AN APPLICATION OF SYNTACTIC PATTERN RECOGNITION TO SEISMIC INTERPRETATION

Lawrence H.T. Le and Edo Nyland

Seismology Laboratory, Department of Physics
University of Alberta
Edmonton, Alberta T2G 2J1

ABSTRACT

Seismic exploration data can be used to image the acoustic impedance variations in the earth. In order to convert such data to an image that more closely matches the vision of geology image enhancement techniques, including pattern recognition methods, must be applied. A syntax-dependent approach employing a string-to-string matching algorithm matches peaks between traces on a seismic record. A filtering process then enforces matching coherence by correcting matches that deviate seriously from the general trend around anomalous pairs. Connected pairs form lateral coherent events which have a confidence measure. These events are objects of any seismic investigation. Clustering techniques can be used to associate the objects with geologic zones. The algorithm performs well in a test run and detects most of the strong reflections.

KEYWORDS: string-to-string matching, seismic records, trace, events, parallel analysis, clustering analysis, zone.

INTRODUCTION

Geophysical exploration of the earth's crust aims to develop an image of the distribution of physical properties within the earth by remote sensing methods of various kinds. Among the many methods used seismic reflection studies are by far the most productive and it is to seismic data that we apply some novel, for geophysicists, approaches. These approaches have merit in other geophysical techniques, but we do not discuss them here.

During the past ten years, substantial achievement has been made in the area of geophysical data acquisition, processing and inversion. The reliability and resolution of information available to geophysical interpreters have increased. In spite of such a large data base of good quality, computer software is not sophisticated enough to help geophysicists make even simple judgement or monitor the data processing flow. Even though computer usage is more pervasive than ever, the nature of the computer program is virtually the same. Geophysical data processing programs are by and large compute intensive processing tools that depend extensively for their control on expert judgment. Very few decision-making aids exist in the lexicon of geophysical image processing systems.

Seismic data contain records of signals that are echoed from acoustic impedance contrasts within the earth. Seismic interpretation is the identification of these contrasts and could be done with expert systems. The final goal of interpretation is to obtain the image of the subsurface structure in terms of physical quantities such as densities and velocities. This requires recognition of the major seismic events on a seismic record and translation of these events to yield the desired quantities.

A seismic record is the response of geophones to a source of mechanical energy. The record of response with time of any single geophone is called a trace. The elastic waves generated by the source are perceived to travel into the earth and be reflected at the impedance discontinuities. Such reflections appear on a seismic record as wavelets of strong amplitudes and form seismic events.

Picking the events is an important and elementary step in any seismic interpretation process and is usually done manually. Statistical picking algorithms have been proposed (Paulson and Merdler, 1968; Schneider, 1971) but pattern recognition (Fu, 1974; Lu, 1982) provides an attractive alternative.

AN AUTOMATED SEISMIC EVENT-PICKING PROCEDURE

A seismic trace can be considered a pattern or string of primitives with features such as amplitude, location and duration. A comparison between two traces can be made on their corresponding strings by the string-to-string matching algorithm which examines all possible mapping between the two strings and determines the optimal mapping which derives one string from the other with the least effort.

Consider a seismic trace as a sequence of cycles (Figure 1) each of which is bounded by two consecutive local minima. The primitives are defined as an upslope followed by a downslope. The amplitude, $y'(t_p)$ is

$$y'(t_p) = \frac{y(t_p) - y^*(t_p)}{2} \qquad (1)$$

The duration is the time difference between A^* and B^* and the location is time, t_p. The time gap between two adjacent transformed peaks is the duration of the corresponding null primitive which has a location in the middle of this gap.

