EVALUATION OF THE EFFICIENCY OF AMPHETAMINES DETECTION BY CLUSTER ANALYSIS

Ştefănuț CIOCHINĂ[1], Mirela PRAISLER[2]

In this paper we are evaluating three signal processing methods applied for improving the efficiency of the detection of illicit amphetamines performed with Principal Component Analysis. The GC-FTIR spectra of the targeted compounds have been preprocessed by using two functions acting as selective amplifiers and one as an amplifying selector. The efficiency of the detection system was evaluated for each feature weight by using dendrograms obtained by agglomerative clustering. A final ranking of the preprocessing functions was performed based on the cophenetic correlation coefficient. The results confirm the usefulness of the proposed signal preprocessing methods.

Keywords: Principal Component Analysis, agglomerative clustering, cophenetic correlation, amphetamines

1. Introduction

Amphetamines are synthetic drugs that are very popular among drug users, due to their effect of stimulating the central nervous system [1]. The chemical structure of amphetamines contains an aromatic ring linked to an amino group *via* an aliphatic side chain. The type of the aromatic ring substitution influences the biological activity and toxicity of amphetamines [2]. The amphetamine analogues having a mono-substituted aromatic ring form the class of stimulant amphetamines (class code M), and those containing a tri-substituted aromatic ring form the class of hallucinogenic amphetamines (class code T).

The aim of this study is to evaluate three signal preprocessing methods applied for improving the efficiency of the detection of illicit amphetamines. The detection system is based on Principal Component Analysis (PCA) combined with Cluster analysis (CA). The operating principle of CA is the grouping of the objects into clusters based on the Euclidean distance between the spectra of the analyzed compounds. At first, any compound (defined by its spectrum) is considered as a single cluster. Then each such cluster is grouped with the most similar "individual" cluster in a new larger cluster, the number of clusters being,

[1] PhD student, Department of Chemistry, Physics and Environment, "Dunarea de Jos" University of Galati, Romania, e-mail: stefanut27@yahoo.com.
[2] Prof., Department of Chemistry, Physics and Environment, "Dunarea de Jos" University of Galati, Romania, e-mail: Mirela.Praisler@ugal.ro

thus, diminished at each stage. Finally, the algorithm results in a classification tree (dendrogram) [3]. We emphasize that no information about the biological activity or toxicity of the compounds, whose spectra were analyzed, was introduced in the input database.

2. Experimental Part

The training database contains 30 GC-FTIR spectra, 7 of which belong to the main illicit stimulant amphetamines (M), and 6 belong to the main hallucinogenic amphetamines (T). The stimulant amphetamines are: amphetamine (AMP), β-phenylethylamine (BPEA), methamphetamine (MAMP), N-ethyl-amphetamine (EAMP), N-n-propylamphetamine (PAMP), α-phenylethylamine (APEA), N-methyl-α-phenylethylamine (MAPEA). The hallucinogenic amphetamines are: 3,4-methylenedioxyamphetamine, 3,4-methylenedioxy-N-ethylamphetamine, 1 - (3,4-methylenedioxy-phenyl)-2-butanamine, 3,4-methylenedioxymethamphetamine, N-methyl-1-(3,4-methylenedioxy-phenyl)-2-butanamine, 3,4-methylenedioxy-N-hydroxyamphetamine. The experimental conditions in which the spectra have been recorded are detailed in previous studies [4, 5]. The spectral database contains also 17 spectra recorded for non-amphetamine compounds (class code N), which represent substances of toxicological or analytical interest: codeine-PFPA, γ-hydroxybutyric acid (TMS), caffeine, γ-valerolactone, γ-butyrolactone, cadaverine, piracetam, bemegride, β-butyrolactone, γ-hydroxyvaleric acid (TMS), cocaine I.S., γ-butyrolactone artifact 1 (thermal degradation product formed during the chromatographic separation of γ-butyrolactone), nicotamide, cadaverine - HFBA, dextromoramide, prolintane and putrescine [4,5]. The spectra were recorded between 4000 - 600 cm^{-1} and the absorbance was measured every 5 cm^{-1}. Thus, each spectrum is a vector with 681 variables and the spectral database is a matrix with 30×681 entries.

