Identification of Inorganic Salts and Determination of Their Concentrations in Aqueous Solutions Based on the Valence Raman Band of Water Using Artificial Neural Networks

S. A. Burikov, T. A. Dolenko, V. V. Fadeev, and A. V. Sugonyaev

Physics Faculty, Moscow State University, Leninskie gory, Moscow, 119992 Russia

e-mail: burikov@lid.phys.msu.su, tdolenko@radio-msu.net, fadeev@lid.phys.msu.su, sugonjaev@lid.phys.msu.su

Abstract—The characteristic features of the valence Raman band of water in the solutions of electrolytes are revealed. These features allow the noncontact recognition of the type of salt and the determination of its concentration in aqueous solutions using artificial neural networks.

DOI: 10.1134/S1054661807040141

INTRODUCTION

The interaction of water molecules with inorganic compounds manifests itself as a variation in the vibrational characteristics of the bound water molecules [1–5]. These variations can be detected with the laser Raman scattering spectroscopy (RS). Many authors [1–5] study the most intense valence Raman band of water. This band was used in [4, 5] to solve the RS inverse problem lying in the determination of the concentration of specific salt in aqueous solution. From the theoretical and practical points of view, it is expedient to find specific features in the characteristics of the water Raman spectra related to each salt. Thus, we can formulate the problem of the noncontact identification of various salts and the determination of their partial concentrations in multicomponent aqueous solutions. This problem can be solved due to the fact that different ions exhibit different interaction with water molecules [5, 6].

This work is devoted to the application of artificial neural networks (ANNs) [7] for the precise analysis of Raman spectra and the solution of the RS inverse problem. In the case under study, the application of ANNs makes it possible to recognize the type of the dissolved salt and to increase the sensitivity of the method in the determination of the impurity concentration in water.

METHODS

Aqueous solutions of alkali–halogen salts with the Na^+, K^+, and NH_4^+ cations and the F^-, Cl^-, Br^-, and I^- anions in the entire range of solubility (at a temperature of 25°C) serve as the objects under study. For the Raman excitation, we use an argon-ion laser with the wavelength λ = 488 nm and a power of 350 mW. The Raman spectra of the valence band are recorded with an optical multichannel analyzer. The width of the spectra is 70 nm, and the spectral resolution is 0.14 nm/channel. Figure 1 demonstrates the valence Raman bands of water for the solutions of the NaI and NH_4F salts. The experimental data from [1–5] show that an increase in the salt concentration in water leads to an increase (decrease) in the intensity in the high-frequency (low-frequency) region of the valence band, so that the maximum of the band shifts towards higher frequencies and the half width of the band decreases (NaI in Fig. 1).

Specific features of the valence Raman band of water related to a variation in the ion concentration and

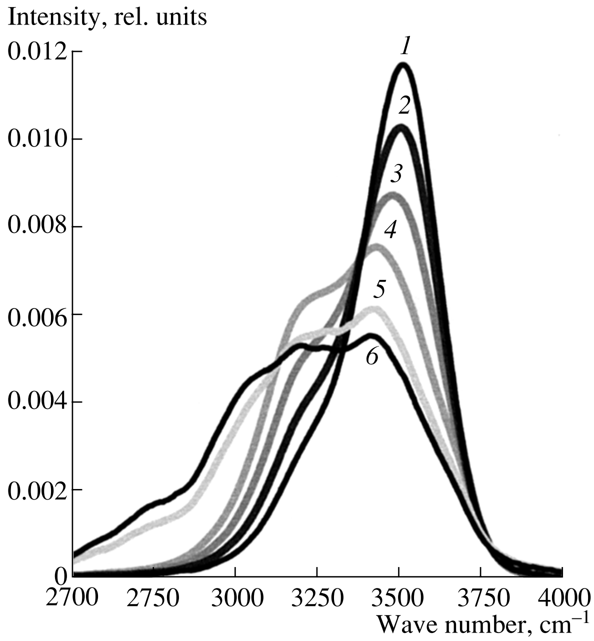

Fig. 1. Valence Raman bands of water in the solutions of electrolytes at various concentrations: (*1*) 6M NaI, (*2*) 4M NaI, (*3*) 2M NaI, (*4*) distillate, (*5*) 7M NH_4F, and (*6*) 12M NH_4F.