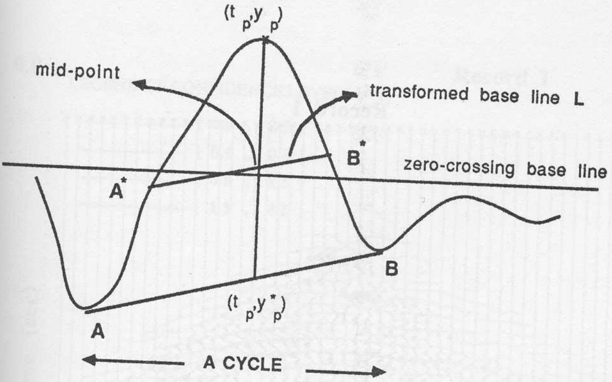

Figure 1. The waveform cycle.

The cost of matching traces is defined by the characteristics of the waveforms and interpreter's criteria for waveform matching. A cost function, or heuristic rule, that appears to mimic interpreter's judgement is

$$c(i,m;j,m+1) = W_1 \left| t_{i,m} - t_{j,m+1} \right| + W_2 \left| d_{i,m} - d_{j,m+1} \right|$$

$$+ W_3 \left| \frac{y'_{i,m} - y'_{j,m+1}}{AVER} \right| + W_4 \,(MDU\text{-}OVL) \tag{2}$$

where

$$AVER = \frac{y'_{i,m} + y'_{j,m+1}}{2} \tag{3a}$$

for substitution and

$$AVER = \frac{\sum_{i=m}^{m+1} \sum_{j=1}^{l_i} y'_{j,i}}{\sum_{i=m}^{m+1} l_i} \tag{3b}$$

for deletion or insertion. c(i,m;j,m+1) is the matching cost between the i^{th} primitives of the m^{th} trace and the j^{th} primitive of the $(m+1)^{th}$ trace. W_1, W_2, W_3, and W_4 are weights and l_i is the total number of peaks of the i^{th} trace.

The first term accounts for the location difference; the second term, the duration difference; the third term, the amplitude difference. The fourth term accounts for the non-overlapping part of the two corresponding transformed waveforms (Figure 2). MDU=*min*($d_{i,m}$, $d_{j,m+1}$), and OVL is the overlap between them. For insertion and deletion, a peak is matched with a null primitive. This heuristic rule is an experience formulation which can not be rigorously proven but reliably guides the matching action. Human interpreters frequently compare two waveforms by noting their physical and temporal differences. The rule attempts to encode such a judgement. More sophisticated functions could be designed if more precise discrimination is needed, but the examples we use do not seem to warrant the extension.

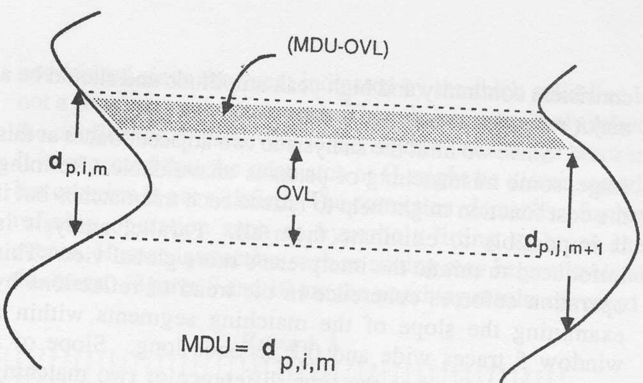

Figure 2. The shaded area is the part of the waveform that does not overlap and adds to the matching cost.

For the same difference in amplitude, matching a pair of strong peaks should imply less cost than matching a weak pair and this criterion is reflected in the cost functions by considering the fraction of the amplitudes in the third term. The fourth term is not independent of the first two terms and a small weight is assigned for its lesser significance.

The weights have limited meaning, since terms in the cost functions are not normalized. The distribution of weights should be such that the average values of those terms indicate their importance. The weights are chosen experimentally in a way that the amplitude term is emphasized. They are 0.15, 0.10, 0.70 and 0.05. Indeed, the weights very much reflect the intuitive knowledge and judgement of an interpreter and their values provide a means for comparing different such judgements. They should be taken as a way to represent this quantitative knowledge.