The database was divided into two classes, i.e. the class of positives (M and T) and the class II of negatives (N). A discriminating feature weight w was determined [5]:

$$\omega_\kappa = \frac{\sum \frac{A_I^2}{N_I} + \sum \frac{A_{II}^2}{N_{II}} - 2\sum\sum \frac{A_I A_{II}}{N_I N_{II}}}{\sum \frac{\left(A_I - \overline{A_I}\right)^2}{N_I} + \sum \frac{\left(A_{II} - \overline{A_{II}}\right)^2}{N_{II}}} \qquad (1)$$

where A_I and A_{II} are the absorptions in the GC-FTIR spectra corresponding to the samples of classes I and II, and N_I and N_{II} are the number of samples in classes I and II respectively. The w and w^2 functions increase the intensity of the absorption measured for each variable (wave number) according to its modeling and / or discrimination power ($w > 1$), leaving unchanged the intensity of the irrelevant

wave numbers, for which $w = 1$ [6,7]. Unlike these functions, $(w-1)^2$ amplifies the important absorptions and it eliminates the absorptions at the irrelevant wave numbers, for which $(w-1)^2 = 0$ [4,5]. In other words, w and w^2 act as selective amplifiers, and $(w-1)^2$ as an amplifying selector.

3. Results and discussion

Previous studies have shown that the best results in the detection of amphetamines performed with PCA are obtained by using the PC1 vs. PC2 score plots [6, 7]. The score plots obtained for the unprocessed GC-FTIR spectra and for the w, w^2 and $(w-1)^2$ preprocessed spectra are presented in Fig. 1. They have been compared in order to evaluate the clustering quality and the corresponding performances of the detection system. The PC2 vs. PC1 score plots indicate that, in all the cases, the compounds cluster in three relatively distinct clusters, specific to stimulant amphetamines (class M), hallucinogenic amphetamines (class T) and to non-amphetamines (class N).

The aim of the signal preprocessing was the improvement of the clustering and associated sensitivity and selectivity of the detection system. A feature weight is considered fit for purpose if the selected preprocessing function leads to clusters located as far as possible one from each other and/or to cluster condensation. In our case, as Fig. 1 shows, the main benefit of spectra preprocessing consists of significantly increased distances between the centers of the three modeled clusters (see Fig. 1). However, the intensity of this effect differs from one feature weight to the other. Consequently, a CA was performed for the PC2 vs. PC1 scores plots in the case of the unprocessed spectra, of the spectra preprocessed with the selective amplifiers w and w^2, and for the spectra preprocessed with the amplifying selector $(w-1)^2$, respectively. The CA was performed with the *Matlab 2012a* software [8], by using the agglomerative clustering technique. This unsupervised pattern cognition method generates dendrograms, i.e. two-dimensional diagrams that show the similarities identified in each iterative clustering step [9].

