

## APPLICATION OF PATTERN RECOGNITION THEORY TO IDENTIFICATION OF THE ROVIBRATIONAL SPECTRAL LINES

A.P. Shcherbakov

*Institute of Atmospheric Optics,  
Siberian Branch of the Russian Academy of Sciences, Tomsk*

*Received March 3, 1997*

*The methods of pattern recognition theory are applied to identification of lines in the molecular rovibrational spectra. The method proposed is based on recognizing systems of spectral lines meeting the Ridberg-Rits combination rule and most close to those calculated using some estimates of rotational constants and transition moments. The expert system was created for analysis of spectra of asymmetric top molecules. This system can automatically seek and assign spectral lines. It is based on the Watson effective vibrational Hamiltonian, the Pade-Borel approximation, generation functions, and allows for the resonance interactions. To construct the decision rule during learning, the system uses the method of potential functions that allows the reconstruction of the decision rule to be done using learning sequence.*

### INTRODUCTION

This paper reports about the expert system developed in the Lab of Molecular Spectroscopy at IAO SB RAS. This system makes the assignment of rovibrational spectral lines automatically.

For a successful solution of the atmospheric optics problems, one needs to know spectral characteristics of the absorption spectral lines of atmospheric gases and pollutants, including their quantum identification.<sup>1</sup> The use of spectroscopic data with wrong identification can cause misinterpretation of the results of atmospheric experiments. For example, in Ref. 2 the anomalously large shift of H<sub>2</sub>O spectral line by air pressure was discovered. This line lying at 13947.2608 cm<sup>-1</sup> was assigned to the [624](301) ← [523](000) transition. As recent analysis has shown, in fact this line a doublet, and so large shift was due to deformation of the sum contour at increasing pressure.<sup>3</sup> It should be noted that misassignment of spectrum is rather frequent, especially in the cases with weak spectra.

In this paper, line assignment is considered as a separate problem of molecular spectroscopy having certain specific features like: recognition under condition that existing theoretical estimates of line position and intensity are not sufficiently accurate and measured spectral line parameters are noisy. In this paper, to solve this problem, we use well developed methods of the pattern recognition theory. The necessity of solving the problems of such a type is rather obvious and is due to the following circumstances.

The methods of line assignment in rovibrational (RV) spectra are based on general quantum-mechanics

rules, but in practice they are determined by how the spectrum is complex. Thus, in simple cases, for example in the fundamental absorption bands of heteronuclear diatomic molecules (HF, HCl, CO, etc.), when there are simple regularities in line position in *P*- and *R*- branches, lines can be assigned by their position relative to the band center without calculating energy levels. In more complicated cases, for example in spectra of asymmetric top molecules, preliminary analysis of line positions and their relative intensities is required as well as the determination of stable (relative to small variations of spectroscopic constants) spectral elements. As a consequence, energy levels and corresponding wave functions should be calculated.

Different methods for assignment of lines in spectra of linear, symmetric top, asymmetric top, and spherical top molecules are described in Ref. 4.

In high vibrational energy levels, when the energy of rotational-vibrational interaction, anharmonic constants become comparable with the vibrational energy, simple methods of line assignment following from the model of small vibrations prove to be inapplicable. In spectra of high vibrational overtones, characteristic features (position of the strongest lines, series of doublets, etc.) may vary from band to band. In such cases, the methods of assignment become more complex, that in fact necessitates development of some general versatile methods for making the assignment.

Another part of the problem under consideration is in the following. Multiatomic molecules: water vapor, methane, ozone, being of great importance in atmospheric studies have complex absorption spectra comprising thousands and even dozens of thousands lines. Analysis of such spectra and line assignment is

too cumbersome, and as P. Jensen<sup>5</sup> said “the time needed for spectrum recording is negligibly small in comparison to the time for its theoretical analysis. B In this connection, it becomes necessary to develop such means for automated processing, which would allow one to quickly and efficiently work with such spectra, assign spectra lines, solve inverse problems and compile data bases.

