Cleary 1

KNOWLEDGE-BASED SYSTEMS for the INTERPRETATION of SEISMIC DATA

John G. Cleary[*], Laurane L. Kramer[*], Michael P. Wingham[**]

[*]Department of Computer Science,
University of Calgary,
Calgary, Alberta, Canada
[**]Teknica Resource Development Ltd.,
Calgary, Alberta, Canada

The domain of geophysical seismic data processing is analyzed for the applicability of artificial intelligence and knowledge engineering techniques. The main area of use for this domain is in the search for hydrocarbon deposits. The domain is characterized by large amounts of data and processing to be done on them. The processes range over a continuum from large volume signal processing to high level skilled interpretation. Many different attributes vary across this range. It is claimed that these attributes determine the style of processing and reasoning appropriate at each stage. A prototype system in the mid-range of the continuum has been developed. The prototype and its characteristics are examined in light of this analysis.

1. PROJECT OVERVIEW

The interpretation of geophysical seismic data involves very large amounts of data processing coupled with intensive and skilled human interpretation. The demands of hydrocarbon exploration have exacerbated this situation with the introduction of 3-D surveys and tomographic techniques. The application of AI techniques to seismic data processing is one phase of a large-scale research and development project, called TEKXPERT, currently being undertaken by TEKNICA RESOURCE DEVELOPMENT LTD. The project aims meet these challenges by utilizing new computer technologies and software techniques including knowledge engineering. TEKNICA is managing the project and providing its substantial human expertise in all areas of hydrocarbon exploration.

At present the majority of seismic interpretation is a patchwork of various and often ill-understood mathematical modelling routines written in FORTRAN and dependent on massive mainframes and array processors. Systems are run in batch mode with the resulting long turnaround times. The overall goal of the project is to develop powerful interactive graphical workstations utilizing local optical disc storage. The hope is that by automating parts of the overall process and by eliminating down time the process can be fundamentally improved. For example: timely access allows much more substantial use of archived seismic and geological data; faster response time encourages greater experimentation with parameters; automating routine tasks frees human processors and interpreters to concentrate on areas requiring specialized skill; knowledge-based programming makes geological knowledge available earlier in the processing.

The Artificial Intelligence (AI) subgroup within TEKXPERT was formed in December, 1984 with a mandate to determine the feasibility of applying AI techniques to seismic data interpretation. Its work is being done in co-operation with the Software Research and Development Group of the University of Calgary. When the subgroup was formed there was little consensus about the possible role of AI in the total project, but two constraints were obvious. First, AI techniques would have to co-exist with the current extensive library of numerical processing software. Consequently, the programs would have to run on hardware compatible with that used in the industry. Second, the human interpreter must remain in complete control of the entire process. Any AI techniques for inferencing and decision making would have to be optional and reversible.

In order to determine the feasibility of using AI techniques and to provide a focus for understanding seismic data interpretation, the AI Unit undertook to identify a small subproblem and develop a prototype solution. Simultaneously, discussions with domain experts and research into the alternatives offered by AI were initiated in order to characterize the domain of seismic interpretation and to identify the AI approaches which could be useful.

The following section very briefly describes the process of seismic data interpretation. Sections three and four discuss some of the characteristics of seismic data and its interpretation which have an impact on the choice of AI techniques. Sections five and six describe the problem of velocity picking and the prototype systems for its solution. Section seven summarizes the paper.

2. SEISMIC DATA PROCESSING

Today, virtually all new hydrocarbon reservoirs are located through seismic surveys. The basic idea is simple. Highly sensitive microphones (geophones) are placed at equal intervals in a straight line along the earth's surface. At selected intervals along the line of geophones a burst of noise (typically a dynamite charge) is introduced into the earth. At the boundaries between strata (horizons) the signal reflects back to the surface and is recorded by the geophones (Figure 1). The seismic signal is therefore a received signal as a function of travel time. The travel time is usually expressed in terms of an average velocity of the signal through the ground, referred to simply as velocity. The interpreter's task is to separate signal from noise, and, based on the signal characteristics and travel time, infer the sub-surface geology.

In fact the real situation is rarely, if ever, as simple as that depicted in Figure 1. The land surface usually prevents the geophones from being set at regular intervals and at fixed elevation in a straight line. These changes obviously affect an interpretation based on travel time. If the horizons are not flat, the reflection path is deflected in proportion to the angle of dip. Figure 2 illustrates the complexities introduced by a single reflector with simple, shallow dip. Signals may reverberate between layers before reaching the surface. The situation is further complicated because hydrocarbon deposits only occur in areas of geologic anomalies.

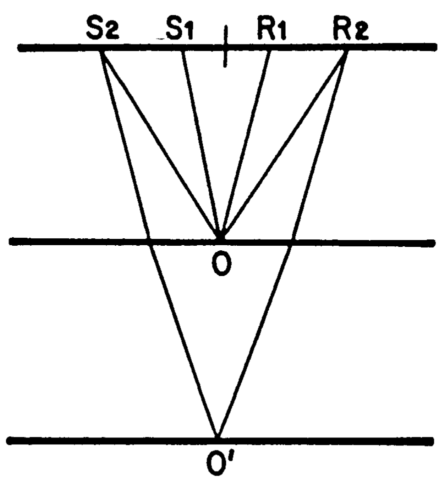


Figure 1. Geometry of a Simple Stack.

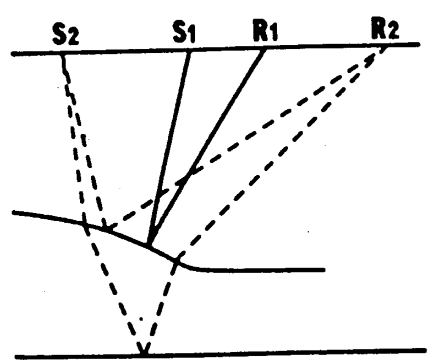


Figure 2. Geometry of a Stack with a dipping reflector.