Received April 6, 2005

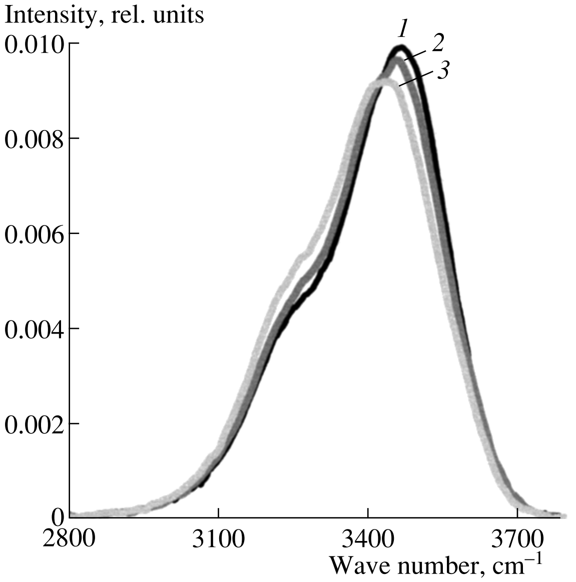

Fig. 2. Valence Raman bands of water for the solutions of potassium salts at a concentration of 3.2 M: (*1*) KI, (*2*) KBr, and (*3*) KCl.

observed in the solutions of fluorides and ammonium salts (Fig. 1) are due to individual features of the NH_4^+ and F^- ions [4, 5].

Figure 2 demonstrates the Raman spectra of water for the solutions of various potassium salts with equal concentrations. It is seen that the Raman intensity of water in the solutions depends on both the salt concentration and type. The results from [1–5] show that the effect of anions on the water Raman band is significantly stronger than the effect of cations. The effect of ions on the Raman spectrum can be ordered in the following way: $Cl^- < Br^- < I^-$ (Fig. 2). This series is in agreement with the ion hydration theory [4–6]. A similar series of the cation effect on the valence Raman band is missing, since the effect is insignificant.

Note proposals from [4, 5] on the determination of concentration C of the known salt in a one-component solution based on the concentration dependence of the valence Raman band of water. Parameter χ equal to the intensity ratio of the high-frequency (ν_{hf}) and low-frequency (ν_{lf}) components of the band serves as a quantitative characteristic of variations in the valence Raman band. The ν_{hf} and ν_{lf} frequencies are chosen using the critical points of the first derivative for the valence Raman band of distilled water and are fixed. Using the calibration curves $\chi(C)$, which are close to linear for each salt under study [4, 5], one can determine the salt concentration in a one-component solution. For this method, the accuracy of the concentration measurements is 0.1–0.2 M [4, 5].

The differences in the effect of different salts and their concentrations on the valence Raman band of water make it possible to develop a method for the identification of salts based on their Raman spectra using ANN [7]. We employ ANN to solve a complex problem lying in the determination of the type of dissolved salt and the measurement of its concentration in one-, two-, and three-component aqueous solutions.

Table 1. Results on the salt classification for the examination set

Real	Recognized as a class					Percentage of truth
	NaI	NaCl	KCl	KBr	KI	
NaI	9	0	0	0	0	100
NaCl	1	4	0	0	0	80
KCl	0	0	12	0	1	92
KBr	0	0	10	4	0	29
KI	0	0	0	0	16	100

RESULTS

1. One-Component Aqueous Solutions

At the first stage of the inverse problem, we solve a classification problem to determine which of the five salts (NaI, NaCl, KCl, KBr, and KI) is present in the solution based on the valence Raman band. The corresponding ANN has five outputs for the salts under study. In the training of the network, the value of the first output corresponds to the salt present in the solution, whereas the remaining four outputs equal zero.

In accordance with the conventional ANN procedure, all of the experimental spectra (574 spectra) are divided into training (409), test (109), and examination (57) sets. A three-layer perceptron architecture is used. 1

The results on the salt classification are presented in Table 1. In the cells of the central part of Table 1, we present the distribution of recognized spectra (network outputs) with respect to salts. In the last column, we present the fraction of true results (in percents) for each salt.

Note the following facts. First, four salts (NaI, KI, NaCl, and KCl) are well identified with a probability of 80–100%, whereas the fifth salt (KBr) is often erroneously classified as KCl. This is due to the fact that the effect of the I^- ions on the valence Raman band of water is significantly stronger than the effect of the Cl^- and Br^- ions (see above). Second, the results on the salt recognition for the three data sets (training, test, and examination) are close to each other, which indicates the representative character of the sets.