In this paper we present an expert system that uses, in contrast to similar ones,<sup>6-9</sup> the algorithms of pattern recognition theory<sup>13-15</sup> and recognition learning that allows obtaining necessary flexibility of the system – possibility to use different methods of assignment with the combination rules applied to analysis of both strong bands and weak, “single lines corresponding to transitions to states with large values of angular momentum or lines of weak vibrational bands, for which the method of combination differences is inapplicable.

## 2. QUALITATIVE CRITERIA FOR LINE ASSIGNMENT

It should be noted that only *ab initio* calculations may give reliable line assignment. However, at present *ab initio* calculations of line centers and intensities in RV spectra of multiatomic molecules are sufficiently accurate for line assignment by simple comparison with the observed characteristics. In the general case, the calculation using empirical functions of potential energy, molecular dipole moment, in its turn, also cannot be directly used for assignment, especially for lines corresponding to transitions to high-excited RV states, when errors in the line position prediction can reach several dozens  $\text{cm}^{-1}$ . Therefore the absence of accurate information must be compensated by using other methods of assignment along with the calculation.

Such methods of assignment are based on search of characteristic peculiarities in spectrum, for example, doublet series with 1:3 intensity ratio (for water-type molecules), search of subbands (in symmetric top molecules), strong Q-branches for linear molecules.<sup>4</sup>

When considering the problem of assignment, general criteria of line assignment should be formulated.

According to the Ridberg-Rits combination rule (as applied to RV spectra of multiatomic molecules), spectrum contains groups of lines corresponding to transitions from different initial states to the same final state. Figure 1 shows, as an example, rotational energy levels of upper and lower vibrational states. Thus, according to the combination rule, we have to find, in the spectrum, the system of lines, whose centers by pairs obey the relationship

$$\left| v_{n_1}^c - v_{n_2}^c \right| - \left| v_{i_1}^c - v_{i_2}^c \right| \leq \Delta v_{n_1} + \Delta v_{n_2}, \quad (1)$$

where  $v_{n_1}^c$ ,  $v_{n_2}^c$ ,  $v_{i_1}^c$ ,  $v_{i_2}^c$ ,  $v_{n_2}^c$  are the calculated and measured line centers;  $n_1$  and  $n_2$  are the line numbers;

$\Delta v_n$  are measurement errors;  $i_1$  and  $i_2$  are the quantum numbers of the initial levels of the transitions  $f \leftarrow i_1$  and  $f \leftarrow i_2$ , respectively. Here it is supposed that the energy of the initial levels (to be certain, we consider it as a ground vibrational state) is determined with higher accuracy than  $\Delta v_n$ . Note that the differences in Eq. (1) are independent of errors in calculation of the upper levels.

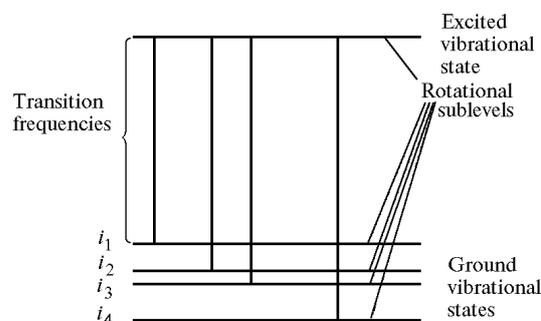


FIG. 1. An example demonstrating the work of combining rule.

In Ref. 6 some general criteria were proposed that can be applied to RV line assignment. The problem is to determine such parameters  $\chi$  of the effective Hamiltonian and parameters  $\mu$  of the effective operator of the dipole moment, such that for every calculated line with the calculated intensity  $S_{if}^c > S_{lim}$  a line can be found in the spectrum, such that

$$\left| S_{if}^c - S_n^e \right| < \Delta S_{if}^c - \Delta S_n^e, \quad (2)$$

$$\left| v_{if}^c - v_n^e \right| < \Delta v_{if}^c - \Delta v_n^e. \quad (3)$$

The parameter  $S_{lim}$  is determined by the device sensitivity threshold.