In the face of these complexities, a sophisticated suite of signal processing algorithms has been developed. The following is a brief description of some of the stages of interpreting a seismic survey. The signals are first modified by a series of algorithms designed to compensate for the attenuation of the signal over time (gain compensation) and to compensate for differences in elevation (static removal). Next the signals are grouped to reflect the same sub-surface point and averaged to remove noise (stacking). Algorithms correct the deflection due to dip (migration). Reflections present across a number of traces are identified as sub-surface horizons (horizon-picking). A representation of the reflections along the entire line called a section (Figure 3) is produced. This is the last step in the processing phase. Next, the horizons are studied to reveal geologic features such as reefs, faults, etc. From the features identified a depositional history is built up. The goal is to identify a depositional environment likely to create hydrocarbons and structural or stratigraphic formations likely to trap them in sufficient quantities to justify exploration. Little computer processing occurs in this final stage.

Each stage is sufficiently complex that a wealth of literature and expertise exists for each one.

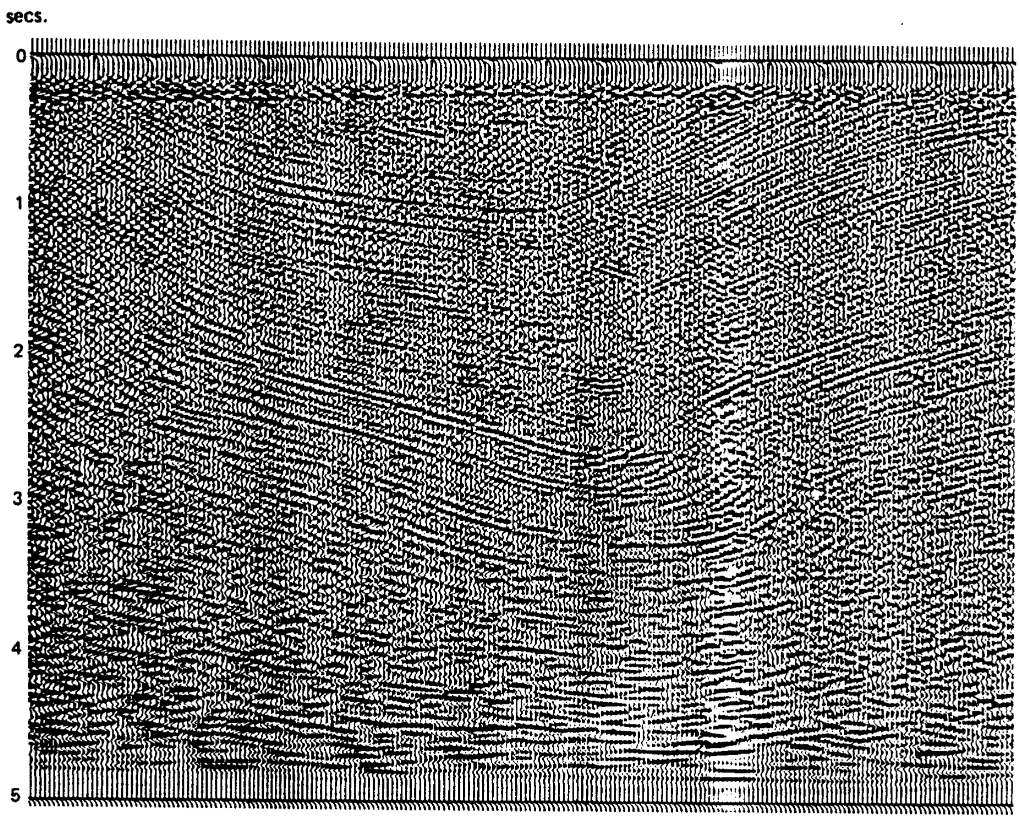


Figure 3. A Seismic Section.

3. CHARACTERISTICS OF SEISMIC DATA INTERPRETATION

Very early it became clear that the processing and interpretation of a seismic survey was not a uniform process. It is rather a complex of processes forming a continuum along several different dimensions. (Figure 4 summarizes these dimensions.) Each characteristic varies steadily and monotonicly from initial signal acquisition to final geological interpretation.

The overall process begins with the newly acquired raw data. The early stages (demultiplexing, gain compensation, static removal) are concerned with improving the signal quality. They are strictly in the domain of signal processing and illustrate the characteristics of the left pole of the continuum. They are almost totally automated and rely on highly structured mathematical algorithms. Their input is the raw seismic data and the output is modified traces with some form of signal correction. An expert wishing to describe the process uses mathematical equations from the realm of geophysics. If the output is examined the form is a graphical presentation similar to Figure 6. In the later stages of the signal processing although the emphasis remains on the use of mathematical algorithms, geological knowledge may be used to set parameters, eliminate alternatives, etc. These stages blend the characteristics of processing and interpretation in varying amounts. The interpretation stages illustrate the right hand side of the continuum. After the section is produced, the interpretation stages are less interested in actual signal manipulation. (Although an interpreter may ask to have an earlier process repeated with different parameters based on the hypothesis being considered.) The focus of the interpretation activity is to build a geological model of the subsurface. The geologist examines the section to identify features such as old beaches and reefs and attempts to reconstruct the depositional history. Instead of mathematical algorithms, the geologist works with spatial and temporal models. The results are presented symbolically. Interpretation may well be done by a team of individuals with individual specialties depending on the type of information available. A team could consist of geophysicists, geologists, geochemists, paleontologists, and regional specialists.

In the following sections some of these dimensions and their implications for the expert system are discussed in more detail.

3.1. Characteristics of the Input Data

Seismic data has certain characteristics which dominate its processing. First is the sheer volume. A land survey may produce a gigabyte of data. The data is noisy. A noise to signal ratio of 4:1 is considered very good. Noise may be random interference or a coherent distortion which cannot be analyzed. Quality control is an art in itself. The actual signal is very ambiguous. There is no 1:1 correspondence of features and signals because each feature represents the interaction of complex forces (Davis [4]). The data is clustered, overlapping and contradictory. The volume of data reduces steadily through the various processing steps. Except, perhaps, when oil is actually discovered, the data remains inexact and uncertain at all stages. Reasoning about the data in this form will likely draw most heavily on AI based systems for signal processing such as HASP/SIAP (Nii [10]). Once the phase of signal correction and enhancement has been passed the techniques of pattern recognition and interpretation utilized by Dipmeter (Smith [11]) and LITHO (Bonnet [2]) will be appropriate.

