Journal of Chromatography, 485 (1989) 557–567
Elsevier Science Publishers B.V., Amsterdam — Printed in The Netherlands

CHROM. 21 969

RECENT ADVANCES IN FUZZY PEAK TRACKING IN HIGH-PERFORMANCE LIQUID CHROMATOGRAPHY

E. P. LANKMAYR*, W. WEGSCHEIDER and J. DANIEL-IVAD
Institute for Analytical Chemistry, Micro- and Radiochemistry, Graz University of Technology, Technikerstrasse 4, A-8010 Graz (Austria)
I. KOLOSSVÁRY and G. CSONKA
Department of Chemistry, Technical University, Budapest (Hungary)
and
M. OTTO
Department of Chemistry, Bergakademie Freiberg, Akademiestrasse 6, 9200-Freiberg (G.D.R.)

SUMMARY

In general, automated optimization procedures for chromatographic separations necessitate a recognition of the eluted signals. An approach based on comparison of peak areas and of the elution order of the peaks has been designed to match the requirement of widely varying chromatographic conditions. This method relies on fuzzy theory and can therefore be applied to uncertain data as they stem from the imprecision of peak areas, the change in the elution order and the uncertainty of overlapped peak areas. The handling of peak overlap has been greatly improved and is successfully demonstrated for the recognition of chromatograms with several overlapped peaks and changing elution patterns.

INTRODUCTION

Recognition of chromatographic peaks based on fuzzy theory was reported recently[1]. Using fuzzy theory, uncertainties in the data can be handled in a mathematically well defined manner[2,3], which is very useful in chromatographic peak tracking where the uncertainties in the peak areas of the same component in different chromatographic runs can be taken into account. Another great advantage of fuzzy theory over traditional techniques is that it can handle cardinal and ordinal variables in tandem. This means that in chromatographic peak tracking not only detector response information, *e.g.*, peak areas and peak heights, but also the peak elution order in different chromatographic runs can be compared simultaneously.

A peak-tracking algorithm based on fuzzy theory and used first for component identification in the UV spectral range[4] was presented in a previous paper[1]. This routine is used with a program for unattended optimization of high-performance liquid chromatographic (HPLC) separations[5]. The optimization program is based on mathematical modelling of the retention behaviour of solutes as a function of the

0021-9673/89/$03.50

composition of the mobile phase and requires peak recognition. The peak identification is carried out by comparing the peak areas and elution order of components of the same solute between the trial and a chosen reference run. The difference is computed as a fuzzy comparison, *i.e.*, the variables (peak area, elution order) are assigned a so-called membership function that characterizes the variability of peak areas and of the solute elution order. Comparison of peak areas in the trial run with those of the reference run is carried out by fuzzy subtraction of both area patterns, yielding a membership function around the peak differences. From this membership function, the so-called membership value, m_s (degree of belief), can be computed and ranges from zero to one. The closer, the membership value to one, the better is the coincidence between the trial and reference peak areas. A detailed explanation of this fuzzy algorithm was given in ref. 1.

Although the change in the elution order of solutes in different mobile phases often causes difficulties in chromatographic optimization, it is very useful to obtain a sufficient chromatographic separation selectivity. The change in peak elution order is normally limited with respect to a definitie solute. Keeping the stationary phase constant, it is only rarely observed that the first-eluting solute will elute at the end of the chromatogram of another run with a different mobile phase composition. This experience is expressed in such a way that the elution order in the trial run is compared with that in the reference run. Thus, in a similar manner to that possible with peak areas, a fuzzy comparison of peak elution order provides additional information from the chromatograms. The appropriate fuzzy sets are described by a membership function of the form

$$m(t) = [1 - c\,|\,t - t_i\,|]^+ \tag{1}$$

where t is the elution rank of the peak tested for membership of the ith peak in the reference run, c is a constant normalizing $m(t)$ to the interval [0–1] and + denotes truncation to 0 if negative membership values occur.