This criterion supposes line assignment based on a set of accurate spectroscopic constants, and it is, in fact, natural to require the assignment to be done within the framework of a unified theoretical model.

The process of obtaining the spectroscopic constants  $\chi$  and  $\mu$  is the iteration one. First, using the initial approximation  $\chi = \chi_0$  and  $\mu = \mu_0$  the strongest lines are assigned, and the parameters  $\chi$  and  $\mu$  are adjusted by solving the inverse problem. Then, using the adjusted parameters weaker lines are sought and the lines found are also used for further adjustment of the parameters, and so on. The use of the inversion results is in most cases sufficiently reliable method to check the line assignment, because misassignment results in an abrupt increase of standard deviation in the parameter fitting and can be readily detected.

## 3. THE PROBLEM OF PATTERN RECOGNITION

The methods of pattern recognition theory were already applied to solving problems of gas analysis and molecular physics,<sup>10-12</sup> for example, to determination of the molecular structure by analyzing IR absorption

spectra, vibrational-electronic spectra, EPR, NMR, and mass spectra. It should be noted that the below methods and results of recognition theory are classic and widely applied to solution of statistical problems in cybernetics. The description is necessary to give an idea about the implementation of this method in the form of a set of algorithms.

Let the object to be assigned (in our case that can be separate lines, groups of lines forming characteristic structures in spectrum) has  $N$  characteristics, then each object will have its image in an  $N$ -dimensional space of description  $X$ .

Let  $U$  be the space of solution,  $U = \{u_1, u_2, u_3, \dots, u_m\}$ , and  $\Lambda$  be the space of situations,  $\Lambda = \{\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_k\}$ . Formally, to solve the problem of pattern recognition theory (PRT) one must construct the mapping

$$u(x) : X \rightarrow U \tag{4}$$

the best in a certain sense. The quality criterion can be, for example, the minimum of Bayes risk, minimum difference between  $u(x)$  and the mapping  $u^*(x)$  created by the tutor in the sense of some introduced distance between two functions, the minimum number of errors and the minimum error probability, etc.

Depending on the quality criterion, different methods can be used (however, under certain conditions, different methods generate the same classes of procedures). At least two different methods to PRT solution may be cited: the probability approach (estimate of mutual probability density of the space of descriptions  $u$  and the space of situations  $\Lambda$ ) and the deterministic approach (reduction of PRT to the problem of construction of separating surfaces in the space of descriptions). However, both these approaches lead to the same class of procedures.<sup>13</sup>

It follows from the above definitions that construction of the decision rule is equivalent to construction of separating surfaces in the sign space that separate one class of objects from the other ones. For simplicity let us consider the case of two classes, with separating surface denoted as  $f(x)$ .

Let  $f^*(x)$  be the function to be sought, which defines the separating surface. In the training process let us seek  $f(x)$  closest (in a certain sense) to  $f^*(x)$ . To do this one needs to introduce the corresponding distance functional. Wide enough class of functionals can be written in the form

$$J = M_X\{Q(f(x), f^*(x))\}, \tag{5}$$

where  $M_X\{f(x)\} = \int_X f(x) p(x) dx$  is the mathematical expectation,  $Q(x, y)$  is some function (for example,  $J = M_X\{(f(x) - f^*(x))^2\}$  will correspond to rms distance and so on). Let  $f^*(x)$  be presented by the series

$$f(x) = \sum_{i=0}^{\infty} c_i \psi_i(x) = f(x, \mathbf{c}), \tag{6}$$

where  $\psi_i(x)$  is the complete set of orthogonal functions. Then PPR is reduced to finding the coefficients  $c_i$  in the series (6). This problem is solved with the help of known stochastic approximation method by Robbins-Monro<sup>13,14</sup>:

$$\mathbf{c}[n+1] = \mathbf{c}[n] + H[n+1] \nabla_{\mathbf{c}} Q\{f(x[n+1], \mathbf{c}[n]), f^*(x[n+1])\}, \tag{7}$$

where  $\mathbf{c}[n]$  is the vector-column of  $c_i$  coefficients obtained at the  $n$ th (previous) step of training. The elements of the diagonal  $H[n]$  matrix give steps value of the iterative procedure,  $\mathbf{x}[n+1]$  is the vector-column of indicators attached to the  $n+1$ th element of the training sequence,  $\nabla_{\mathbf{c}}$  is the gradient-vector on  $\mathbf{c}$ -parameters. The conditions and proofs of the procedure (7) convergence are presented in Refs. 13, 15.