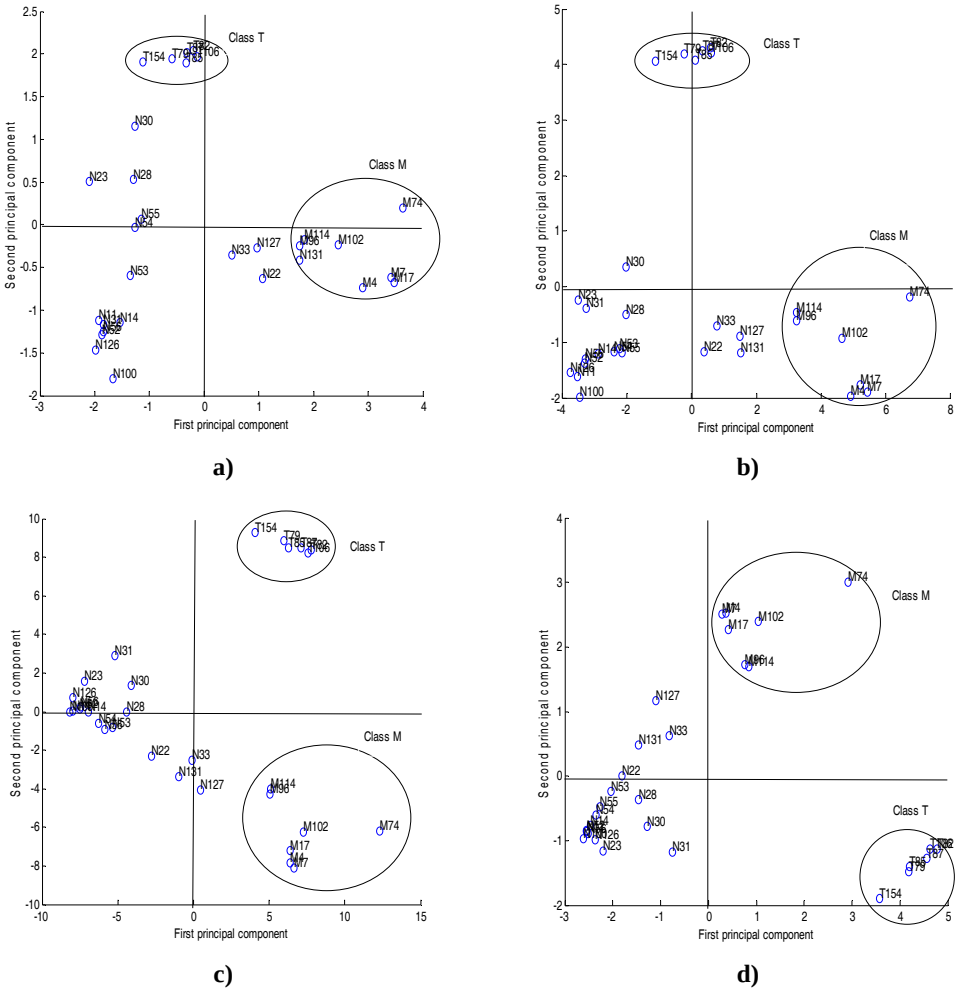


Fig. 1. PC1 vs. PC2 score plots associated to the GC-FTIR spectra of stimulant amphetamines (class M), hallucinogenic amphetamines (class T) and negatives (class N) for: a) unprocessed spectra; b) spectra preprocessed with the *w* selective amplifier; c) spectra preprocessed with the w^2 selective amplifier*;* d) spectra preprocessed with the $(w-1)^2$ amplifying selector.

The dendrograms may be determined by graphical or by geometric methods. The most important graphical methods are the single linkage method, the complete linkage method, the average linkage method and the weighted linkage method. Very good results are also obtained geometrical methods such as the Ward linkage method, the centroids linkage method, and the median linkage method. All these methods have been tested and their performances were evaluated based on the cophenetic correlation coefficient [10]. The best results were obtained with the *average linkage* algorithm. The dendrograms obtained

with this algorithm yield the best sensitivity in terms of negatives and the best selectivity in terms of positives.

The dendrograms obtained with the (PC1, PC2) scores obtained for the case of the unprocessed spectra and of the spectra preprocessed with the w, w^2 and $(w-1)^2$ functions (see Fig. 2, Fig. 3, Fig. 4 and Fig. 5) confirm that the detection of illicit amphetamines occurs according to the biological activity of amphetamines even in the case of the unprocessed GC-FTIR spectra. The same three branches, specific to the stimulant amphetamines (class M), hallucinogenic amphetamines (class T) and non-amphetamine class (class N) are present in all dendrograms. In other words, the classification obtained by hierarchical clustering confirms the results obtained by PCA, which is a non-hierarchical clustering method.