In the trial run, the membership value $m_e(t)$ is assigned by intersecting the membership function for the elution order of peaks in the reference run $m_b(t)$ with that of the trial peaks $m_a(t)$, where $m_a(t)$ is taken as crisp, *i.e.*, with $m_a(t) = 1$ for the elution rank t and zero for all other ranks. Intersection means simply the selection of the best coincidence

$$m_e(t) = \min [m_b(t), m_a(t)] \tag{2}$$

Results from comparison of the peak areas and the elution order of peaks can be aggregated by taking some mean m_m of m_s and m_e. In this work, the arithmetic mean was chosen.

Handling overlapping peaks gives rise to new problems. Summing individual reference peaks for representing unresolved trial peaks was described first by Issaq and McNitt[6]. This idea was transferred and "fuzzyfied". In contrast to the handling of peak areas, the comparison of the peak elution order is much more critical with overlapped peaks. In this paper, an improvement of the fuzzy algorithm outlined briefly in eqns. 1 and 2 and a logic designed for its application in automated HPLC optimization are presented.

EXPERIMENTAL

The chromatographic experiments were performed with Waters Assoc. M6000 pumps, a Waters WISP 710A autoinjector and two detectors, a Waters Assoc. M440 and a Schoeffel SF 770 UV detector, connected in series. For data acquisition and system automation, a Waters Assoc. 840 chromatographic data station with a Digital Equipment 350 personal computer was used.

A Spherisorb ODS-2 (5 μm) column (150 $\times$ 4.6 mm I.D.) was used throughout. The flow-rate was kept constant at 1 ml/min. Solvents and chemicals used for the experiments were of analytical-reagent grade, purchased from various sources. The sample mixture contained phenol, benzaldehyde, *m*-dinitrobenzene, 2,4-dimethylphenol, *p*-iodophenol and phenetole.

The peak-tracking software is written in compiled Microsoft Quick Basic V 4.0 and was used with an MS-DOS computer.

THEORY

The problem with overlaps arises from the fact hat in this instance the number of peaks in the reference run is not identical wih that in the trial run. For instance, if there are ten reference peaks, but only seven trial peaks because of overlap(s), then the last trial peak has a rank of seven, which is, however, misleading. It is much more probable that this last trial peak coincides with the last reference peak which is described by the membership function $m_b(t)$ with a maximum value at rank ten (see eqn. 1). Thus, the intersection in eqn. 2 probably gives a smaller membership value than it should. This shift in the trial peak ranks propagates along the whole chromatogram and the ranks lose their meaning.

This drawback of the peak-tracking algorithm can be partly overcome by introducing formal peak ranks. The formal peak ranks can also be non-integral numbers. The improved algorithm is formulated as follows:

(1) The chromatogram showing the maximum number of signals is always chosen as a reference.

(2) Peak tracking is executed in several stages. In the so-called zeroth stage the membership values for all of the trial peaks to each reference peak are computed, but taking only the peak areas into account, *i.e.*, only m_s is computed. If at least one of the membership values of a trial peak is greater than a predefined threshold value, then this trial peak will be considered as a single peak. On the other hand, all the trial peaks that have no membership value greater than the threshold value will be considered in the later stages as potentially overlapped peaks.

(3) After dividing the trial peaks into a set of single and a set of potentially overlapped peaks, formal peak ranks are assigned (#Rpks $\equiv$ number of reference peaks, #Tpks $\equiv$ number of trial peaks, and #Chs $\equiv$ number of detection channels):

First trial peak $\equiv$ 1;

Last trial peak $\equiv$ #Rpks.

For the other trial peaks, the following formula is used:

DIF = number of the potentially ovelapped trial peaks;

If DIF $\neq$ 0 then SKIP = 1 + (#Rpks $-$ #Tpks)/DIF;

If the *i*th trial peak is a single peak, then the formal peak rank of the $(i + 1)$th trial

peak is greater by one, otherwise greater by SKIP than that of the *i*th trial peak.

(4) First the trial peaks are compared with the single reference peaks, then with the linear combinations of two, and then three reference peaks, etc. Here the peak areas as well as the elution order is taken into account by the following weighting equation:

$$m_m = \frac{\#\mathrm{Chs} \cdot m_s + \dfrac{m_e}{1 + \#\mathrm{Rpks} - \#\mathrm{Tpks}}}{\#\mathrm{Chs} + \dfrac{1}{1 + \#\mathrm{Rpks} - \#\mathrm{Tpks}}} \tag{3}$$

Eqn. 3 gives the arithmetic mean if no overlap occurs, *i.e.*, $\#\mathrm{Rpks} = \#\mathrm{Tpks}$, and if single-channel monitoring is used.