If  $Q(f(x, \mathbf{c}), f^*(x)) = (f(x, \mathbf{c}) - f^*(x))^2$ , then (7) can be presented as

$$\mathbf{c}[n+1] = \mathbf{c}[n] + \gamma^T[n+1] \times \left( \sum_i c_i[n] \psi_i(x[n+1]) - f^*(x[n+1]) \right) \Psi(x[n+1]), \tag{8}$$

where  $\Psi(x[n+1])$  denotes vector-column composed from the  $\psi_i(x)$  basis functions calculated at  $x[n+1]$  point; T stands for transposition.

Equation (7) defines simple and general enough way for separating surface construction during the training process. It is obvious that such an approach can also be easily applied in the case when the objects to be recognized are from several classes. Due to their simplicity and algorithm conveniences the stochastic approximation method procedures are very popular among specialists in the field of automation. They provide the possibility to use the separating function (6) and to correct its parameters, when necessary, at every stage  $n$  using Eq. (7).

The potential function method (PFM) is one of the methods applied to training problems for the separating surface construction (and for the distribution density evaluation also). This method's procedures belong to the Robbins-Monro<sup>14</sup> class of extremization procedures.

In the so called computer realization of the PFM approximation  $f(x)$  of the  $f^*(x)$  function at the  $N+1$ th step is taken as

$$f[n+1](x) = f[n](x) + \gamma[n+1] K(x[n+1], x), \tag{9}$$

where the potential function  $K(x, y)$  satisfies the following conditions:

$$\begin{aligned} K(x, y) &\geq 0, \quad K(x, y) \leq K(x, x), \quad \forall x, y, \\ K(x, y) &< \infty, \quad K(x, y) = K(y, x). \end{aligned} \tag{10}$$

In Eq. (9)  $\gamma [n]$  are the numbers that determine the extent to which the separating surface changes during the training process, they can be called the training doses. The training doses should be chosen in accordance with the following rule:

$$\gamma [n] = \begin{cases} > 0, & \text{if } f [n](x) \leq 0, f^*(x) > 0, \\ < 0, & \text{if } f [n](x) > 0, f^*(x) \leq 0, \\ 0, & \text{in other case;} \end{cases} \quad (11)$$

$$\gamma [n] = r(f(x), f^*(x)),$$

and  $r(x, y)$  is some function such that  $|r(x, y)| \leq A|x - y| + B$ ,  $A$  and  $B$  are constants.

It follows from Eq. (9) that the  $f [n](x)$  function (that means  $f(x)$  at the  $n$ th step of training) can be written in the form

$$f [n](x) = \sum_{i=1}^n \gamma [i] K(x[i], x). \quad (12)$$

Let us take the  $K(x, y)$  as follows:

$$K(x, y) = \sum_{i=1}^M \lambda_i^2 \psi_i(x) \psi_i(y), \quad (13)$$

where  $\psi_i(x)$ ,  $i = \overline{1, \infty}$  is the set of orthonormal functions. One can prove that such a function satisfies the above mentioned conditions (10) (see Ref. 13). Let us denote  $\varphi_i(x) = \lambda_i \psi_i(x)$ , then  $f(x) = \sum c_i \varphi_i(x)$ . The procedure

$$\mathbf{c}[n + 1] = \mathbf{c}[n] + \gamma [n + 1] \Phi(x[n + 1]), \quad (14)$$

where  $\gamma [i]$  is the numerical sequence satisfying the conditions (11), is called the perceptron realization of the PFM. The PFM procedures extremize the functional

$$J = M_x \{f(x) [\text{sign } f(x) - \text{sign } f^*(x)]\}. \quad (15)$$

It is also necessary to note that the above described methods work only if the hypothesis assuming that the patterns are compact is valid. To say briefly, the idea of compact patterns presented in Ref. 15 denotes, that the separating surface sought should be smooth enough since only in this case one can manage with comparably small number of the terms in the series (6).