Every peak of any trace is matched to all peaks of the adjacent traces. The string-to-string matching algorithm is applied to compute the cost of all three types of transformations and to pick the matching which has the least cost. At this stage we associate with each matching segment joining two peaks of two adjacent traces a confidence measure defined as

$$F_{vm}(i,m;j,m+1) = \frac{\sqrt{y'_{i,m}\, y'_{j,m+1}}}{c(i,m;j,m+1)} \tag{4}$$

where F_{vm} is the matching confidence of the v^{th} matching pair of the m^{th} match. All the matched peaks are then connected to designate a coherent event each of which is also assigned a merit of confidence G_z. The value

$$G_z = \sum_m F_{vm} \tag{5}$$

is the coherence confidence of the z^{th} event and consists of the sum of all the matching confidences of the pairs that make up this event. Both the matching confidence and the coherence confidence measure the uncertainty about the decision made by the matching algorithm. The coherence confidence discriminates the major events from the rest in the record. Coherence confidence reflects the reliability and importance of a coherent event and results in four levels of confidence in the examples. A coherent event with high confidence has consistent continuity and high peak amplitude and should be a major reflection event.

Since we limit the analysis to two adjacent traces at this stage, some mismatching of peaks is unavoidable. Refining the cost function might help to reduce such mismatches but it is impossible to eliminate them all. Parallel analysis is introduced to mimic the interpreter's more global view. This operation enforces coherence in the trend of reflections by examining the slope of the matching segments within a window 5 traces wide and 0.2 seconds long. Slope of a segment is defined as the time difference of two matching peaks. Within the window, the weighted average slope, $SLOPE_w$ and the weighted standard deviation, STD_w of the slope of the matching segments are calculated. These weighted parameters are obtained by weighting the slope with the corresponding matching confidence, F_{vm}. In this manner, the average slope within the window is mostly dominated by the matching segments of high confidence. A matching segment is chosen as a possible mismatch if $|SLOPE - SLOPE_w| > STD_w$ and $F_{vm}(i,m; j,m+1) < T_1$ where T_1 is the threshold chosen to be the maximum value in the lowest matching confidence bracket. If these conditions are met a peak in the old matching pair has to be matched with a new peak, if available, within a bounded interval. Crossing of matching segments is prohibited. A new match is accepted if the result is improved. A new match is also added if the slope of the new pair follows the existing trend in the window.

A seismic record can further be divided into zones with homogeneous features within zones and significant differences between them. Coherence confidence is the major feature of zones. An event of high confidence and a group of adjoining coherent events with high confidence are likely to be the seismic responses arising from similar strata. We take these as zones and the region of a record between them is also taken as zone. The partition into zones is a simple clustering procedure using set of coherent events with highest confidence as seeds for clusters. Zoning is begun at the event of the highest confidence in this set. An event will join this zone if and only if it is a member of the set, it lies within 0.1 second close to the zone, and no more than 4 coherent events laterally spanning a total length less than n lie between it and the seed of the zone. The event length threshold n is 24 (50% of the maximum length of a coherent event, 48 traces).

When no coherent event can join this zone, another coherent event of next highest confidence in the set is chosen as the seed of a new zone. If no other events satisfy the criteria, the seed is a zone by itself. The region between two zones with high coherence confidence is also a zone, in spite of the fact that it may consist of a collection of low coherent events. Zone features such as average length, average confidence, average peak amplitude, and average seismic attributes of the coherent events can then be extracted.

EXAMPLES

Figures 3, 7 and 9 are portions of three seismic records taken at adjacent shot points.