(a) In the first stage, single reference peaks are assigned single trial peaks and then the assigned trial peaks and their reference peak are excluded from further investigation.

(b) With the linear combinations of the remaining reference peaks, the peak areas are "fuzzy" added (see ref. 1), and the peak elution order is also "fuzzy" handled as follows: the membership value $m_e(t)$ is computed by eqn. 2, where the membership function $m_b(t)$ is the intersection of the individual $m_{bi}(t)$s, *i.e.*, $m_b(t) = \min [m_{bi}(t), m_{bj}(t), \ldots]$, where $i, j, \ldots$, are the ranks of the reference peaks involved in the actual linear combination.

(5) Each potentially overlapped trial peak is compared with each linear combination of the reference peaks not recognized as a single peak and the linear combination with the highest membership value m_m is assigned.

RESULTS AND DISCUSSION

The improved peak-tracking algorithm was tested with the HPLC separation of aromatic sample species. For the separation a reversed-phase system was chosen. The mobile phases consisted of various mixtures of methanol, acetonitrile (ACN) and tetrahydrofuran (THF) with water as indicated in Fig. 1. Thus, we expected to obtain significant alterations in chromatographic selectivity according to the concept of the solvent selectivity triangle[7], including peak overlaps and crossovers. Additionally, the changed solvation conditions of the analytes throughout the experiments were expected to contribute to the imprecision of measured peak areas. According to the aromatic nature of the sample solutes, detection wavelengths were selected at 254 and 280 nm.

The retention data and peak areas determined at ten different mobile phase compositions in Fig. 1 are reported in Table I. The data are listed according to the elution order of the signals and represent a mere integration report without signal identification. The solvent peaks originating from the sample solvent were recorded by extra injections and could, therefore, be subtracted automatically from the chromatograms. The retention data and the peak areas were used to compute the membership functions with the peak tracking program. The results are presented in Table II.

The theoretically expected true assignment is compared with the calculated assignment and the associated membership value for three different cases: first, the