The Rozenblatt's perceptron<sup>15</sup> is one of the realizations of the PFM. The Rozenblatt perceptron uses, in the potential function methods terminology, the basis functions of the form:

$$\varphi_i(x) = \lambda_i \psi_i(x) = \text{sign} \left[ \theta \left( \sum_j a_j^i x_j \right) \right], \quad (16)$$

where  $\theta(x)$  is some function,  $a_j^i$  are numerical

coefficients. The orthogonality of this functions set is shown in Ref. 13. The training algorithm of the Rozenblatt's perceptron second layer  $u_k = \sum_i \mu_i^k \varphi_i(x)$ ,  $k = \overline{1, K}$ , where  $K$  is the number of classes, has the following:

$$\mu^k[n + 1] = \mu^k[n] + T_k[n + 1] \Phi(x[n + 1]). \quad (17)$$

Here  $\mu^k[n]$  is  $\mu^k$  coefficients vector obtained after the  $n$ th step of training,  $\Phi(\cdot)$  is the basis functions vector at the  $x$  point, and  $T_k$  numbers are chosen in accordance with the rule:

$$T_k = \begin{cases} 1, & \text{if } u_k(x) \leq 0, x \in U_k \\ -1, & \text{if } u_k(x) > 0, x \notin U_k \\ 0. & \end{cases} \quad (18)$$

A more detailed description of the method may be found in Ref. 15.

#### 4. THE APPLICATION OF THE PATTERN RECOGNITION METHODS TO ANALYSIS OF SPECTRA

As was already noted, whichever identification method is used, the comparison and evaluation are necessary of the observed and calculated spectral features. Usually several line sets can be found in the spectrum, similar to the sought one, and each of the variants is analyzed using some additional criteria. The only correct set is chosen by a spectroscopist using a number of indicators, i.e. definite combinations of the line characteristics (intensity correlation, deviation of the observed centers and intensities from calculated ones; accuracy of the combination rule fulfillment in comparison with the measurements precision, and so on – can serve as an example). Such a problem is a typical one for the pattern recognition, that is why the well elaborated pattern recognition theory and mathematical statistics methods and algorithms can be naturally applied to the spectrum analysis automation. The application of the pattern recognition methods allows one to introduce quantitative criteria of the “analogy” of the recognized spectral elements with some theoretical ideas, and, hence, to verify the assignment of lines.

To apply the pattern recognition theory it is necessary to specify the objects (i.e. the characteristic features) to be recognized and their indicators. The analysis shows that to make the line assignment it is necessary to use simultaneously the Ridberg-Ritz combination rule which is the only accurate one free of model ideas (molecular Hamiltonian, potential and dipole moment functions parameters and so on), along with the calculated lines characteristics (their positions and intensities). This provides reliability and generality of the recognition rules. At the same time other peculiarities which can be in the spectral doublets,

strongest lines, subbands, and so on, depend on the molecular type, rotational constants, accidental resonances and so on.

Thus the pattern recognition problem reduces to the determination of the “similarityB of the observed and calculated characteristics within the array of lines satisfying approximately the combination rule. It is obvious that in order to obtain a flexible system the use of algorithms with training to recognition is desirable in this case.

