CHAPTER 4
RESULT AND DISCUSSION

4.1   Simulation Section

4.1.1   Ability of Factor Analysis in Multicomponent Analysis

Five simulated spectra of pure components were generated by using the Gaussian function as same as the distribution data in real infrared spectra. The Gaussian peaks with various peak widths and intensities were use to demonstrated the characteristic of each component.   The five simulated spectra were shown in Figure 4.1.
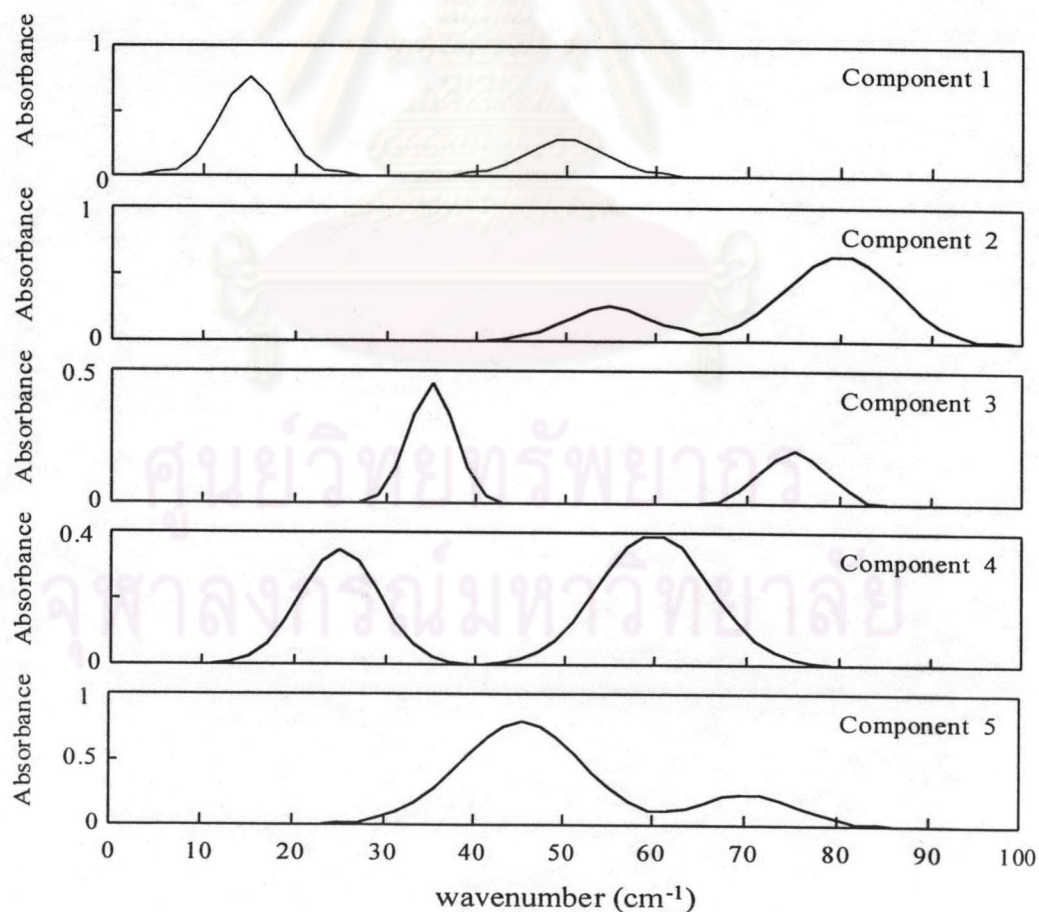

Figure 4.1   Simulated spectra of five pure components.

First the three-component system of components 1, 2, and 3 was studied. The simulated spectra of the mixture at twenty different compositions were generated and shown in Figure 4.2. By using the rule of five of the number of calibration sample in calibration set, fifteen samples were used as a calibration sample. Five testing samples were used for the test. Table 4.1 shows the results obtained by factor analysis compared with the actual data.
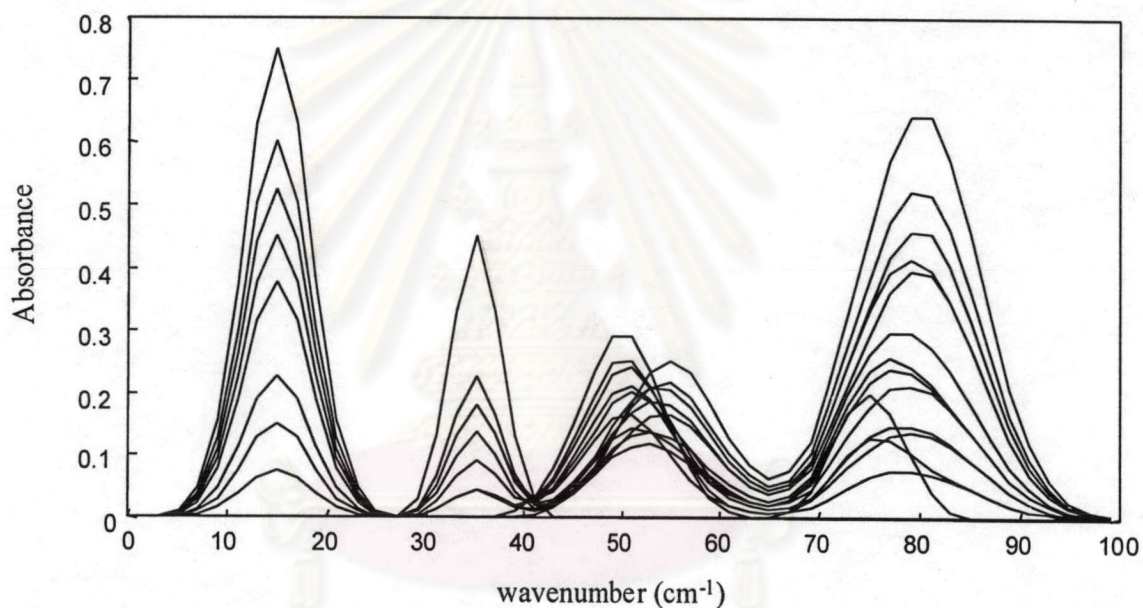