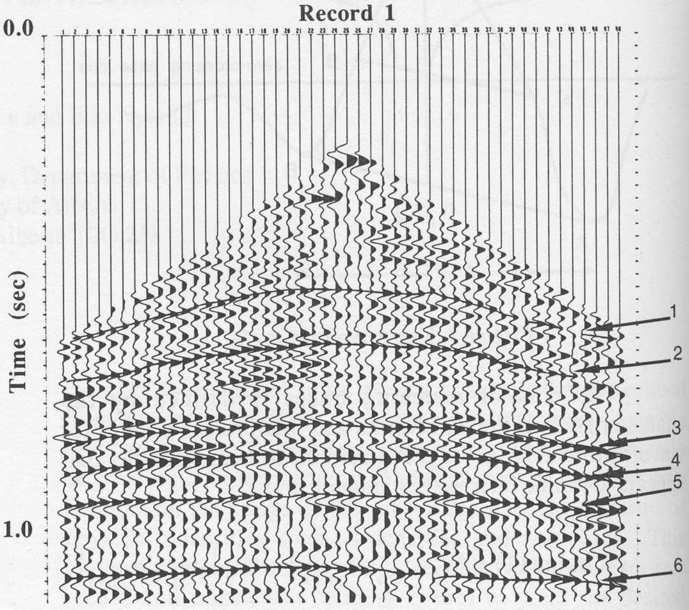

Figure 3: Seismic record 1 with visually recognized major reflections.

Six consistent events, numbered as 1 to 6 in the figures can be used to judge the competence of the automated algorithm. In Figure 4, events (on record 1) of highest confidence are shown in thick solid lines, while those of least confidence are dotted lines. Matches at A, B, C, D and G do not conform to the general trend of the events in their vicinity. Figure 5 indicates that the mismatches have low matching confidence.

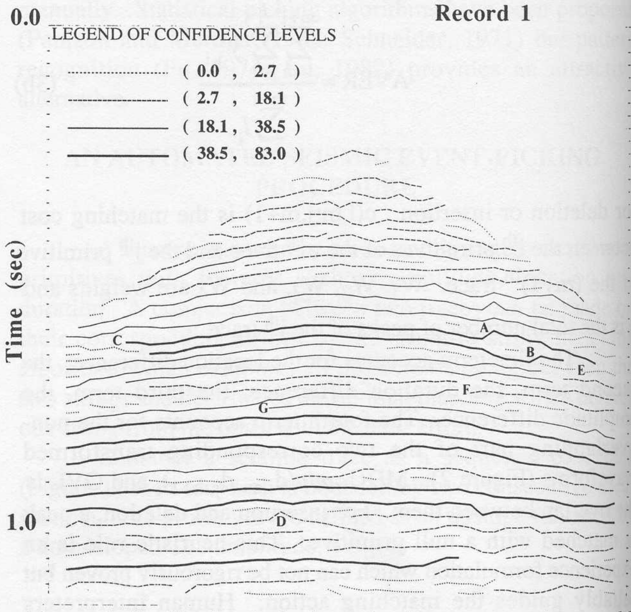

Figure 4: Profile shows the lateral coherent events in record 1 with coherence confidence indicated.

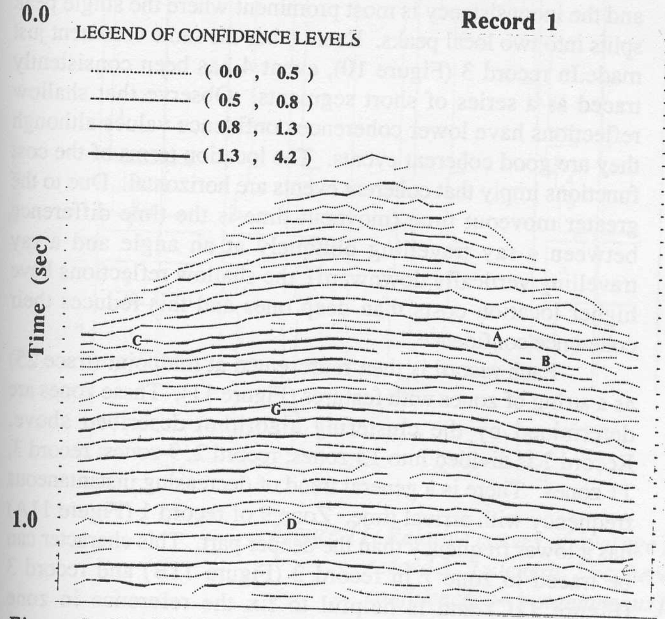

Figure 5: Profile shows the matching segments in record 1 with matching confidence indicated.