Each line from the set under consideration has definite characteristics, for example, deviation from the calculation on frequency, intensity, accuracy of the combination rule fulfillment and so on. The following characteristics have been found to be expedient as an indicators during the system creating process:

1) deviation of the measured line frequency from the calculated one

$$x_0 = \ln(1 + |v^e - v^c| k_0^0) k_0^1; \tag{19}$$

2) deviation of the measured intensity from the calculated one

$$x_1 = \ln(1 + |s^e - s^c| k_1^0) k_1^1; \tag{20}$$

3) accuracy of the combination rule fulfillment for a given line

$$x_2 = \ln(1 + |E^c - \bar{E}^e| k_2^0) k_2^1; \tag{21}$$

4) deviation of the measured line intensity from the calculated average (over the total set) one

$$x_3 = \ln(1 + |s^e - \bar{S}^c| k_3^0) k_3^1; \tag{22}$$

5) proportionality, on the average, of the measured and calculated line intensity

$$x_4 = 1 - K\{s^c, s^e\}, \tag{23}$$

where  $K\{s^c, s^e\}$  is the coefficient of correlation between the calculated and found intensities for the set checked:

$$K\{x, y\} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2}};$$

6) the ratio of the summed intensity for all the lines found to the calculated total intensity

$$x_5 = \ln \left( 1 + \ln \left( \left| \frac{\sum s^e}{\sum s^c} \right| \right) k_4^0 \right) k_4^1; \tag{24}$$

7) accuracy of the combination rule validity throughout the group.

$$x_6 = \sqrt{\frac{1}{n-1} \sum_i (E_i^e - \bar{E}^e)^2} \times 1000. \tag{25}$$

The constants  $k_j^i$  are chosen so that the range of indicators change within the segment  $[-10; 10]$  containing the coordinate system origin. This provides the classes reflection being compact in the indicators space. The logarithms in Eqs. (19)–(25) were used in order to narrow the ranges of possible  $x_n$  values.

Training and checking indicators informative ability were carried out using H<sub>2</sub>O high resolution absorption spectrum in 1.4 μm region as an example with line identification being previously known.<sup>16</sup> The indicators selection was made in turn using the  $\nu_1 + \nu_3$  strong band as an example. All indicators, but one, were dropped and by varying  $k_j^i$  coefficients it was achieved that specifier indicated correct variants.

The  $\gamma[n]$  values in Eq. (11) were taken in the form

$$\gamma[n+1] = \begin{cases} +r & \text{if } f(x, \mathbf{c}) \leq 0 \text{ and line is correct} \\ 0 & \\ -r & \text{if } f(x, \mathbf{c}) > 0 \text{ and line is incorrect} \end{cases} \tag{26}$$

$$r[n] = 0.2 f(x[n+1]) + 1.0 \cdot 10^{-7}. \tag{27}$$

(After submission of this paper new expressions have been derived instead of Eqs. (26) and (27) providing more stable work of the algorithm.)

The basis functions in the vectors  $\mathbf{X} = \mathbf{R}_7$  space are the product of the form

$$\psi_n(x) = \prod_{i_1+i_2+\dots+i_{10} \leq 3}^{s=1.7} \psi_{i_s}(x_s), \tag{28}$$

where  $\psi_i(x) = N_i \exp\{-x^2/2\} H_i(x)$ ;  $H_i(x)$  are Hermitian polynomials;  $N_i = \sqrt{2^{i+1}/i!}$  are the normalizing factors. The  $\lambda_i$  values in (13) were taken to be equal to 1.

The Rozenblatt perceptron has also been tested in the expert system in addition to the above described realizations. The vector space  $\chi$  was taken as the X-description space having the form

$$\chi = (\chi_1^1, \chi_2^1, \dots, \chi_{20}^1, \chi_1^2, \chi_2^2, \dots, \chi_{20}^2, \dots, \chi_1^6, \chi_2^6, \dots, \chi_{20}^6), \tag{29}$$

$\chi_j^i$  components,  $K = \overline{1, 20}$  are calculated as follows:

$$\chi_j^i(x) = \begin{cases} 1 & \text{if } x_i \in [-10 + K - 10 + K + 1], \\ 0 & \text{if } x_i \notin [-10 + K - 10 + K + 1]; \end{cases} \tag{30}$$

$x_i$  values are calculated using formulas (19)–(35).