In addition to the traces, human interpreters utilize many other sources of data whenever available. All surveys are accompanied by a field geologist's report describing the terrain of the shot, rock outcroppings, the distance between shot points, the orientation of the line etc. Most surveys are accompanied by satellite photographs. In well developed oil fields there are well logs from nearby wells and intersecting seismic lines from previous shots. These sources have already been interpreted. They are used to guide or inform the interpretations of the current seismic line. Usually the signals are not compared directly, but the knowledge of the subsurface given by one is used. For example, if a sonic log has indicated that the strata at a certain depth is shale then the known characteristics of a seismic signal passing through shale can be used to refine the processing and interpretation. Applying these other forms of knowledge is an uncertain business because there is no guarantee that an area is consistent or, in fact, that the interpretation being applied is itself accurate. The use of external information increases steadily from the initial stages when little or none is used to the final interpretation where a wide and varied range of geological and geophysical information is brought to bear.

Domain

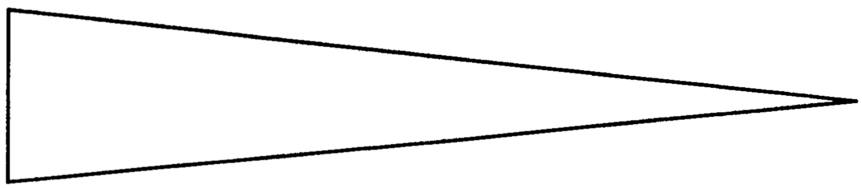
geophysics — geology

physics of sound transmission — deposition and change over time

Language

numeric and graphical — symbolic and graphical

Size



AI techniques

pattern analysis — heuristic search — inference — deep reasoning

Sources of input data

seismic traces — check shots, sonic logs — local geology — regional geology

Processing steps

gain compensation — velocity picking — sonic log correlation — horizon picking — depositional environment — interpretation — basin studies

Current systems

prototype

DIPMETER

LITHO

PROSPECTOR

Figure 4. The Attributes of Seismic Data Interpretation.

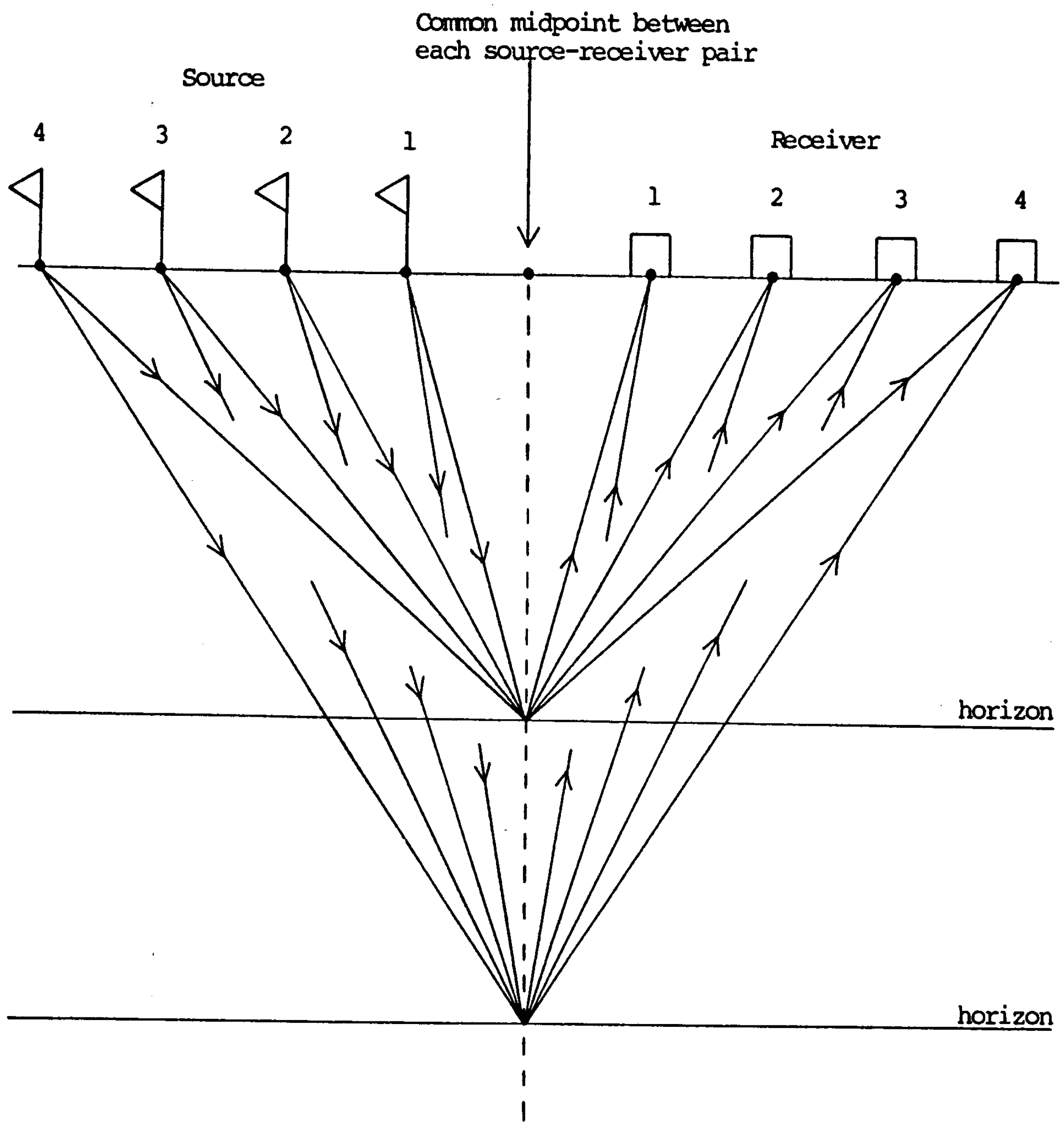


Figure 5. Source-receiver Pairs for a CDP Gather Group.

3.2. The Language of the Problem

A very important aspect of any knowledge based system is the language in which its problems are stated and solved. The language not only determines the possible internal representations of the system it is also the medium of explanation and interaction with the user. The form of language steadily changes during processing. Initially it is signals varying in one dimension (time) -- these are grouped spatially but no notice is taken of this in early steps. By the horizon picking stage this is altered to signals varying in depth embedded in the single dimensional of an underground section (the other two dimensions of the subsurface world are still ignored). As well symbolic information begins to appear at this point. For example, the statement "this is a transform fault" is symbolic but embedded within the subsurface world. The final stages of interpretation are increasingly symbolic and abstract. However, they remain imbedded within the three dimensional world of subsurface phenomena. Indeed, some of the final reasoning steps are purely ones of fitting data into a three dimensional context. This stage also includes spatial and temporal reasoning about depositional history.