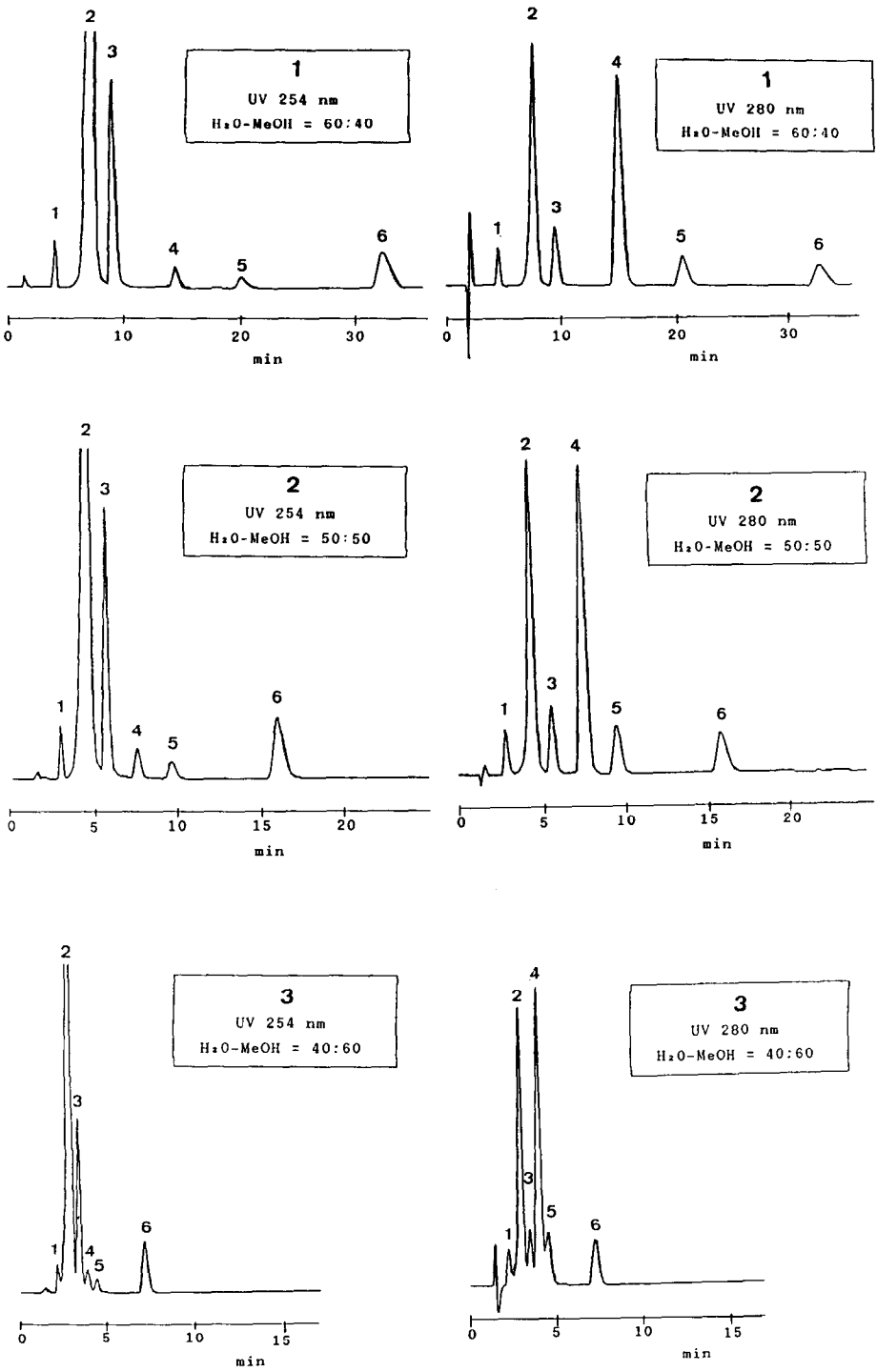

Fig. 1.

(Continued on p. 562)

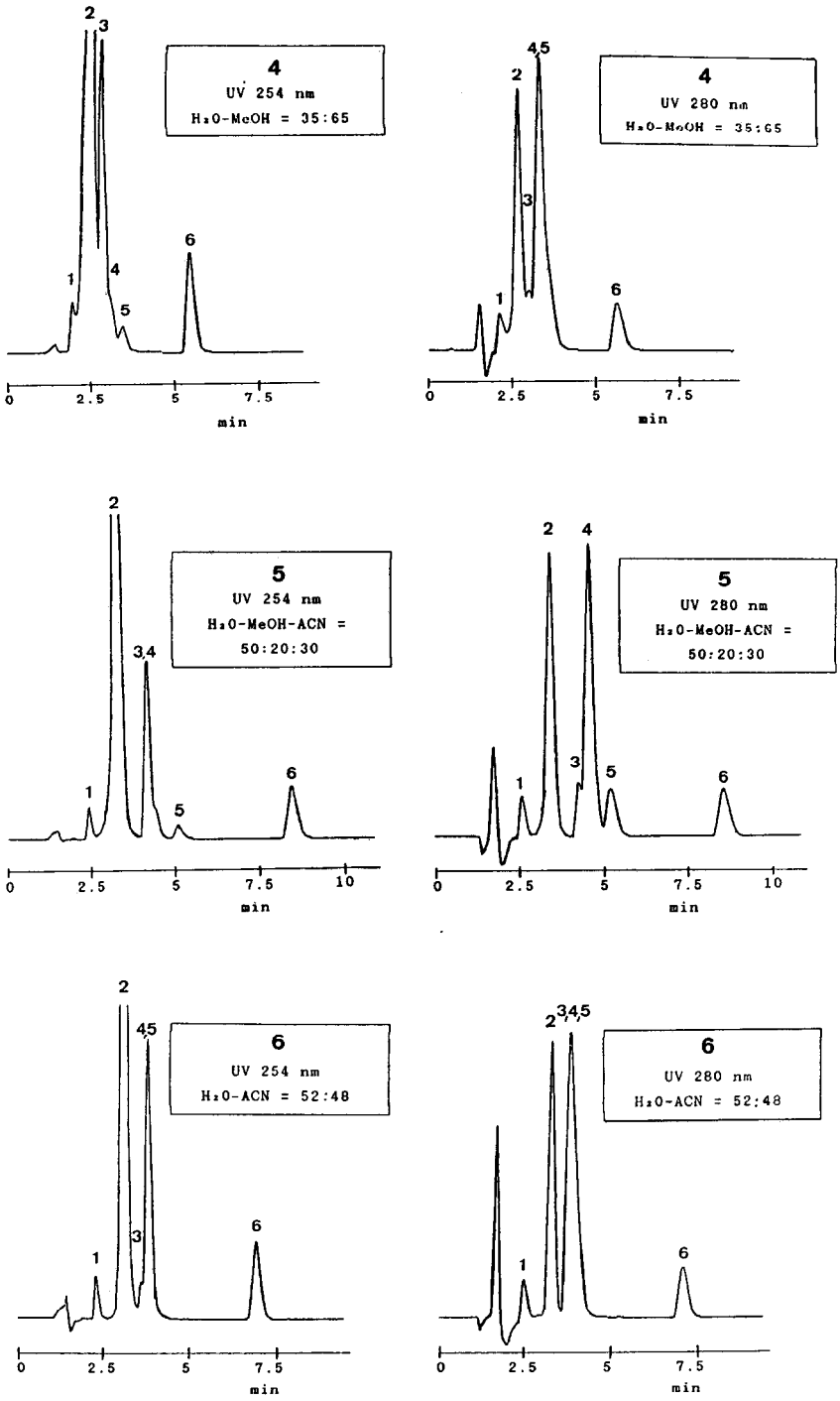
4
UV 254 nm
H₂O-MeOH = 35:65
4
UV 280 nm
H₂O-MeOH = 35:65
5
UV 254 nm
H₂O-MeOH-ACN = 50:20:30
5
UV 280 nm
H₂O-MeOH-ACN = 50:20:30
6
UV 254 nm
H₂O-ACN = 52:48
6
UV 280 nm
H₂O-ACN = 52:48
min