In PFM terminology the basis functions were taken for a given case as

$$\varphi_k(x) = \text{sign} \left[ \sum_{j=1}^{20} \sum_{i=1}^6 a_{ij}^k \Theta_{[-10+j; -10+j+1]}(x_i) \right], \quad (31)$$

where  $k = \overline{1, 120}$ ;  $\Theta_{[a; b]}(x)$  is the indicator of the  $[a; b]$  segment, it equals to 1, if  $x \in [a; b]$ , and equals to 0 in other cases;  $a_{ij}^k = \delta_{k(20i+j)}$ ;  $\delta_{ij}$  is the Kronecker symbol.

The transitions recognition operates as follows. At first search is carried out of all lines satisfying the combination rule (1). The variants found are processed with the specifier that finds, and marks the correct ones; then the regime is switched on of the variants view, and the best one (containing line with the largest  $f(x)$  function value) is displayed on the screen.

## 5. EXPERT SYSTEM FOR ROVIBRATIONAL LINES ASSIGNMENT

Using the above described methods expert system has been created providing spectra assignment in an automated and semiautomated regimes. The block-diagram of the system is presented in Fig. 2.

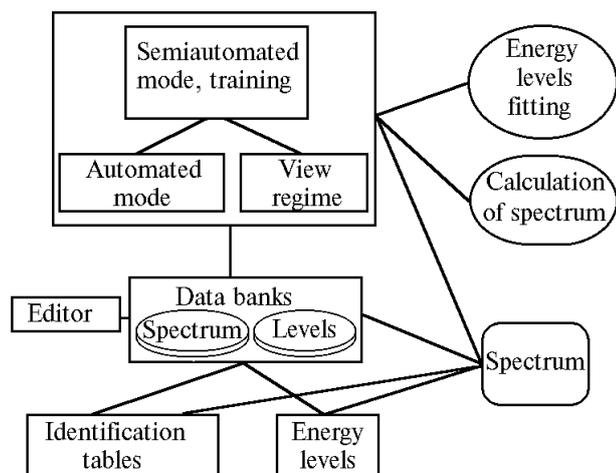


FIG. 2. Block-diagram of the expert system for rovibrational line assignment.

The system contains the blocks for calculation of the energy levels, wave functions, line positions and intensities; the recognition and training block; special-purpose databases; the blocks responsible for data bases correction and control. Special initialization block contains the programs providing spectra and the necessary files preparation and control before running.

The recognition blocks are adaptable that means that the built-in dialogue means between the system and a user makes it easy to make additional training of the system for a concrete spectrum or to train it anew during the work.

The system can operate with molecules of different types (two cases of  $C_s$  and  $C_{2v}$  symmetry are realized in the present version). Watson-type effective Hamiltonian,<sup>17</sup> Pade-Borel approximants,<sup>18</sup> and generating functions method<sup>19</sup> can be used for the energy levels, line frequencies and intensities calculations. The maximum size of the analyzed spectrum is 32000 lines.

The upper states Hamiltonian constants (the initial approximation), ground state and dipole moment parameters, and analyzed spectrum, - are the initial data for the system operation.

The results are accumulated in a special databases for the obtained experimental upper states levels and for the spectrum along with the assignment. There is a possibility of obtaining completed tables of "experimental Energy levels, and files containing centers, intensities, and quantum identification of the lines.

The operation of the system is performed as follows. At the beginning of the work with the help of the initialization program a user sets the names of files containing the spectrum (line frequencies and intensities list), sets the model parameters of upper and lower states, dipole moment constants, names databases, identifies the molecule type, kind of the model used, the search mode, and indicates some conditions of the experiment. Then it is possible to enter the spectrum analysis regime. At first, rovibrational state to work with is fixed. Then the user can switch on the line search using lower state energy levels either calculated or experimental.

In the case when at least one of the combination difference variants is found the variants view regime starts and the best variant (if the system is trained) is placed on the screen as a set of quantum numbers, lower state energy levels, calculated and experimental frequencies and intensities, upper state experimental energy levels appearing as a sum of lower state energy and measured line positions. In this mode of operation it is possible to view, correct, add or drop the found lines assignments, include in or exclude from the energy levels database the upper state experimental energy level value obtained as an average for all lines marked by the user. Each line can have up to three assignments.