After the salt identification, we determine its concentration at the second stage. For this purpose, we employ five neural networks (one network per one salt). Each ANN has a single output, which corresponds to the salt concentration. We compare the results obtained with the five-layer perceptron using the dependence 1

Table 2. Mean absolute errors for the determination of salt concentrations with various methods

1

Salt	Mean absolute error, M		
	five-layer perceptron	dependence χ_{12} (C)	GMDH
NaI	0.08	0.10	0.19
NaCl	0.07	0.12	0.22
KCl	0.09	0.10	0.10
KBr	0.06	0.10	0.06
KI	0.06	0.13	0.14

χ(C) and the group method of data handling (GMDH) [8]. GMDH is a simulation algorithm that creates a polynomial model of the dependence under study (the dependence of the Raman intensities at the spectral frequencies on the salt concentration). Table 2 shows the errors of the salt concentration measurement with the
1 above methods. Note that the five-layer perceptron is superior to the χ(C) dependence and GMDH.

2. Two-Component Aqueous Solutions

To determine the concentrations of two salts in a single aqueous solution with ANN based on the valence Raman band of water, we employ two (experimental and quasi-model) approaches [9].

In the first (experimental) approach, all of the experimental Raman spectra of water in the two- and one-component solutions are used for the ANN learning. The sets are low-representative, since the accumulation of large experimental data arrays is a difficult task. In this approach, the learning of the neural network on the experimental curves takes into account various interactions in the solutions. In addition, the network is trained on real apparatus noise.

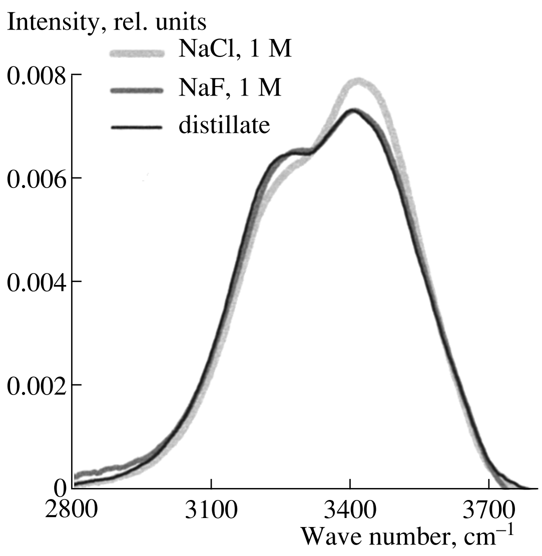

Fig. 3. Valence Raman bands of distilled water and water in the solutions of NaCl and NaF.

In the second (quasi-model) approach [9], the model spectra are used to obtain representative training and test sets. The model spectra represent numerically simulated spectra of the two-component solution that contain the spectra of the one-component solutions with regard to the salt concentrations. Evidently, this approach makes it possible to obtain a sufficient number of curves. However, the accuracy of the solution to the inverse problem strongly depends on the error of the model that is used for the calculation of the curves. In addition, the network is not trained on real noise. Thus, it is expedient to compare the two approaches and to reveal their advantages.

We study two two-component solutions: NaCl + NaF for the concentration range 0–1 M (for each salt) and KI + KCl for the concentration range 0–1.5 M (for each salt).

2.1. NaCl + NaF solution

In this solution, one component (NaCl) significantly affects the valence Raman band of water, whereas the other component (NaF) weakly affects this band. To demonstrate this difference, we present (Fig. 3) the valence Raman bands of distilled water and water in the NaCl and NaF solutions with concentrations of 1 M.

To analyze the two-component (NaCl + NaF) solution using the ANN, we employ the experimental approach. All of the experimental Raman spectra of water for the two-component and one-component solutions of NaCl and NaF (386 spectra) are arbitrarily divided into three sets: training, test, and examination (310, 56, and 20 spectra, respectively).

At the first stage, we solve two inverse one-parameter problems on the determination of the concentrations of each component with neglect of the effect of the other component on the shape of the valence Raman band of water. The experimental curves of the one- and two-component solutions are used for learning of the neural network that has a single output, whose value corresponds to the concentration of either NaCl or NaF. We perform the training of several ANN architectures:
three- and five-layer perceptrons, generalized regres- 2
sion neural network (GRNN) with the sequential and genetic search for the smoothing parameter [10], and
Ward network (three-layer perceptron with various 1
activation functions).

Table 3 shows the results obtained for both one-parameter inverse problems. It is seen that GRNN provides the best accuracy in the determination of the NaCl concentration in the two-component solution with neglect of the effect of the NaF salt on the valence Raman band of water. The Ward network makes it possible to accurately determine the NaF concentration in the two-component solution with neglect of the effect of the NaCl salt on the valence Raman band of water.