Figure 6 is the parallel analysed matching result. The matches at A, B and C have been improved while the mismatch at G remains unchanged. The correction at D has been erroneously made since the slope of the 'improved' result does not follow the surrounding trend.

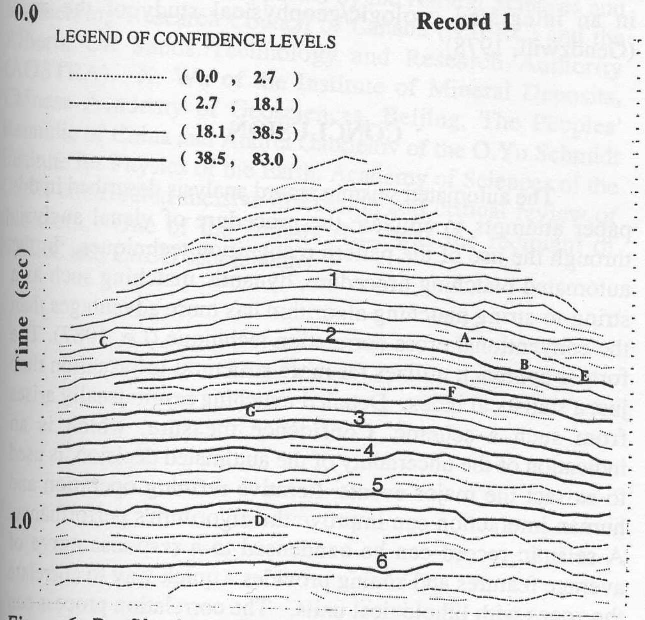

Figure 6: Profile shows the lateral coherent events in record 1 after parallel analysis.

However, the error does not destroy the general performance of the scheme and thus can be ignored since the resultant event is short and of low confidence. New pairs have been added at E and F. Joining at E has nicely completed the tracing of the major event labelled as 2 while joining at F is not desired since the event indicated by the thick solid line is not a consistent coherent event. The join has mistakenly raised the confidence of that event to the highest confidence bracket. It is expected that the mismatch at G might be disconnected, but criterion is not satisfied. The correction demands a further refining operation such as hyperbola-fitting to break the events. Human interaction such as editing may be profitable at this stage. Figures 8 and 10 are the matching results of record 2 and 3 after parallel analysis. In the results, a variation of ±0.02 of each weight does not change the results significantly.

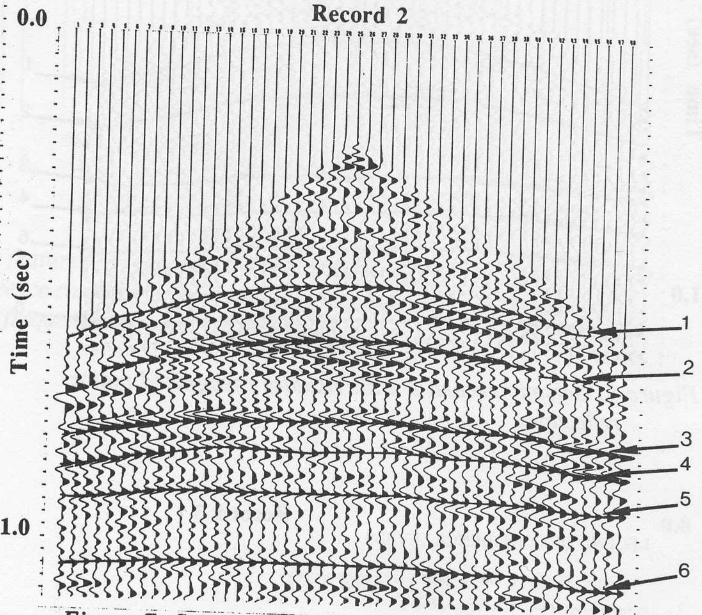

Figure 7: Seismic record 2 with visually recognized major reflections.