This data is almost all explained and viewed by the users graphicly. This poses interesting problems for an expert system. For example, how is it to generate explanations graphicly. Systems such as DIPMETER (Davis [4]; Smith [11]) are criticly dependent on the graphical presentation of their results for their success. However, the more complex task of generating explanations does not seem to have been attempted in this context.

It is anticipated that one of the largest challenges facing the TEKXPERT system will be the processing and organization of graphical data.

3.3. Characteristics of the Interpretation Process

While the interpretation may involve many disciplines the process splits into two separate stages: the processing is done by the geophysicist and the interpretation by a geologist. Here it is useful to consider the distinctions made by Hawkins [6]. A geophysicist is a member of the behavioral school. From a theoretical or empirical basis an equation is developed relating several variables of petrophysical significance. Sets of such equations are used to solve for unknowns or in the case of multiple solutions, probabilistic methods are used to derive mathematically optimal solution. A geologist is from the classificatory school. The traces are examined for patterns associated with known geological features. The features are combined to form recognizable classifications. The geologist looks for a best fit and then rationalizes or amends the description to encompass anomalies. Of all of the problems this may be the best understood. Aikens [1] has presented an effective technique for handling prototypical knowledge. The work on PROSPECTOR (Duda [5]) also illustrates a method of dealing with geological models.

3.4. General Expertise

Expertise in the process of seismic interpretation exists in two forms. The first is specific expertise in a given area. For example, a geophysicist may become highly expert at analyzing the migration paths. The second form of expertise is well described by Hawkins [6]. This is the role of the petroleum explorationist as mediator between the specific disciplines. The primary role of the mediator is to recognize the implications of the findings from one discipline for another and to be able to translate the language of one into the other.

An integrated system for seismic interpretation will need to provide both the individual tasks and also provide some mediation. At present, the processing and interpretation process are often quite separate. The role of such an automated system would be to request reprocessing of earlier steps in the interpretation as a result of incorrect or erroneous results in later stages. It lies outside the continuum along the processing stages, and would require considerable abstract knowledge of the effects of the various processing steps and alternatives within them. It seems likely that small parts of this task (for example redoing velocity picking as a result of attempts at horizon picking) will be difficult tasks in themselves.

4. IMPLICATIONS FOR EXPERT SYSTEMS

It is instructive to look at current expert system theory to understand what relationship our problem has to first-generation expert systems. We will consider the framework proposed by Stefik et al [13] which identifies the type of task to be done, the characteristics of the task and the implications for implemented systems. The framework has been modified somewhat to highlight the dimensions most crucial to the seismic problem.

1. The Solution Space Size. The solutions, both to the final interpretation and the separate stages, are essentially infinite. Consequently exhaustive search is inappropriate. However, at least in the early stages proposing all feasible alternatives is also not appropriate. The approach taken by humans seems to be to commit to one path and follow it until something definitely contradicts it and then re-do whatever processing is necessary.
2. Reliability of the Data and the Domain Knowledge. No data is ever reliable. Further, the knowledge brought to the problem itself is not fixed. Instead there are widely differing techniques of analysis used not only from company to company but interpreter to interpreter. Because the knowledge is both proprietary and conflicting any knowledge representation used is going to require easy modification. This is true not only of the static domain knowledge but the control structure as well. The uncertainty of the data would seem to indicate some use of a technique such as fuzzy logic or Bayesian analysis, however, at this time it is impossible to determine an appropriate strategy. Humans appear to either commit to one approach or in the later stages, to hold separate and equal hypotheses in mind until deciding completely for one.
3. Availability of the Data. All data is available at the start of a consultation. However, for any given seismic line the type of data may be radically different and dictate a different solution path. In Alberta, for example, the geology is well understood. In the well developed fields there are relevant sonic logs and other seismic surveys. When these exist features determined by them will affect the selection of processing strategies and parameters.
4. Factorability of the Solution Space. No method of pruning early alternatives exists.
5. Existence of Evaluators for Partial Solutions. The entire question of determining the correctness of a solution is unclear. In the final analysis a solution is correct if it finds oil, an expensive method of verification. Some of the sub-problems do have ways of checking the correctness of their results.
6. Subproblems Interact. Ideally the entire process should be highly interactive. The goal of the final TEKXPERT project is to speed response sufficiently so that latter stages could inform the choice of parameters for earlier ones. An interpreter could for example, select certain traces to process through to final section and on the basis of the interpretation, reprocess the entire line.
7. Requirement for Efficient Guessing. The data may be sufficiently ambiguous that at a given stage the most appropriate technique is to simply guess and continue on until the guess is confirmed or disproven.
8. Multiple Knowledge Sources. Each subproblem is itself a separate knowledge source. Further, within a particular problem it may be necessary to bring several sources to bear.
9. Computational Complexity. The sheer volume of data to be processed may well turn out ot be the most significant factor in the entire project.

5. VELOCITY PICKING

The problem we chose to prototype was the processing after compensation has been done for signal attenuation and before horizons have been identified in the data. The main process here is the assignment of a velocity (average velocity over entire signal path) to each prominent reflection signal in the data. Hence the name for this process "velocity picking". Once this has been done a number of different signals can be averaged. This results in a "picture" of the subsurface with an enhanced signal:noise ratio which is then searched (usually manually) for geologically significant features. To see in more detail what is involved in this step it is necessary to look carefully at the geometry of the signals being processed. Lindseth [7] gives a detailed discussion of the geophysics of the process.

By varying the distance between source and receiver but keeping the midpoint the same, one gets a set of traces with each consisting of reflections form the same vertical line (Figure 5). Such a set of traces is called a common-depth point (CDP) gather group. Since the traces consist of reflections from the same vertical line, each trace should contain the same signal component plus a random noise component. By stacking (taking the average of) the traces the random noise should cancel thus greatly improving the signal to noise ratio. This process is based on the assumptions of sources ,and receivers at the same elevation and flat lithology. Most of the complexity of this process comes from the fact that the earth doesn't cooperate.