To simultaneously determine the concentrations of both salts (NaCl and NaF) in the solution based on the valence Raman band of water, we use a two-output ANN. The output values correspond to the concentrations of the first and second salts. We perform the learning of the following ANN architectures: three- and five-
2 layer perceptrons, GRNN with the sequential and genetic search for the parameters, and Ward network
1 (three-layer perceptron with various activation functions). The results obtained for this two-parameter problem are also presented in Table 3. It is seen that GRNN with both methods for the search for the smoothing parameter provides the highest accuracy. Using GRNN, we can simultaneously determine the NaCl and NaF concentrations in the two-component solution with errors of 0.05 and 0.13 M, respectively. In any case, the accuracy for the NaF salt is lower than the accuracy for the NaCl salt, since fluorides weakly (in comparison with chlorides) affect the shape of the valence Raman band of water.

2.2. KI + KCl solution

In this solution, both components strongly affect the shape of the valence Raman band of water (Fig. 4). It is seen that the effect of potassium iodide on the valence Raman band is stronger than the effect of potassium chloride.

To realize the experimental approach, we randomly divide the experimental Raman spectra of water in the KCl and KI solutions (131 spectra) into training, test, and examinations sets (92, 26, and 13 spectra, respectively). Two ANN outputs correspond to the concentrations of the first and second salts. Evidently, a relatively high accuracy of these parameters indicates that the neural network well knows the characteristic features of both components of the solutions.

The best results in the solution of this problem are
1 obtained with the five-layer perceptron with 32, 16, and 8 neurons in the hidden layers. The experimental approach makes it possible to determine the salt concentration in the two-component solution with accuracies of 0.07 and 0.24 M for KI and KCl, respectively. Obviously, the KI salt, whose effect on the shape of the Raman spectrum of water is stronger, is better identified with the solution to the inverse problem.

To simulate the Raman spectra of water for the two-component solution of the KI and KCl salts, we employ 19 experimental Raman spectra of the KI solution in the concentration range 0.2–3.75 M and 22 spectra of the KCl solution in the concentration range 0.1–3.5 M. We numerically simulate 463 Raman spectra of water in the two-component solution for the concentration range 0–1.5 M on the assumption on the additive effect of salts on water molecules.

Table 3. Mean absolute errors for the concentrations of the NaCl and NaF salts in solutions

ANN architecture	One-parameter problem		Two-parameter problem		
	NaCl	NaF	NaCl	NaF	
Three-layer perceptron (8 neurons)	0.10	0.13	0.08	0.13	1
Three-layer perceptron (64 neurons)	0.08	0.15	0.10	0.14	1
Five-layer perceptron	0.10	0.13	0.09	0.13	1
GRNN with the sequential search for parameters	0.06	0.14	0.05	0.14	
GRNN with the genetic search for parameters	0.06	0.15	0.06	0.13	
Ward network	0.09	0.1	0.08	0.13	

All of the calculated curves are randomly distributed between the training, test, and examination sets (325, 92, and 46 spectra, respectively). To solve the problem, we employ the three-layer perceptron. It is demon- 1
strated that the quasi-model approach enables one to determine the salt concentration in the two-component solution based on the simulated Raman spectra of water with accuracies of 0.01 and 0.02 M for KI and KCl, respectively. This high accuracy results from the fact that we can simulate a large number of the Raman spectra of water and, hence, significantly increase the repre- 3
sentativity of the training and test sets for the network learning when the additivity assumption is well satisfied. As in the previous case, the concentration accuracy depends on the strength of the effect of salt on the shape of the valence Raman band of water.

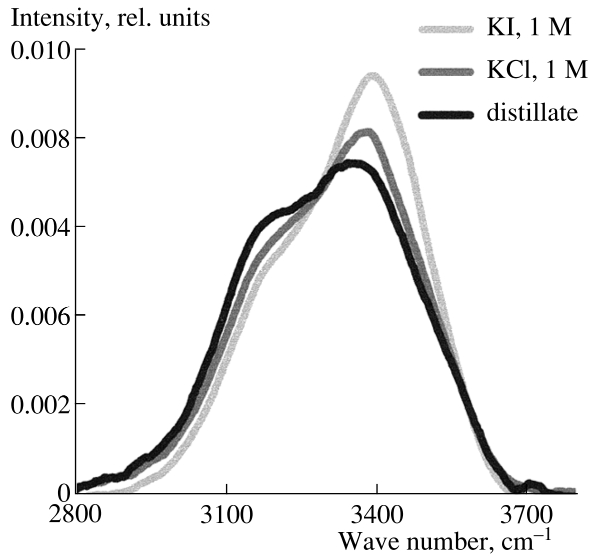

Fig. 4. Valence Raman bands of distilled water and water in the solutions of KCl and KI.