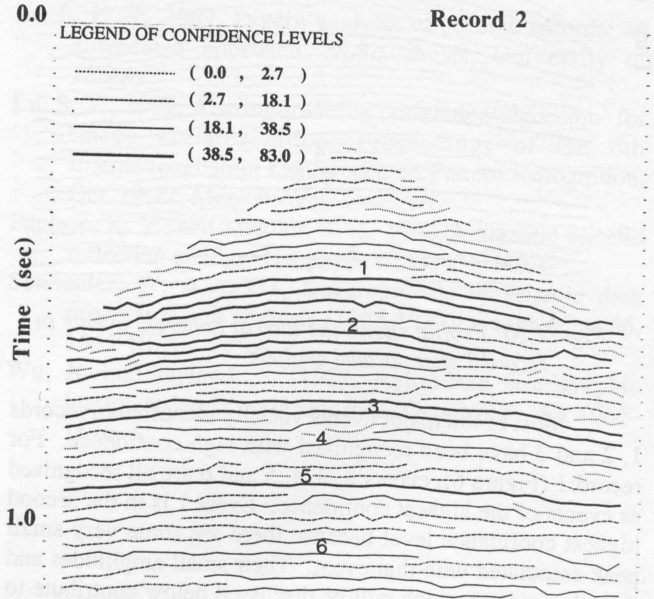

Figure 8: Profile shows the lateral coherent events in record 2 after parallel analysis.

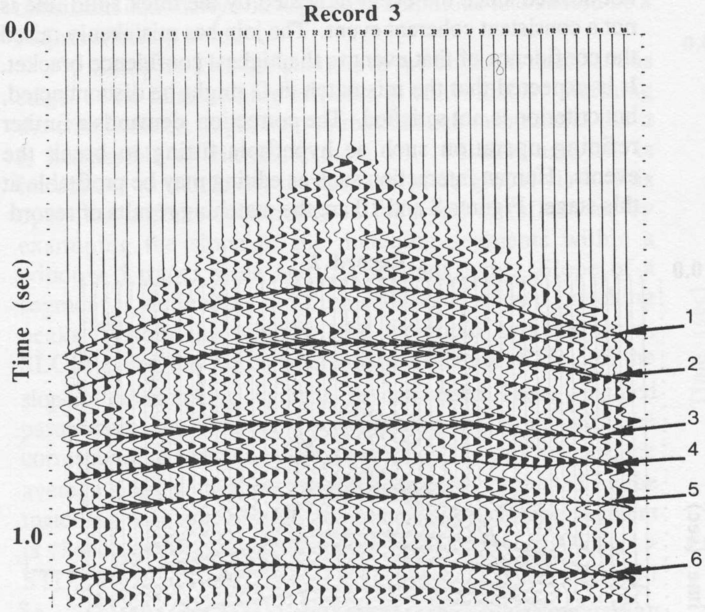

Figure 9: Seismic record 3 with visually recognized major reflections.