Because the signal reflected from a particular horizon follows a different length path through the earth for each source-receiver combination in the CDP gather group, one reflection event will appear at different times on each trace. The event will appear earlier on the near traces (source-receiver near mid-point) and later on the far traces.

Because we want to stack the signals corresponding to a common event (reflection off a lithologic horizon), we must determine what two-way travel times on each trace correspond to that event. By examining Figure 6, one can see that it is often possible to trace events across the CDP gather group; they appear as dark lines (event lines) curving down to the left.

The curving lines are hyperbolas whose shape is determined by the depth of the corresponding horizon and the average velocity to that depth. The problem tackled by the prototype is the assignment of velocities at each time in the traces. Once this has been done all the CDP gather group traces can be corrected to line up the events and stacked (averaged) to remove noise. This gives a single trace for each vertical line in the ground. These traces then form the input to next stage the detection of subsurface horizons running across many traces.

Many factors make this process non-trivial: there may be multiples, reflections which have reverberated between more than one layer, these must be disregarded; the original traces are noisy; it is important to assign accurate velocities otherwise much information will be lost in the final stacked traces; it is important to follow small local variations in velocity as they can be important in detecting the small structural changes indicative of hydrocarbon deposits. Matching the events across the traces is currently done by visual inspection. The main problem with this is that it is not possible to do the many accurate determinations needed to extract all the information potentially available.

6. THE PROTOTYPE

The prototype system for velocity picking separates the problem into two stages. The first stage, the pattern analyzer, uses heuristics and pattern matching to detect events in the CDP traces, match hyperbolas to them and then assign a velocity and depth to the event. The second stage, the curve selector, takes the velocity/depth pairs and assigns a consistent curve of velocity with depth, in the process rejecting some pairs from consideration. The communication between the two stages is one way and includes a goodness of fit measure and an estimate of the signal strength of the event with each pair.

In the context of Figure 4 this velocity picking system sits at the point where the first symbolic reasoning appears. The pattern analyzer is almost completely numeric, dealing with traces and large volumes of data. The curve selector is operating in a partially symbolic domain, with a relatively small number of discrete objects (events), although their attributes (time, velocity, goodness of fit, signal strength) are numeric. Correspondingly, the operation of the pattern analyzer is numeric and heuristic with no discrete and separable knowledge. The operation of the curve selector is a search process across a finite number of solutions.

The following sections give more details of the operation of the pattern analyzer and the curve selector.

6.1. The Pattern Analyzer

The Pattern Analyzer accepts data which has had preliminary processing (filtering, signal enhancement, deconvolution) applied. The first problem is 'seed' generation i.e. deciding how to start looking for event lines. An event line will follow a peak, trough or zero-crossing. Generally the near traces have less noise than the far ones, so we look at the near traces first and pick seeds down to a suitable amplitude threshold. If one of the seeds is incoherent noise it will not grow and will be discarded. In the gather groups there are often events that are visible for a few traces, disappear and then reappear further out; and in other cases the event might not be visible at all on the near traces. The former case can be handled by allowing the growing process to skip over a few traces and continue farther out by extrapolation. The latter case can be handled by generating seeds periodically across the trace, say every twelfth trace. The next step consists of growing as long a line as possible from each seed. The basic idea is to find an event of corresponding phase and amplitude on the next trace at some acceptable displacement in time. This done by doing a correlation between the adjacent traces and picking the displacement (within a predefined window) which gives the highest correlation. Sometimes two lines generated from different seeds will need to be merged when they are found to be almost identical. Figure 7 shows the event lines in the format generated by the prototype. Once the lines have been identified, a hyperbola is fitted to the event line using a least-squares technique. From each hyperbola the associated depth and velocity are calculated and passed with the set of correlation measures to the curve selector. The hyperbolas fitted to the event lines of Figure 7 are shown in Figure 8.

6.1.1.

Given the depth-velocity pairs generated by the pattern analyzer (Figure 9), the curve selector isolates those pairs which represent the final velocity function against depth. First, pairs which are obviously unacceptable are eliminated, these are pairs whose: times are below the geologic basement; velocities are unreasonably high or low; velocities are radically different from neighbouring pairs. As well some pairs are redundant, that is they have essentially identical velocities and depths, these are merged into one pair.

The next step is to build a graph consisting of the possible paths through the remaining depth-velocity pairs, where each path represents a possible stacking velocity function. Each point (depth-velocity pair) on the tree is weighted according to the following criteria: its proximity to other points; its proximity to the expected path of the stacking velocity function; its resemblance to a simple multiple (it has a velocity similar to that of another point at half the time). The first two criteria increase a point's weight while the last one decreases it.

Each path segment joining two points is then weighted according to: the weights of the points it joins; the closeness in time between the two points; the direction of the path segment - a velocity increase with depth is preferable to a velocity decrease; the change in velocity - small changes are preferable to large ones; the change in average velocity between two points implies a local velocity in the intervening rock - there are physical constraints on the possible values for such local velocities. Finally a single path of maximum weight is selected through the graph. This constitutes the final velocity/depth curve which is then used to correct the raw traces and produce a single stacked trace. The corrected traces are shown in Figure 10.

7. SUMMARY

Seismic data processing involves a great range of characteristics in its various stages. In building an intelligent automated system for this domain a wide range of knowledge based techniques will be necessary. There is a continuous flow of information from the raw seismic data through to high level geological interpretation. This flow continuously changes its many characteristics. For example the language of the early stages is graphical and numerical of the later stages symbolic and linguistic. On top of this there must be an element of feedback and control of the various steps. This element must have an overall appreciation of the effects of choosing various options at each stage and be able to effectively vary them to reprocess the data. To build a complete automated system will require a strong awareness of these varying characteristics at the different stages. There appear to be many significant problems remaining to be solved, particularly, in the area of representing and explaining data which is viewed by its users as mainly graphical and visual.