However, in the case of the unprocessed spectra, there are four false M positives (N131-putrescine, N22-cadaverine, N127-prolintane and N33-dextromoramide). This misclassification can be explained by the similarity of the chemical structures of these compounds with elements present in the molecular skeleton of the stimulant amphetamines. For example, both putrescine and cadaverine have in their structures amino groups linked to an aliphatic chain similar to that present in the molecular structure of amphetamines [11]. Prolintane is an antidepressant that may be used legally in many European countries [12]. It contains an aromatic ring linked by an aliphatic chain to a heterocyclic ring containing a nitrogen atom instead of the amino group encountered in the structure of amphetamines. Dextromoramide contains two aromatic rings and two heterocycles linked to a side chain. This compound is an analgesic opioid used only in cases of terminal cancers [13]. At the same time, a false positive T (codeine - PFPA - N30) is also recorded in the case of unprocessed spectra.

The dendrogram associated to the spectra preprocessed with the w selective amplifier indicates that the modeling and the discrimination power have been improved for the clusters associated with hallucinogenic amphetamines (class T) and with negative compounds (class N). As opposed to the case of unprocessed spectra, in this case there are no false T positives: codeine-PFPA (N30) is correctly recognized by the system as a negative. In conclusion, the increased distance between the T and N clusters (see Fig. 3) generated by the w selective amplifier do result in an improved sensitivity of class N and selectivity of class T. On the other hand, no improvement is recorded for stimulant amphetamines: the increased distance between the center of the clusters T and N is not large enough to avoid being counterbalanced by the increased dispersion of these two clusters. The four false M positives (cadaverine, prolintane, putrescine and dextromoramide) are grouped in this case in a compact mini-cluster located at a distance of only 3.811 units from the nearest group of stimulant amphetamines.

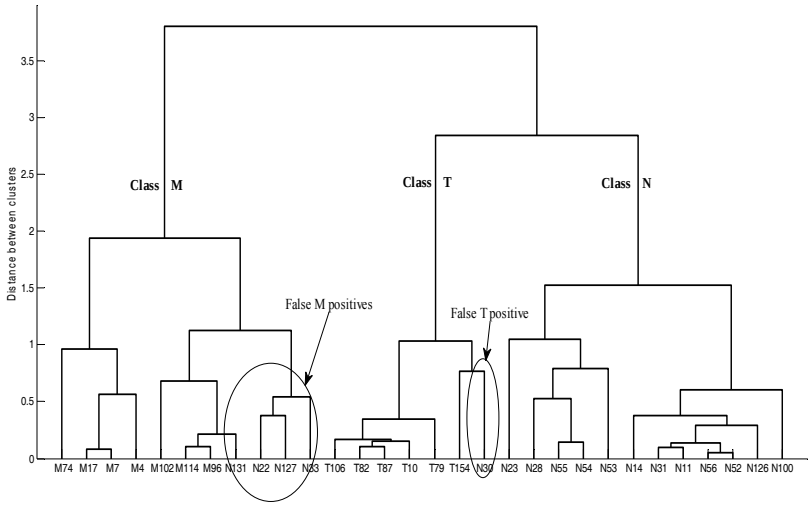


Fig. 2**.** Dendrogram built with the PC1 and PC2 scores associated to the spectra of stimulant amphetamines (class M), hallucinogenic amphetamines (class T) and negatives (class N) in the case of unprocessed spectra.