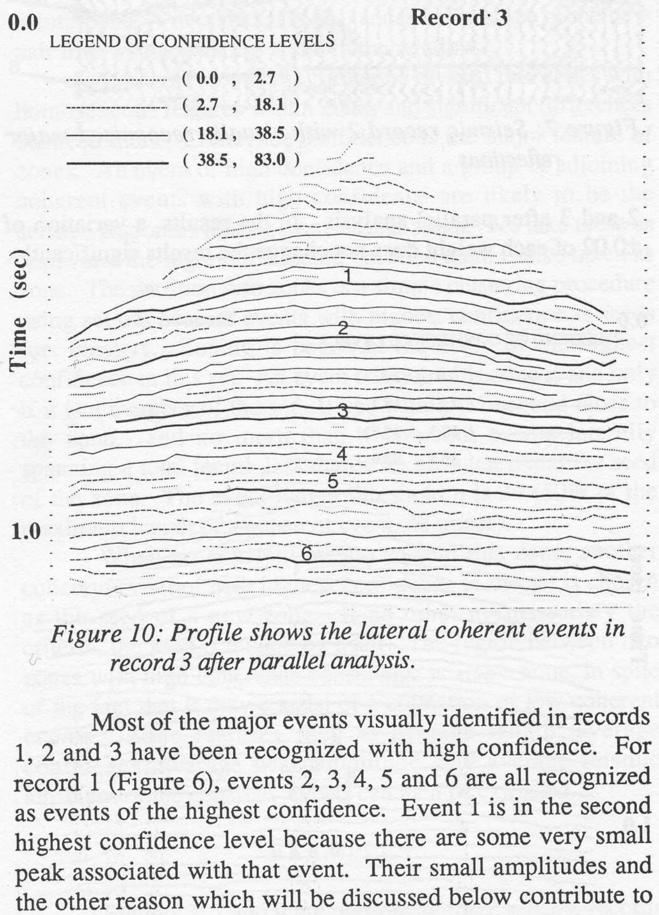

Figure 10: Profile shows the lateral coherent events in record 3 after parallel analysis.

Most of the major events visually identified in records 1, 2 and 3 have been recognized with high confidence. For record 1 (Figure 6), events 2, 3, 4, 5 and 6 are all recognized as events of the highest confidence. Event 1 is in the second highest confidence level because there are some very small peak associated with that event. Their small amplitudes and the other reason which will be discussed below contribute to the overall lower confidence of the whole event. Nevertheless, extracting this event is not a major problem. In record 2 (Figure 8), however, tracing event 4 has some difficulty. Some places in the event are incorrectly matched and the inconsistency is most prominent where the single peak splits into two local peaks. Event 6 has the same comment just made.In record 3 (Figure 10), event 4 has been consistently traced as a series of short segments. Observe that shallow reflections have lower coherence confidence values although they are good coherent events. The location terms of the cost functions imply that coherent events are horizontal. Due to the greater moveout time (moveout time is the time difference between a ray travelling obliquely at an angle and a ray travelling vertically downward), the shallow reflections have higher location costs than deep ones and this reduces their coherence confidence.

Each record is then represented at midpoint (trace 25) as a string of zones with features (Figure 11). These zones are determined by the clustering algorithm described above. Record 1 is divided into 13 zones; record 2, 9 zones; record 3, 11 zones. There is a general trend of decreasing instantaneous frequency with arrival time. Zone 9 of record 1 (Figure 11A) has a lower frequency than the deeper part. This character can be traced to zone 7 in record 2 (Figure 11B) and record 3 (Figure 11C) and is helpful to fix the reference in zone identification.

The zones in three records show good correlation. Zones 2, 4, 6, and 8 in record 2 can be followed as zones 2, 4, 6, and 8 in record 3. There is some difficulty in correlating record 1 with the other two records because some events have been erroneously recognized as major events; nevertheless, zones 2, 8, 10 and 12 correlate with zones 4, 6, 8 and 10 in record 3. These zones appear to correspond to lithological units such as the Upper Devonian Interval (4 in B), and the top of the Winnipegosis (10 in A and 8 in B and C) as discussed in an integrated geologic/geophysical study of the area (Gendzwill, 1978).

CONCLUSION

The automated seismic record analysis described in this paper attempts to emulate the procedure of visual analysis through the use of the pattern recognition techniques. In the automated matching procedure, dynamic matching such as a string-to-string matching algorithm has more advantages than the conventional cross correlation technique (Le, 1987). The former approach utilizes far more structural information than just a statistical index. Detailed matching result usually arises from such a scheme. Confidence measure, which is an indication of the uncertainty of the automated decision, is used to extract the major events. Iterative refining operation and human interaction can improve the algorithm's performance. A seismic record can be condensed to a response curve of average features and zoning provides a quick way to correlate the zones with lithological units. The correlation process can be done by string-to-string matching in the same fashion that works for well log analysis (Wu and Nyland, 1987). The algorithm is not, at the present state, typically knowledge-based, but it can be used as a framework upon which more knowledge and rules can be accumulated. The success of such a knowledge-based recognition system will not only free interpreters from routine picking process but raise the level and productivity of seismic data processing and interpretation.