If none variant is found, the regime starts of the nearest line view, where it is possible to identify a single lines. In this regime there it is also possible to work with the databases like in the combination differences view mode. Both modes allow the spectrum view to be performed in succession. Regime of the successive viewing of spectrum also allows the work with databases. After returning from the view modes it is possible to choose another rovibrational state and to continue the work, or to call the initialization program (mentioned above) and to choose the search regime; to view the databases contents with the results, and to place their contents into a text file in the form of tables. It is possible also to switch on the automatic mode, beginning from the current state and up to some

other state, using different regimes for isolated lines search (drop, analyze, or address to a user).

This system has been used for line identification of HDO, D<sub>2</sub>O, H<sub>2</sub>O, H<sub>2</sub><sup>18</sup>O absorption spectra.

#### ACKNOWLEDGMENT

The author is grateful to L.N. Sinitsa, A.D. Bykov, O.V. Naumenko, and T.M. Petrova for assistance, helpful remarks, and for the programs enabling line parameters calculation.

This work was supported by Russian Foundation for Basic Research, grant N-96-07-89321.

#### REFERENCES

1. V.E. Zuev, Yu.S. Makushkin, and Yu.N. Ponomarev, *Modern Problems in the Atmospheric Optics*. Vol. 3. *Spectroscopy of the Atmosphere* (Gidrometeoizdat, Leningrad, 1987), 247 pp.
2. J. Bösenberg, *Appl. Opt.* **24**, No. 21, 3531-3534 (1985).
3. B.E. Grossman and E.V. Browell, *J. Mol. Spectrosc.* **138**, No. 3, 562-595 (1985).
4. G. Gerzberg, *Vibration-Rotation Spectra of Polyatomic Molecules* (Foreign Literature Press, Moscow, 1949), 403 pp.
5. P. Jensen, BRUKER FT-IR application note 35.
6. A.C. Pine, G. Dresselhaus, B. Palm, W. Davies, and S.A. Glought, *J. Mol. Spectrosc.* **67**, Nos. 1-3, 386-415 (1977).
7. I.N. Kozin, S. Klee, and P. Jensen, *J. Mol. Spectrosc.* **157**, 409-422 (1993).
8. O.V. Naumenko, "Methods and some results of the inverse spectroscopic problem solution for highly excited vibrational states of the asymmetric top molecules," *Cand. Phys.- Math. Sci. Dissert.*, Tomsk (1986).
9. S. Urban and J. Behrend, in: *Proceedings of the Forteen Colloquim on High Resolution Molecular Spectroscopy* (1995), p. B30.
10. M.E. Eljashberg, L.A. Gribov, and V.V. Serov, *Molecular Spectral Analysis and Computer* (Nauka, Moscow, 1980), 305 pp.
11. M.Yu. Kataev and A.A. Mitsel, *Avtometrija*, No. 4, 15-20 (1985).
12. M.Yu. Kataev, A.A. Mitsel, and S.P. Tarasova, *Atm. Opt.* **3**, No. 8, 763-771 (1990).
13. M.A. Aizerman, E.I. Braverman, and L.I. Rozonoer, *Potential Function Method in the Computer Training Problems* (Nauka, Moscow, 1970), 850 pp.
14. M.B. Nevelson and R.Z. Khas'minskii, *Stochastic Approximation and Recurent Estimation* (Nauka, Moscow, 1972), 295pp.
15. L.L. Levin, *Introduction Into Pattern Recognition Theory*. (State University, Tomsk, 1982), 250 pp.
16. A. Robert, *Appl. Optics.* **33**, No. 21, 4851-4867 (1994).
17. J.K.G. Watson, *J. Chem. Phys.* **46**, 1935-1948.
18. O.L. Polyanskii, *J. Mol. Spectrosc.* **112**, 79 (1985).
19. V.I. Starikov, S.A. Tashkun, and V.I. Tyuterev, *J. Mol. Spectrosc.* **151**, 130-147 (1992).