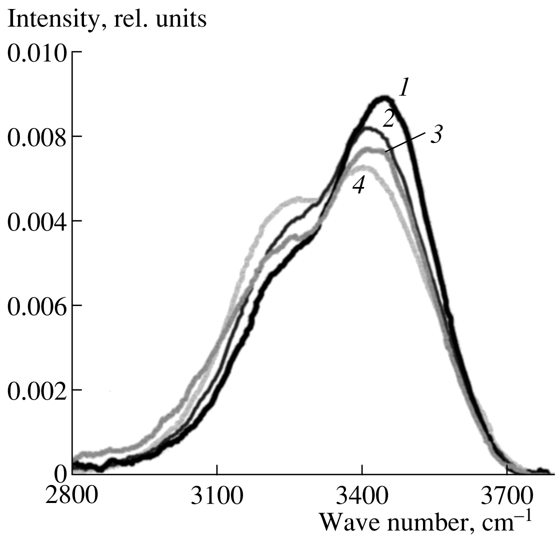

Fig. 5. Valence Raman bands of water in the solutions of (*1*) KI, (*2*) NaCl, and (*3*) NH_4Br and (*4*) of distilled water.

It was demonstrated that, when the simulated curves are used, the accuracy with which the concentrations are determined from the real Raman spectra depends on the error of the calculation model. To verify this statement (the main assumption on the additive effect of two salts on the shape of the valence Raman band of water at concentrations of up to 1.5 M) we supply an examination set that consists of the experimental Raman spectra of water to the neural network that is trained on the simulated spectra. The results of this test show that the quasi-model neural network determines the KI and KCl concentrations in the two-component solution using the real Raman spectra with accuracies of 0.17 and 0.26 M, respectively. The comparison of this result with the result of the experimental approach shows the validity of the assumption on the additive effect of two salts in the two-component solution on the valence Raman band of water at a salt concentration of no greater than 1.5 M.

Table 4. Mean absolute errors for the concentrations of the NaCl, KI, and NH_4Br salts in the three-component solutions

	ANN architecture	Three-parameter problem		
		NaCl	KI	NH_4Br
1	Three-layer perceptron (8 neurons)	0.27	0.14	0.17
1	Three-layer perceptron (64 neurons)	0.26	0.13	0.15
1	Five-layer perceptron (16–8–4 neurons)	0.25	0.14	0.16
1	Five-layer perceptron (32–16–8 neurons)	0.22	0.13	0.13
	GRNN with the sequential search for parameters	0.36	0.22	0.17
	Ward network	0.29	0.14	0.15

3. Three-Component Aqueous Solutions

We simultaneously determine the salt concentrations in three-component solutions (i.e., we solve the three-parameter inverse problem) using ANN. We employ the NaCl, KI, and NH_4Br solutions to prepare mixtures in the concentration range 0–2.4 M with a concentration step of 0.4 M. The salts under study differently affect the shape of the valence Raman band of water (Fig. 5). Potassium iodide provides the strongest effect. As was mentioned, the high-frequency shift of the band is virtually absent in the presence of ammonium salts.

To simultaneously determine the concentrations of the components, we use various architectures of the three-output ANNs, so that each output corresponds to a single salt concentration, and employ the experimental approach. All of the experimental spectra (320 spectra) are randomly divided into the training, test, and examination sets (320, 59, and 20 spectra, respectively). Table 4 demonstrates the results on the salt concentrations in the three-component mixture obtained with the ANN. It is seen that the best concentration accuracy is reached with the five-layer perceptron. The 1
error is lower for the salts with a stronger or more specific effect on the valence Raman band of water. In the case under study, potassium iodide provides the strongest increase in the intensity of the high-frequency component of the water Raman band and the high-frequency shift of this band whereas the presence of ammonium bromide leads to an increase in the intensity of the low-frequency component almost in the absence of the shift of the maximum.