Fig. 1.

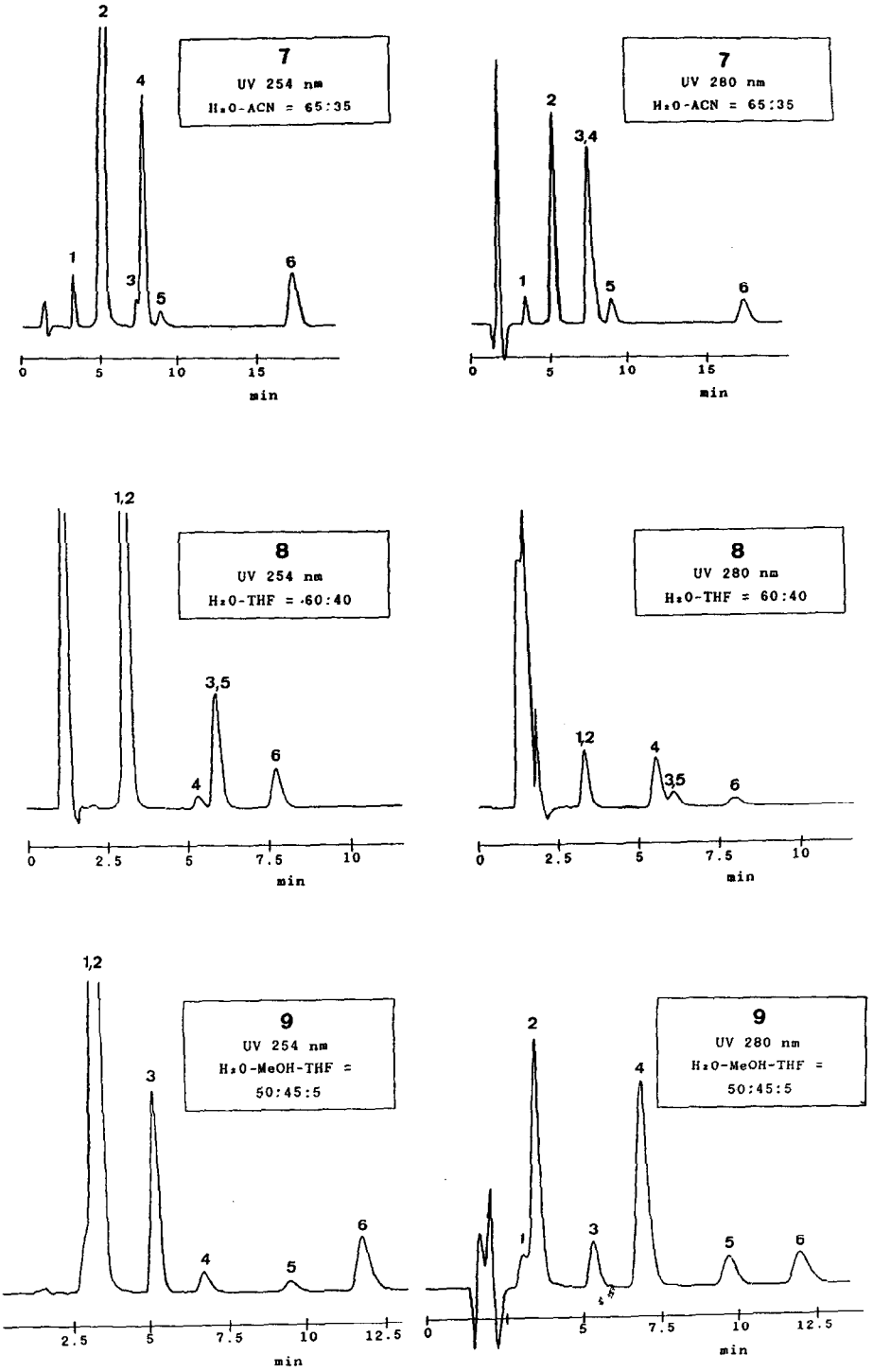

Fig. 1.

(Continued on p. 564

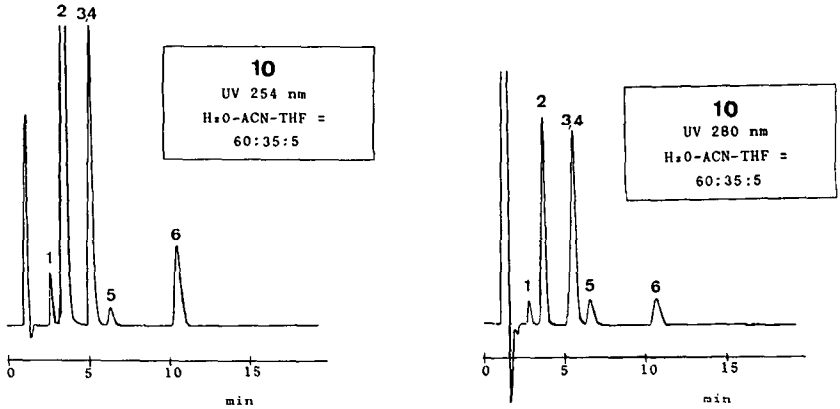

Fig. 1. Chromatographic separation of aromatic compounds with two-channel detection. Peaks: 1 = phenol; 2 = benzaldehyde; 3 = *m*-dinitrobenzene; 4 = 2,4-dimethylphenol; 5 = *p*-iodophenol; 6 = phenetole. MeOH = Methanol; **1–10** correspond to runs I–X.

data obtained at 254 nm alone; second, only the data recorded at 280 nm; and third, the combined data from the two detection channels were used for the peak-tracking procedure. A parabolic membership function was chosen with symmetrical spreads of 10% each and a threshold value of 0.75. The highest membership values for the individual signals and their linear combinations are reported and wrong assignments are indicated by numbers in parentheses. As chromatographic run 1 resulted in a chromatogram with six solute signals, a number which was not exceeded by subsequent separations, it was chosen arbitrarily as the reference run. Peak assignment is expectedly less precise with the data from single-channel detection than from two-channel detection. Incorrect assignments can usually be recognized by either multiple assignment of one compound in a chromatogram or by a missing compound in the list. With the data from the UV channel at 254 nm, *e.g.*, compound 4 is missing in runs 5, 7 and 10, whereas compound 2 is assigned twice in run 8 with the data recorded at 280 nm. The combination of the information from the two detection channels improves the assignment considerably. Only the overlap of components 1 and 2 is not recognized instantly. The area of component 1 corresponds roughly with the deviation of the area of component 2. Consequently, the highest value of the membership function is attributed to component 2 alone. The next best numerical value, however, is computed for the combination of components 1 and 2 and, owing to the absence of component 1 in the assignment list of runs 8 and 9, the overlap can be detected correctly.