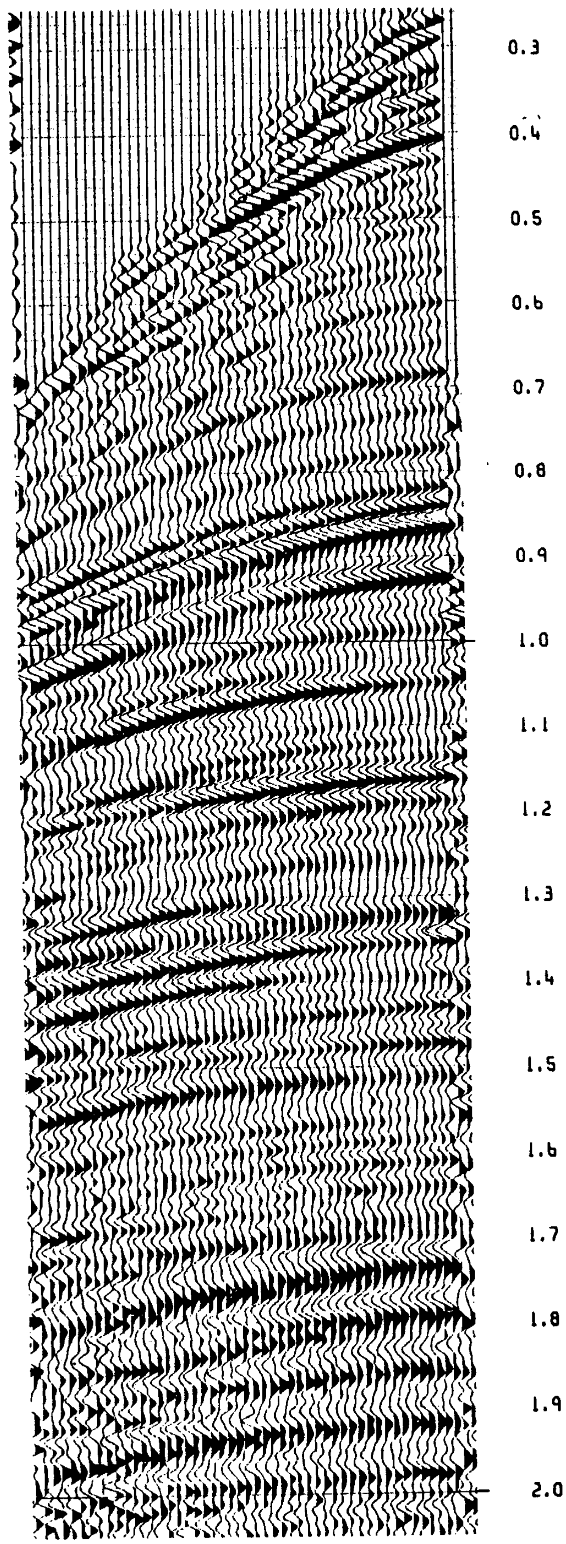


Figure 6. Traces from a CDP Gather Group.

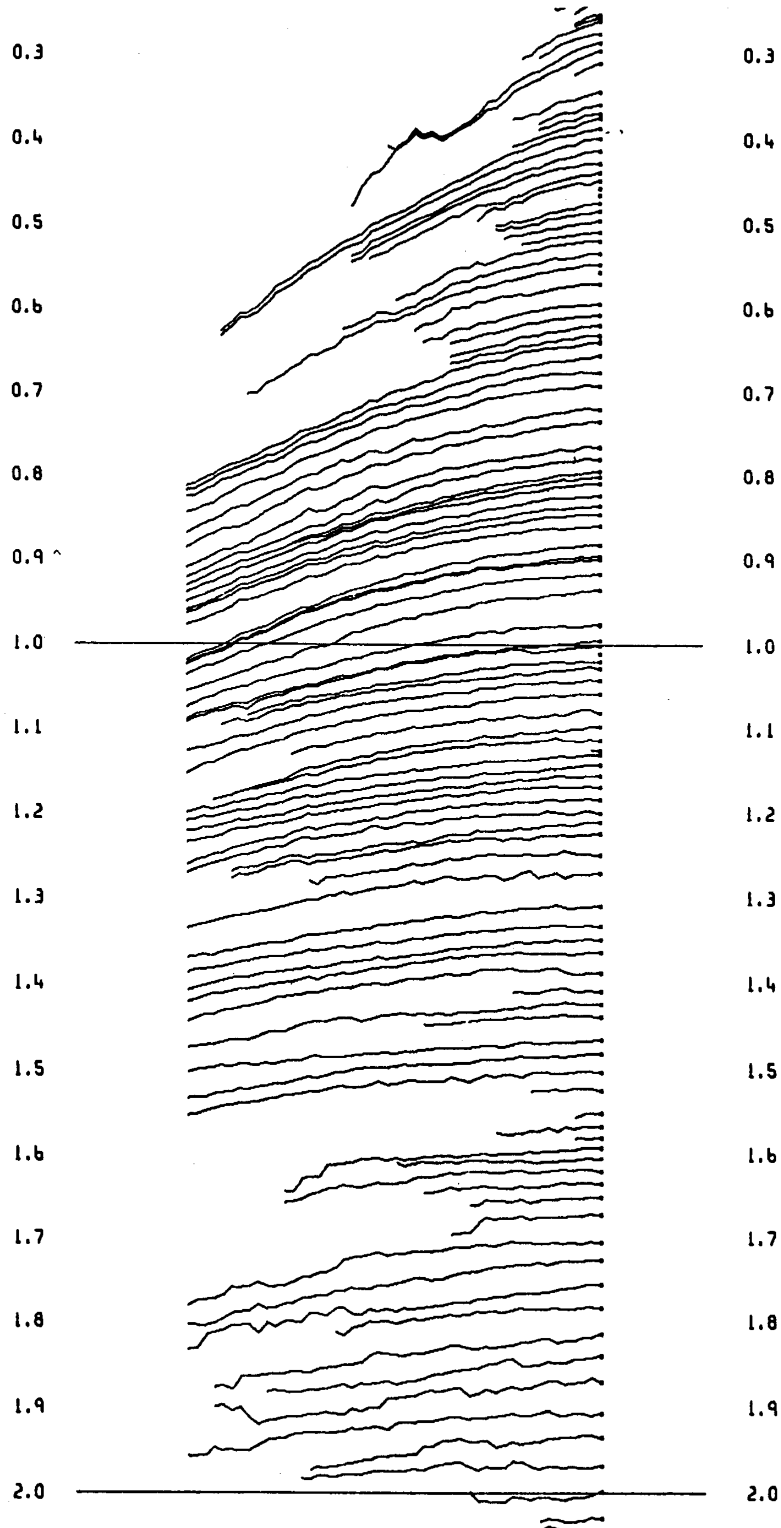


Figure 7. Event Lines for the CDP Gather Group of Figure 6.