The concentration errors obtained for the three-component solution are slightly worse than the concentration errors for the two-component solution, since (i) the inverse problem involves three (rather than two) desired parameters and (ii) the concentrations of the three components range from 0 to 2.4 M and the concentration step for each salt is 0.4 M, while the concentration ranges are 0–1 and 0–1.5 M and the concentration steps are 0.1 M for the two-component solutions. It follows from the above data on the number of spectra in each ANN set that the data representativities in the 4
ANN sets used to solve the two- and three-parameter problems are almost equal. Nevertheless, the three-parameter inverse problem is more complicated and the representativity of the sets used for its solution must be 3
higher than that for the two-parameter problem. The accuracy with which the three desired parameters are determined can possibly be increased with an increase in the number of the experimental spectra in the training, test, and examination sets.

CONCLUSIONS

A method to identify salts in one-component aqueous solutions and to determine the salt concentrations in one-, two-, and three-component solutions using ANNs is proposed. We plan to employ and to develop this method for the identification of salts and the measurement of their partial concentrations in solutions with a large number of components which can lead to a significant increase in the efficiency of the method, for example, in monitoring of the disposal of sewage that contains inorganic impurities.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research, project no. 03-02-16628.

REFERENCES

1. P. Terpstra, D. Combes, and A. Zwick, "Effect of Salts on Dynamics of Water: a Raman Spectroscopy Study," J. Chem. Phys. **92**, 65 (1989).
2. F. Rull and J. A. De Saja, "Effect of Electrolyte Concentration on the Raman Spectra of Water in Aqueous Solutions," J. Raman Spectroscopy **17**, 167 (1986).
3. K. Furić, I. Ciglenečki, and B. Ćosović, "Raman Spectroscopic Study of Sodium Chloride Water Solutions," J. Mol. Struct. **550–551**, 225 (2000).
4. T. A. Dolenko, I. V. Churina, V. V. Fadeev, and S. M. Glushkov, "Valence Band of Liquid Water Raman Scattering: Some Peculiarities and Applications in the Diagnostics of Water Media," J. Raman Spectroscopy **31**, 863 (2000).
5. S. A. Burikov, T. A. Dolenko, P. A. Velikotnyi, A. V. Sugonyaev, and V. V. Fadeev, "The Effect of Hydration of Ions of Inorganic Salts on the Shape of the Raman Stretching Band of Water," Optika i spektroskopiya **98**, 269 (2005) (Optics and Spectroscopy **98**, 235 (2005)).
6. V. V. Pal'chevskii, *Aqueous Solutions of Electrolytes* (Izdatel'stvo Leningradskogo universiteta, Leningrad, 1984) [in Russian].
7. Mohamad H. Hassoun, *Fundamentals of Artificial Neural Networks* (MIT Press, Cambridge, Massachusetts, 1995).
8. H. Madala and A. Ivakhnenko, *Inductive Learning Algorithms for Complex Systems Modeling* (CRC Press, Boca Raton, 1994), pp. 27–65.
9. I. V. Gerdova, T. A. Dolenko, I. V. Churina, and V. V. Fadeev, "New Approaches in the Solution of Inverse Problems of Laser Spectroscopy Using Artificial Neural Networks," Izvestiya RAN. Seriya fizicheskaya **66**, 1116 (2002).
10. *Neuroshell 2 User Manual* (Ward Systems Group Inc., Frederick, MD, 1998).

Viktor V. Fadeev. Born 1935. Graduated from the Physics Faculty of Moscow State University in 1959. Received candidate's degree in 1967 and doctoral degree in 1983. Professor of the Physics Faculty, Moscow State University. Scientific interests: optics, laser physics, spectroscopy, and inverse problems. Author of more than 300 papers and a discovery diploma. Laureate of the USSR State Prize.

Tat'yana A. Dolenko. Born 1961. Graduated from the Physics Faculty of Moscow State University in 1983. Received candidate's degree in 1987. Senior Researcher of the Physics Faculty, Moscow State University. Scientific interests: laser spectroscopy, Raman spectroscopy of aqueous media, inverse problems, and artificial neural networks. Author of 57 papers and an invention certificate. Laureate of the Lenin Komsomol Prize.

Sergei A. Burikov. Born 1978. Graduated from the Physics Faculty of Moscow State University in 2002. Junior Researcher of the Physics Faculty, Moscow State University. Scientific interests: optics, Raman spectroscopy of aqueous media, inverse problems, and artificial neural networks. Author of 14 papers.

Aleksandr V. Sugonyaev. Born 1982. Graduated from the Physics Faculty of Moscow State University in 2005. Scientific interests: Raman spectroscopy of aqueous media, inverse problems, and artificial neural networks. Author of 5 papers.

SPELL: 1. perceptron, 2. perceptrons, 3. representativity, 4. representativities