CONCLUSION

An improved logic for handling peak overlap and signal crossover in fuzzy peak tracking is described. Implementation of this routine into an automated method development procedure for HPLC separations should eventually be accomplished. The peak-tracking procedure was tested with a reversed-phase separation system encountering peak overlap and changing elution order. It has been demonstrated that

TABLE I

RETENTION TIMES (t_R) AND PEAK AREAS (A) DETERMINED FROM THE CHROMATOGRAMS SHOWN IN FIG. 1

Run	*Parameter*	*Signal*					
		1	*2*	*3*	*4*	*5*	*6*
I	t_R (min)	4.1	7.2	9.3	15.0	20.8	33.0
	$A_{254\ nm}$	1130	62 283	9222	1115	768	4022
	$A_{280\ nm}$	690	8134	1692	8439	1470	1599
II	t_R (min)	3.0	4.5	5.7	7.5	9.5	15.9
	$A_{254\ nm}$	1143	58 665	9544	1187	730	4071
	$A_{280\ nm}$	811	8413	1761	8912	1588	1784
III	t_R (min)	2.2	2.9	3.5	4.0	4.6	7.5
	$A_{254\ nm}$	1213	45 850	9231	1343	846	3812
	$A_{280\ nm}$	928	7757	1549	8801	1856	1783
IV	t_R (min)	2.0	2.6	3.6	5.7		
	$A_{254\ nm}$	1326	41 370	11 850	3912		
	$A_{280\ nm}$	931	7357	12 250	1754		
V	t_R (min)	2.4	3.3	4.2	5.2	8.6	
	$A_{254\ nm}$	1101	41 670	9882	827	3845	
	$A_{280\ nm}$	917	7724	10 200	1586	1923	
VI	t_R (min)	2.4	3.2	4.0	7.2		
	$A_{254\ nm}$	1005	35 850	10 910	3866		
	$A_{280\ nm}$	943	7784	12 250	2124		
VII	t_R (min)	3.2	5.1	7.8	8.8	17.2	
	$A_{254\ nm}$	1092	46 660	9596	740	3893	
	$A_{280\ nm}$	842	7974	10 150	1490	2004	
VIII	t_R (min)	3.2	5.5	6.1	8.0		
	$A_{254\ nm}$	39 370	974	9621	3756		
	$A_{280\ nm}$	8563	9508	3548	2058		
IX	t_R (min)	3.5	5.3	6.9	9.8	12.1	
	$A_{254\ nm}$	49 530	8826	1162	736	3947	
	$A_{280\ nm}$	8792	1616	8814	1494	1809	
X	t_R (min)	2.9	3.8	5.4	6.7	11.0	
	$A_{254\ nm}$	1032	40 130	9731	598	3777	
	$A_{280\ nm}$	630	7614	10 140	1323	1784	

TABLE II

PEAK ASSIGNMENT WITH SINGLE-CHANNEL DETECTION AT 254 AND 280 nm AND WITH TWO-CHANNEL DETECTION

Reference = run 1 (for conditions, see Fig. 1).