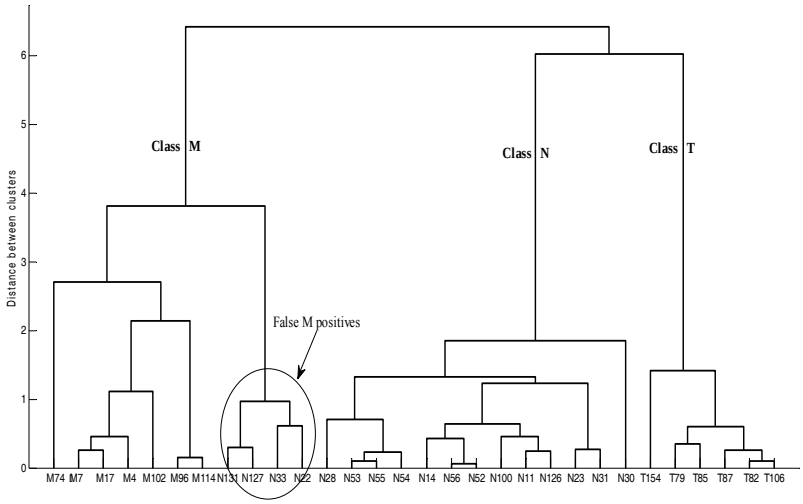


Fig. 3**.** Dendrogram built with the PC1 and PC2 scores associated to the spectra of stimulant amphetamines (class M), hallucinogenic amphetamines (class T) and negatives (class N) in the case of spectra preprocessed with the *w* selective amplifier.

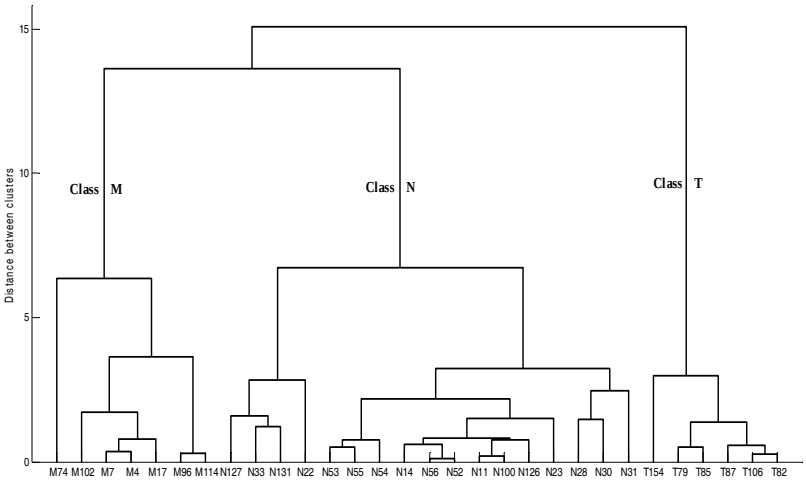


Fig. 4**.** Dendrogram built with the PC1 and PC2 scores associated to the spectra of stimulant amphetamines (class M), hallucinogenic amphetamines (class T) and negatives (class N) in the case of spectra preprocessed with the w^2selective amplifier.

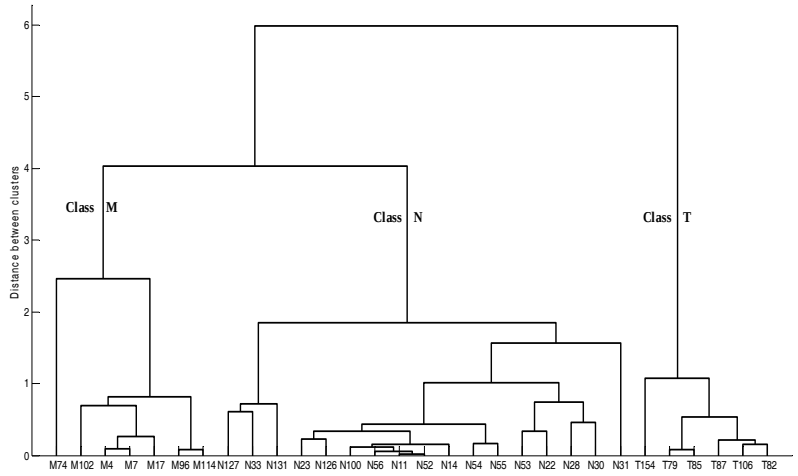


Fig. 5**.** Dendrogram built with the PC1 and PC2 scores associated to the spectra of stimulant amphetamines (class M), hallucinogenic amphetamines (class T) and negatives (class N) in the case of spectra preprocessed with the amplifying selector $(w\text{-}1)^2$.

