50, 168 )#
204
normallimit ( female, f50, 168 )#
205
normallimit ( male, f85, 284 )#
206
normallimit ( female, f85, 284 )#
207
Vx Vy Vy1 Vz C K sound.measured ( x, y, z )
    K predupperbound ( y, y1 )
    lessthan ( z, y1 )
    < expert > sound.range ( x, y, belowupperbound )#
208
Vx Vy Vy1 Vz C K sound.measured ( x, y, z )
    K predupperbound ( y, y1 )
    lessthan ( y1, z )
    < expert > sound.range ( x, y, aboveupperbound )#
209
Vx Vy Vy1 Vz C K sound.measured ( x, y, z )
    K predlowerbound ( y, y1 )
    lessthan ( z, y1 )
    < expert > sound.range ( x, y, belowlowerbound )#
210
Vx Vy Vy1 Vz C K sound.measured ( x, y, z )
    K predlowerbound ( y, y1 )
    lessthan ( y1, z )
    < expert > sound.range ( x, y, abovelowerbound )#
211
Vx Vy Vy1 Vy2 Vz C K sound.measured ( x, y, z )
    K predupperbound ( y, y1 )
    K predlowerbound ( y, y2 )
    K lessthanorequal ( z, y1 )
    lessthanorequal ( y2, z )
    < expert > sound.range ( x, y, normalrange )#
212
< expert > Vx C K decision.assess ( x, crackles )
    K sound.range ( x, f50, aboveupperbound )
    sound.range ( x, f85, aboveupperbound )
    conclude ( severe, crackling, x )#
213
< expert > Vx C K decision.assess ( x, crackles )
    K sound.range ( x, f50, normalrange )
    sound.range ( x, f85, aboveupperbound )
    conclude ( mild, crackling, x )#
```

```
214
< expert > Vx C K decision.assess ( x, crackles )
    K sound.range ( x, f50, normalrange )
    sound.range ( x, f85, normalrange )
    conclude ( insignificant, crackling, x )#
215
< expert > Vx C K decision.assess ( x, breathsound )
    K sound.range ( x, f50, aboveupperbound )
    sound.range ( x, f85, geupperbound )
    sound.spectrum ( x, major.shift.to.high.frequency )#
216
< expert > Vx C K decision.assess ( x, breathsound )
    K sound.range ( x, f50, aboveupperbound )
    sound.range ( x, f85, belowlowerbound )
    sound.spectrum ( x, abnormal.concentration )#
217
< expert > Vx C K decision.assess ( x, breathsound )
    K sound.range ( x, f50, normalrange )
    sound.range ( x, f85, leupperbound )
    sound.spectrum ( x, normal )#
218
< expert > Vx C K decision.assess ( x, breathsound )
    K sound.range ( x, f50, normalrange )
    sound.range ( x, f85, aboveupperbound )
    sound.spectrum ( x, moderate.shift.to.high.frequency )#
219
< expert > Vx C K decision.assess ( x, breathsound )
    K sound.range ( x, f50, belowlowerbound )
    sound.measured ( x, f85, leupperbound )
    sound.spectrum ( x, shift.to.low.frequency )#
220
< expert > Vx C K decision.assess ( x, breathsound )
    K sound.range ( x, f50, belowlowerbound )
    sound.range ( x, f85, aboveupperbound )
    sound.spectrum ( x, abnormally.broad )#
221
< expert > Vx C K decision.assess ( x, crackles )
    K sound.range ( x, f50, aboveupperbound )
    sound.measured ( x, f85, leupperbound )
    conclude ( moderate, crackling, x )#
222
< expert > Vx C observed ( x, crackles )
    decision.assess ( x, crackles )#
223
< expert > Vx C observed ( x, breathsound )
    decision.assess ( x, breathsound )#
224
< expert > Vx C observed ( x, quiet )
    decision.assess ( x, breathsound )#
225
< expert > Vx C observed ( x, breath )
    decision.assess ( x, breathsound )#
```

This is a listing of the 'Inspire.all' rule base. The first order rule base for pulmonary function interpretation is held in the file 'Inspire.fol' and the corresponding belief logic rule base is held in the file 'Inspire.ibq'.