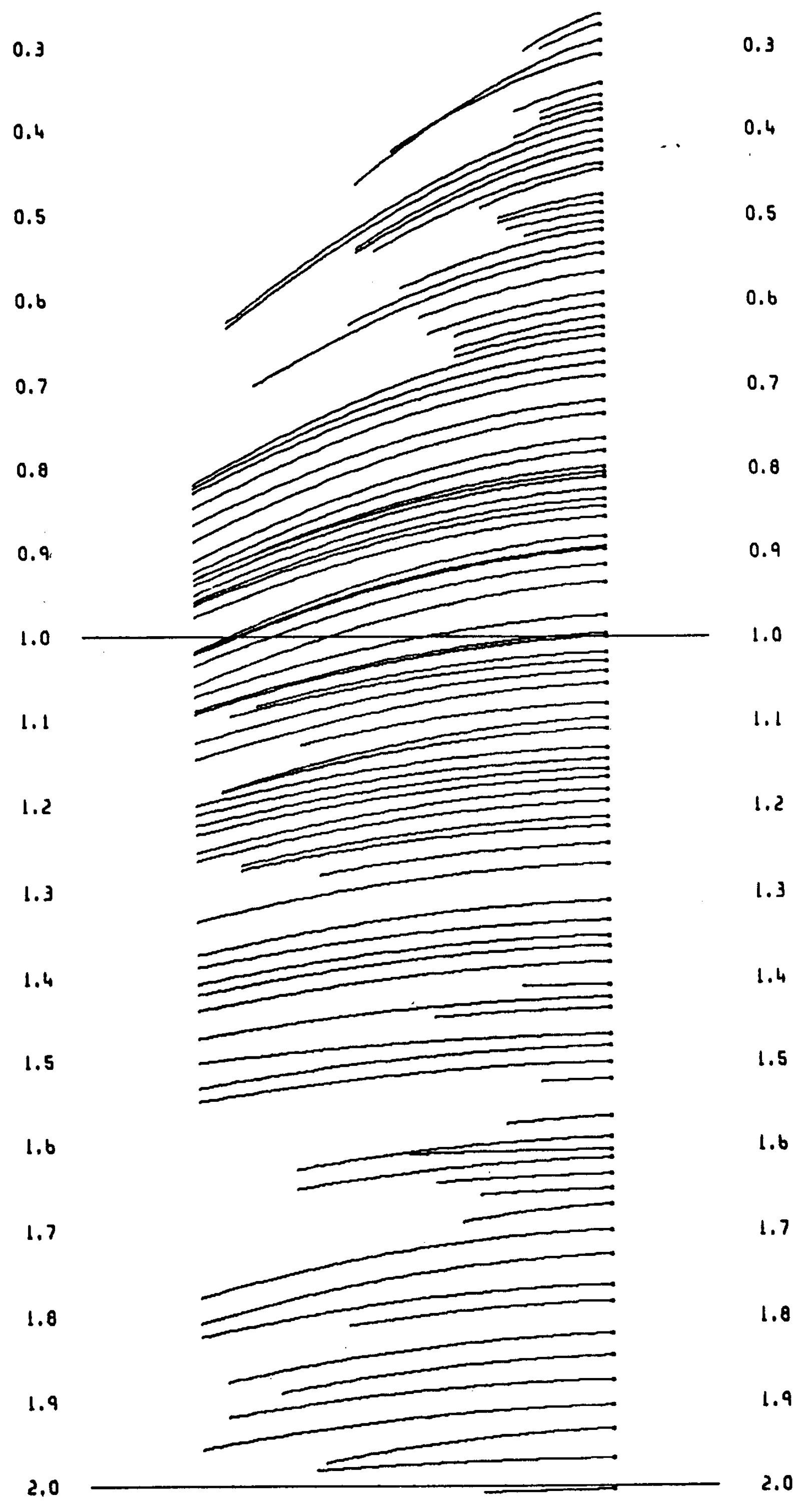


Figure 8. The Hyperbolas Fitted to the Event Lines of Figure 7.