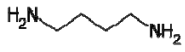
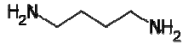
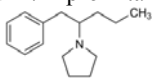
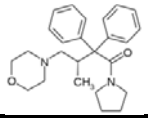
The most striking positive effect generated by the w^2 selective amplifier is the improvement of the selectivity of stimulant amphetamines (M) and of the

sensitivity in detecting non-amphetamines (see Fig. 4). Unlike previous cases, cadaverine, prolintane, putrescine and dextromoramide are correctly recognized by the detection system as negatives (N). The same correct classification rates are obtained in the case of spectra preprocessed with the $(w\text{-}1)^2$ amplifying selector (see Fig. 5), which also generates three compact homogenous clusters: no false positive or false negative are encountered.

Nevertheless, it is noticed that the visual inspection of the dendrograms and the assessment of the correct classification rates did not allow a full ranking of the methods in terms of discrimination efficiency. A more specific ranking method was needed in order to identify objectively which feature weight generates the best detection system, i.e. the w^2 selective amplifier or the $(w\text{-}1)^2$ amplifying selector.

Table 1

Performance evaluation for the agglomerative clustering detection systems

Method	**Cophenetic correlation coefficient**	**False M positives**	**False T positives**
Unprocessed spectra	0,8356	N131 – putrescine  N22 – cadaverine  N127 – prolintane  N33 – dextromoramide 	N30 - codeine - PFPA
Spectra preprocessed with the *w* selective amplifier	0,8531		no false T positive
Spectra preprocessed with the w^2 selective amplifier	0,9231	no false positive or false negative	
Spectra preprocessed with the $(w\text{-}1)^2$ amplifying selector	**0,9408**	**no false positive or false negative**	

Consequently, the performances in predicting the class identity of an unknown were assessed by using the cophenetic correlation coefficient, which is a measure of the rigor with which a dendrogram maintains the distances between the pairs of original non-modeled objects. The cophenetic coefficient determines the correlation between the initial Euclidean distances between the analyzed objects and the distances resulting from dendrogram configuration. It is generally

considered that the algorithms characterized by a cophenetic correlation coefficient larger than 0.8 are adequate for classification. The closer this coefficient is to 1, the more adequate the algorithm is [14].

The cophenetic correlation coefficients specific to the dendrograms presented in Fig. 2, Fig. 3, Fig. 4 and Fig. 5 are listed in Table 1. The coefficients characterising the dendrograms obtained for the case of the amplifying selector w^2 and for the selective amplifier $(w-1)^2$ are only slightly smaller than 1, which explains why no false positives or false negatives are encountered in these cases. However, the best results for discriminating the illicit amphetamines according to their biological activity are obtained by preprocessing their GC-FTIR spectra with the $(w-1)^2$ feature weight. The associated dendrogram is characterized by the largest cophenetic correlation coefficient, i.e. 0.9408.

4. Conclusion

In this paper we presented the evaluation of three signal processing methods performed for improving the efficiency of the detection of illicit amphetamines based on PCA. The GC-FTIR spectra of the targeted compounds were preprocessed by using two functions acting as selective amplifiers and one as an amplifying selector. A remarkable result obtained with both unprocessed and processed spectra is that the presented detection system discriminates amphetamines according to their biological activity (stimulant and hallucinogenic) although no information about the biological activity or toxicity of the compounds was included in the spectral database.

However, PCA - being an unsupervised pattern recognition technique, it does not define boundaries for the modeled clusters. PCA itself does not allow an objective assessment of the sensitivity and / or selectivity of the class identity assignment, i.e. of the number of false positives or negatives corresponding to each of the modeled class (cluster). Consequently, the effect of the preprocessing functions on the efficiency (correct classification rates) of the detection system was assessed by using dendrograms generated by agglomerative clustering.

The dendrograms indicated that the selected spectra preprocessing methods lead to a significant improvement of the discrimination efficiency. The positive effect of spectra preprocessing with the w selective amplifier is an improved sensitivity in discriminating the negatives (N) and selectivity in detecting the hallucinogenic amphetamines (T). The best correct classification rates are obtained when the GC-FTIR spectra are preprocessed with the w^2 selective amplifier, or with the $(w-1)^2$ amplifying selector. Both these functions are maximizing the sensitivity and the selectivity of the detection system for stimulant, as well as for hallucinogenic amphetamines.