Run No.	*Signal*	UV_{254} *Assignment*			UV_{280} *Assignment*		$UV_{254} + UV_{280}$ *Assignment*	
		Theor.	*Comput.*	(m_m)	*Comput.*	(m_m)	*Comput.*	(m_m)
II	1	1	1	(0.926)	1	(0.862)	1	(0.873)
	2	2	2	(0.916)	2	(0.923)	2	(0.888)
	3	3	3	(0.923)	3	(0.922)	3	(0.897)
	4	4	4	(0.916)	4	(0.918)	4	(0.890)
	5	5	5	(0.919)	5	(0.911)	5	(0.882)
	6	6	6	(0.926)	6	(0.894)	6	(0.894)
III	1	1	1	(0.913)	1	(0.777)	1	(0.817)
	2	2	2	(0.771)	2	(0.920)	2	(0.698)
	3	3	3	(0.927)	3	(0.906)	3	(0.901)
	4	4	4	(0.845)	4	(0.922)	4	(0.893)
	5	5	5	(0.901)	5	(0.814)	5	(0.772)
	6	6	6	(0.919)	6	(0.894)	6	(0.883)
IV	1	1	1	(0.795)	1	(0.662)	1	(0.715)
	2	2	2	(0.525)	2	(0.765)	2	(0.494)
	3	3 + 4 + 5	3 + 4 + 5	(0.761)	3 + 4 + 5	(0.794)	3 + 4 + 5	(0.793)
	4	6	6	(0.885)	6	(0.855)	6	(0.860)
V	1	1	1	(0.898)	1	(0.711)	1	(0.782)
	2	2	2	(0.632)	2	(0.850)	2	(0.547)
	3	3 + 4	[3]	(0.802)	3 + 4	(0.860)	3 + 4	(0.843)
	4	5	5	(0.883)	5	(0.882)	5	(0.859)
	5	6	6	(0.894)	6	(0.794)	6	(0.842)
VI	1	1	1	(0.834)	1	(0.652)	1	(0.686)
	2	2	2	(0.498)	2	(0.826)	2	(0.402)
	3	3 + 4 + 5	3 + 4 + 5	(0.804)	3 + 4 + 5	(0.766)	3 + 4 + 5	(0.799)
	4	6	6	(0.883)	6	(0.675)	6	(0.803)
VII	1	1	1	(0.897)	1	(0.781)	1	(0.819)
	2	2	2	(0.707)	2	(0.871)	2	(0.653)
	3	3 + 4	[3]	(0.813)	3 + 4	(0.847)	3 + 4	(0.846)
	4	5	5	(0.897)	5	(0.873)	5	(0.878)
	5	6	6	(0.897)	6	(0.758)	6	(0.835)
VIII	1	1 + 2	[2]	(0.466)	[2]	(0.795)	[2]	(0.444)
	2	4	4	(0.817)	[2]	(0.798)	4	(0.808)
	3	3 + 5	3 + 5	(0.717)	[1 + 3 + 5]	(0.697)	3 + 5	(0.762)
	4	6	6	(0.871)	6	(0.703)	6	(0.796)
IX	1	1 + 2	[2]	(0.672)	[2]	(0.798)	[2]	(0.658)
	2	3	3	(0.895)	2	(0.811)	3	(0.874)
	3	4	4	(0.896)	3	(0.812)	4	(0.874)
	4	5	5	(0.895)	4	(0.817)	5	(0.877)
	5	6	6	(0.900)	5	(0.848)	6	(0.865)
X	1	1	1	(0.873)	1	(0.873)	1	(0.847)
	2	2	2	(0.610)	2	(0.886)	2	(0.538)
	3	3 + 4	[3]	(0.849)	3 + 4	(0.819)	3 + 4	(0.849)
	4	5	5	(0.696)	5	(0.862)	5	(0.802)
	5	6	6	(0.887)	6	(0.859)	6	(0.854)

a correct recognition of signals can be obtained even in difficult cases by increasing the information from the chromatographic detector(s).

REFERENCES

1 M. Otto, W. Wegscheider and E. P. Lankmayr, *Anal. Chem.*, 60 (1988) 517.
2 H. Brandemer and M. Otto, *Mikrochimica Acta*, 2 (1986) 93.
3 M. Otto, *Chemometr. Intell. Lab. Systems*, 4 (1988) 101.
4 M. Otto and H. Brandemer, *Anal. Chim. Acta*, (1986) 191.
5 E. P. Lankmayr, W. Wegscheider and K. W. Budna, *J. Liq. Chromatogr.*, 12 (1989) 35.
6 H. J. Issaq and K. L. McNitt, *J. Liq. Chromatogr.*, 5 (1982) 1771.
7 L. R. Snyder, *J. Chromatogr. Sci.*, 16 (1978) 223.