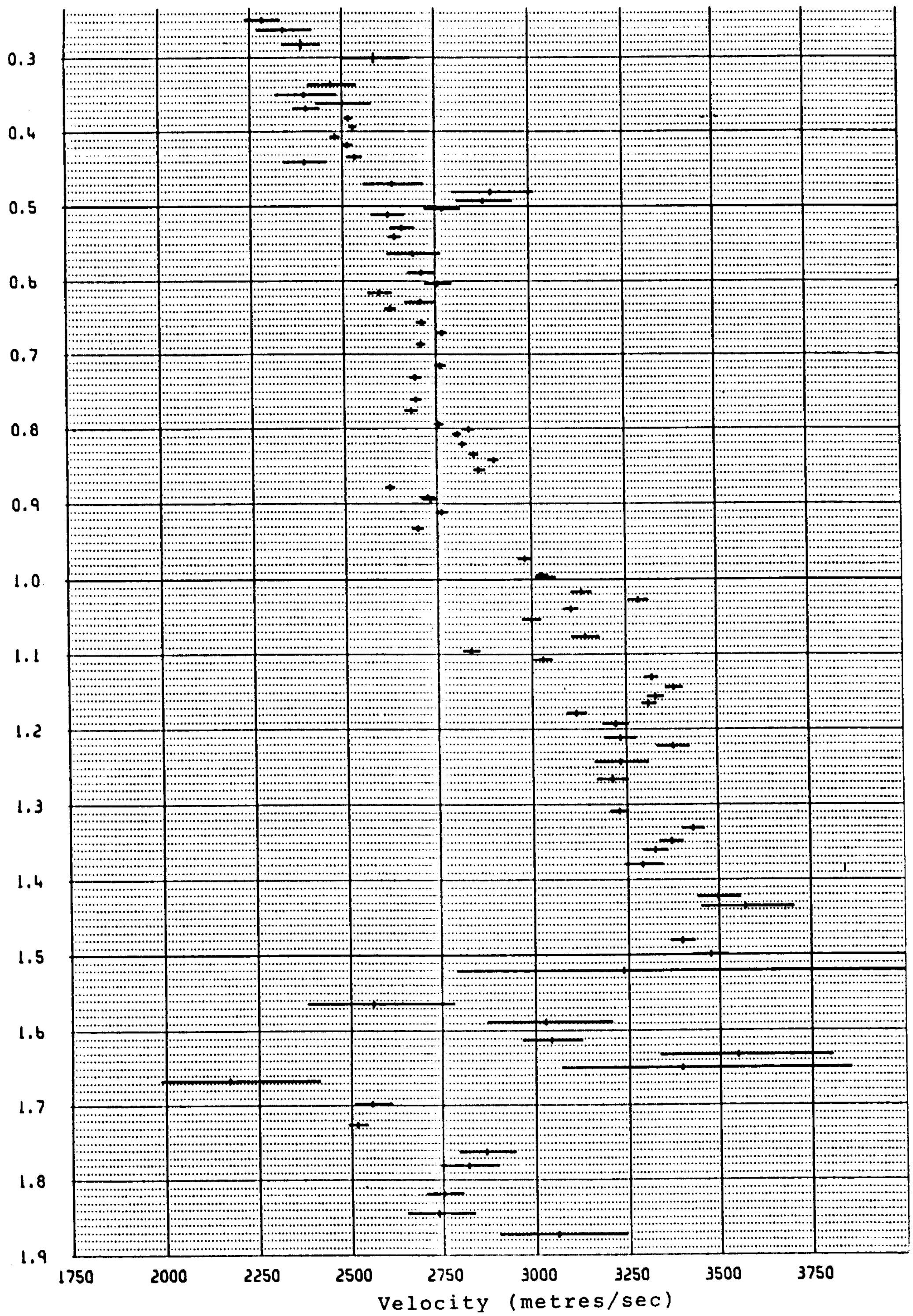


Figure 9. Time Velocity Pairs generated from Figure 8. The cross-bars represent 90% confidence intervals for time and velocity.

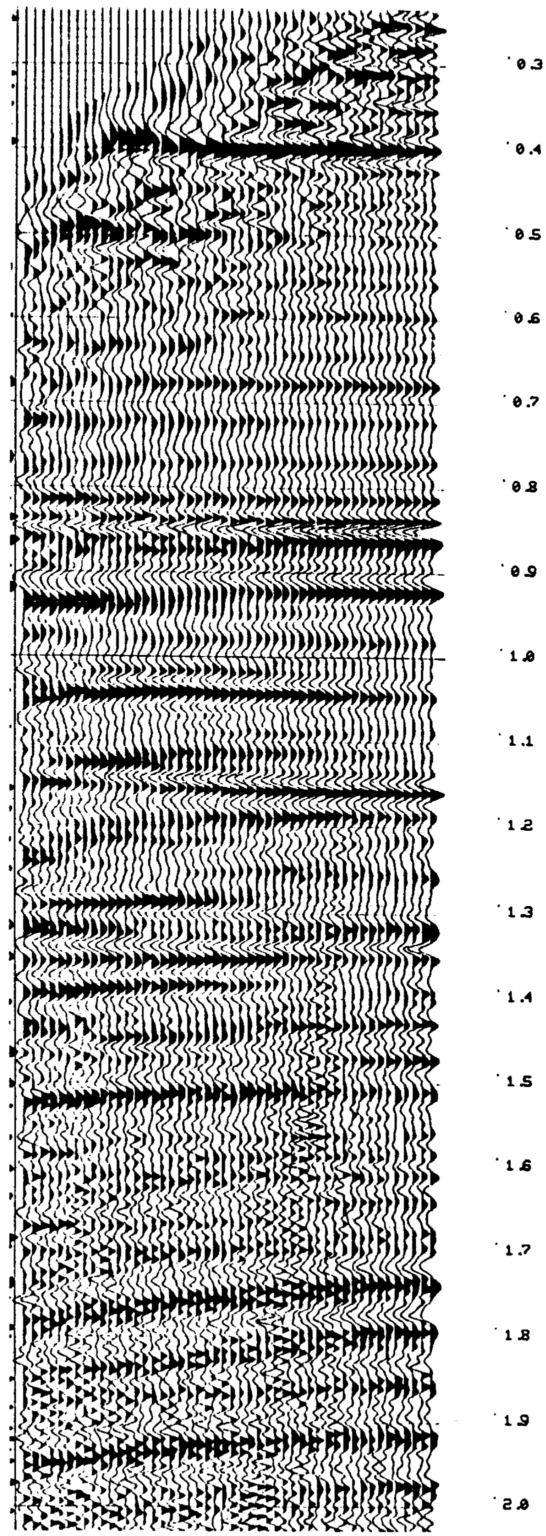


Figure 10. Velocity Corrections applied to the CDP Gather Group of Figure 6.

8. ACKNOWLEDGEMENTS

This work was supported in part by the Natural Sciences and Engineering Research Council of Canada.

9. REFERENCES

[1] Aikens J., Prototypical knowledge for Expert Systems, Artificial Intelligence, 20(1983) 163-210.

[2] Bonnet, A. and Dahan, C., Oil-well data interpretation using Expert System and Pattern Recognition Techniques, in: Proceedings of Eighth International Joint Conference on Artificial Intelligence, (1983) 185-189.

[3] Davis, R., Austin, H., Carlbom, J., Frawley, B., Pruchnik, P., Sneiderman, R. and Gilbraith, J., The Dipmeter Advisor: Interpretation of Geologic Signals, in: Proceedings on Seventh International Joint Conference of Artificial Intelligence, (1981) 846-849.

[4] Duda, R., Hart, P., Barret, P., Gasching, J., Konolige, K., Reboh,R. and Slocum, J., Development of the Prospector consultation system for Mineral Exploration, Final Report on SRI Projects 582/06415, SRI International, (1978).

[5] Hawkins, J., An Analysis of Expert Thinking, Int. Jrnl. of Man Machine Studies 17(1982).

[6] Lindseth, R., Digital Processing of Geophysical Data - A Review, Teknica Resource Development Ltd., (1982).

[7] Nii, P., Feigenbaum, E., Anton, J. and Rockmore, A.J., Signal-To-Symbol Transformation: HASP/SIAP Case Study, AI Magazine Spring(1982) 23-35.

[8] Smith, R. and Young, R., The Design of the Dipmeter Advisor System, Schlumberger - Doll Research, (1984).

[9] Stefik, M., Aikens, J., Balzer, R., Benoit, J., Birnbaum, L. and Hayes-Roth, F., Organization of Expert Systems, Artificial Intelligence 18(1982) 135-172.