A final ranking of the feature weights was obtained based on the cophenetic correlation coefficients specific to the dendrograms determined by agglomerative clustering. They indicated that the best clustering quality is obtained for the $(w\text{-}1)^2$ amplifying selector, which becomes the feature weight of choice for detecting the stimulant and the hallucinogenic amphetamines, and for distinguishing them from non-amphetamines.

R E F E R E N C E S

[1] *** European Monitoring Centre for Drugs and Drug Addiction, European drug report 2013, Lisbon, Portugal, www. emcdda.europa.eu

[2] *S. Gosav, M. Praisler, D.O. Dorohoi and G. Popa*, "Structure – Activity Correlations for Illicit Amphetamines using ANN and Constitutional Descriptors", in Talanta - The International Journal of Pure and Applied Chemistry, **vol.70**, 2006, pp. 922-928

[3] *B.G.M. Vandeginste*, Handbook of chemometrics and qualimetrics, Part B (vol. 20B), Elsevier, 1998

[4] *M. Praisler, I. Dirinck, J. Van Bocxlaer, A. De Leenheer and D.L. Massart*, "Computer-aided Screening for Hallucinogenic and Stimulant Amphetamines with Gas Chromatography - Fourier Transform Infrared Spectroscopy (GC - FTIR)", in Journal of Analytical Toxicology, **vol. 25**, no. 1, 2001, pp. 45-56

[5] *S.Gosav, R. Dinica and M. Praisler*, "Choosing Between GC-FTIR and GC-MS Spectra for an Efficient Intelligent Identification of Illicit Amphetamines", in Journal of Molecular Structure, **vol. 887**, no. 1-3, 2008, pp. 269-278

[6] *S. Ciochina and M. Praisler,* "Feasibility Study Regarding the Detection and Discrimination of Amphetamines based on Vibrational Spectra", in Proceedings of the International Scientific Workshop "European Union – Space of liberty, Security and Justice", Police Academy, Bucharest, 23 May 2013, Ed. Universitara, pp. 480-490

[7] *S. Ciochina and M. Praisler,* "Optimized Detection of Amphetamines based on Principal Component Analysis", accepted for publication in the Annals of Constanta Maritime University, 2013

[8] *** Statistics Toolbox 7 User's Guide, 2010

[9] *D. Mullner,* "Modern Hierarchical, Aglomerative Clustering Algoritms", in arXiv:1109.2378v1, Sept. 2011, pp. 1-29

[10] *S. Ciochina and M. Praisler,* "Cluster Analysis Applied for the Discrimination of Illicit Amphetamines", in Annals of "Dunarea de Jos" University of Galati, Fascicle II Mathematics, Physics, Theoretical Mechanics, **vol. 1**, no. 2, 2013, pp. 69-76

[11] *L.G. Apollonio, I.R. Whittall, D.J. Pianca, J.M. Kyd and W.A. Maher,* "Matrix Effect and Cross-reactivity of Selected Amphetamine-type Substances, Designer Analogues and Putrefactive Amines using the Bio-Quant direct ELISA Presumptive Assays for Amphetamine and Methamphetamine", in Journal of Analytical Toxicology, **vol. 31**, no. 4, 2007, pp. 208-213

[12] *A.N. Nicholson, B.M. Stone and M.M. Jones,* "Wakefullness and Reduced Rapid Eye Movement Sleep: Studies with Prolintane and Pemoline", in British Journal of Clinical Pharmacology, **vol. 10**, no. 5, 1980, pp. 465–472

[13] *A.J. Grace,* "Analgesia in Terminal Malignant Disease", in British Medical Journal, **vol. 4**, no. 2, 1979, pp. 6185-6333

[14] *L. Kaufman and P.J. Rousseeuw*, Finding groups in data. An introduction to cluster analysis, Willey, 2005.