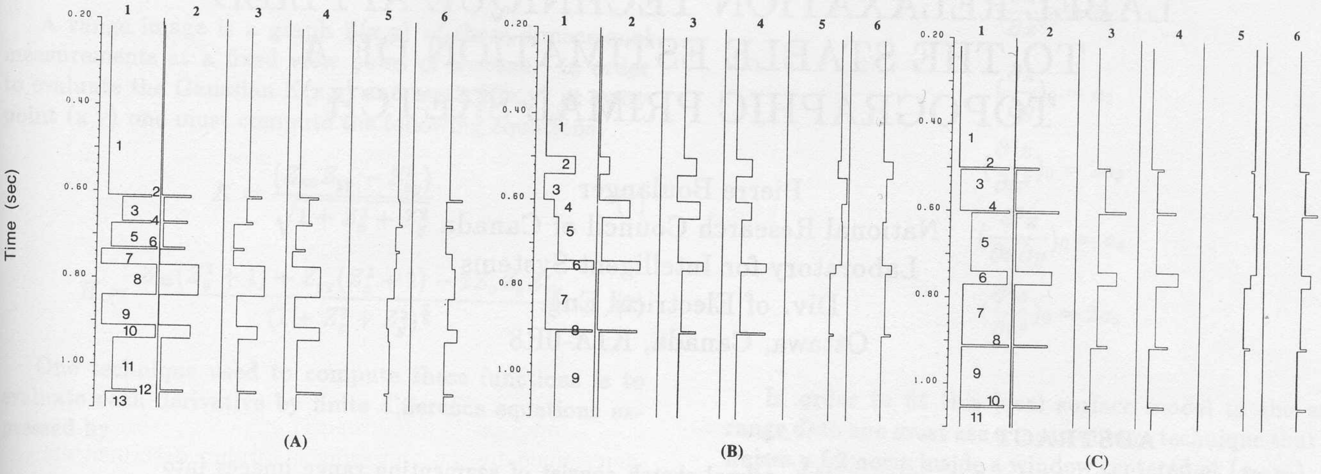

Figure 11: Relative curves of average features (A) record 1, (B) record 2 and (3) record 3. 1=length; 2=coherence confidence; 3=amplitude; 4=matching confidence; 5=instantaneous frequency; 6=envelope amplitude.

ACKNOWLEDGEMENTS

This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Alberta Oil Sands Technology and Research Authority (AOSTRA). X. Wu of the Institute of Mineral Deposits, Chinese Academy of Geosciences, Beijing, The Peoples' Republic of China and Andrei Gabrielov of the O.Yu Schmidt Institute for Physics of the Earth, Academy of Sciences of the USSR contributed incisive discussion and critical review of this work. One of the authors (LL) was the recipient of NSERC and CSEG scholarships.

REFERENCES

Fu, K. S., 1974, Syntactic methods in pattern recognition: Academic Press Inc.

Gendzwill, D.J., 1978, Winnipegosis mounds and Prairie Evaporite Formation of Saskatchewan - seismic study: American Association of Petroleum Geologists Bulletin, **62**, 73-86.

Le, L. H. T., 1987, Pattern analysis of seismic records: an automated approach, M.Sc. thesis, University of Alberta.

Lu, S. Y., 1982, A string-to-string correlation algorithm for image skeletonization: Proceedings of the 6th International Joint Conference on Pattern Recognition, Oct. 19-22, Munich, West Germany.

Paulson, K. V., and Merdler, S. C., 1968, Automatic seismic reflection picking: Geophysics, **33**, 431-440.

Schneider, W. A., 1971, Developments in seismic data processing and analysis (1968-1970): Geophysics, **36**, 1043-1073.

Wu, X. and Nyland, E., 1987, Automated stratigraphic interpretation of well logs data: Geophysics, **52**, 1665-1676.