Figure 4.2   Simulated spectra of three-component system of
components 1, 2, and 3 at twenty compositions.

Table 4.1   Comparison between true value of volume fraction and results obtained
by factor analysis of three-component system.

Mixture	Component	Volume fraction	Result from factor analysis	Aactual value – Calculated value
1	1	0.5000	0.4971	0.0029
	2	0.4000	0.4035	-0.0035
	3	0.1000	0.0994	0.0006
2	1	0.1000	0.1007	-0.0007
	2	0.1000	0.0992	0.0008
	3	0.8000	0.8007	-0.0007
3	1	0.4000	0.3993	0.0007
	2	0.1000	0.1007	-0.0007
	3	0.5000	0.5000	0.0000
4	1	0.2000	0.1989	0.0011
	2	0.3000	0.3025	-0.0025
	3	0.5000	0.4986	0.0014
5	1	0.5000	0.4978	0.0022
	2	0.3000	0.3029	-0.0029
	3	0.2000	0.1993	0.0007

The systems of four-component and five-component were studied by the
same procedure as that in the three-component system. Tables 4.2 and 4.3 show the
results obtained by factor analysis of four- and five-component system, respectively.

Table 4.2   Comparison between true value of volume fraction and results obtained
by factor analysis of four-component system.

Mixture	Component	Volume fraction	Result from factor analysis	Aactual value – Calculated value
1	1	0.2000	0.2004	-0.0004
	2	0.5000	0.5003	-0.0003
	2	0.2000	0.2001	-0.0001
	4	0.1000	0.0993	0.0007
2	1	0.1000	0.1024	-0.0024
	2	0.1000	0.1004	-0.0004
	3	0.3000	0.3003	-0.0003
	4	0.5000	0.4969	0.0031
3	1	0.4000	0.4009	-0.0009
	2	0.1000	0.1003	-0.0003
	3	0.2000	0.2002	-0.0002
	4	0.3000	0.2981	0.0019
4	1	0.2000	0.2014	-0.0014
	2	0.4000	0.4002	-0.0002
	3	0.1000	0.1003	-0.0003
	4	0.3000	0.2981	0.0019
5	1	0.1000	0.1004	-0.0004
	2	0.7000	0.7002	-0.0002
	3	0.1000	0.1001	-0.0001
	4	0.1000	0.0993	0.0007

Table 4.3   Comparison between true value of volume fraction and results obtained
by factor analysis of five-component system.

Mixture	Component	Volume fraction	Result from factor analysis	Aactual value – Calculated value
1	1	0.1000	0.0931	0.0069
	2	0.4000	0.3956	0.0044
	3	0.0000	0.0000	0.0000
	4	0.3000	0.3028	-0.0028
	5	0.2000	0.2086	-0.0086
2	1	0.1000	0.0945	0.0055
	2	0.1000	0.1065	-0.0065
	3	0.5000	0.4915	0.0085
	4	0.2000	0.2010	-0.0010
	5	0.1000	0.1065	-0.0065
3	1	0.2000	0.1903	0.0097
	2	0.1000	0.1038	-0.0038
	3	0.1000	0.0865	0.0135
	4	0.5000	0.5050	-0.0050
	5	0.1000	0.1144	-0.0144
4	1	0.1000	0.0948	0.0052
	2	0.2000	0.2055	-0.0055
	3	0.2000	0.1893	0.0107
	4	0.1000	0.1087	-0.0087
	5	0.4000	0.4017	-0.0017
5	1	0.2000	0.1915	0.0085
	2	0.2000	0.2047	-0.0047
	3	0.2000	0.1869	0.0131
	4	0.2000	0.2042	-0.0042
	5	0.2000	0.2022	-0.0022

The quality of the prediction is investigated by the analysis of the following predicted residual error sum of square (PRESS) and the variance of prediction (s^2) (Appendix C). Table 4.4 shows PRESS and s^2 of the prediction when factor analysis was applied to the mixture of three-, four-, and five-component system.

Table 4.4   PRESS and s^2 of factor analysis in multicomponent analysis.

System	PRESS	s^2
Three-component system (component 1, 2, and 3)	4.7×10^{-7}	3.3×10^{-7}
Four-component system (component 1, 2, 3, and 4)	4.0×10^{-4}	6.9×10^{-2}
Five-component system (component 1, 2, 3, 4, and 5)	0.2045	0.0081

From Table 4.4, as the number of component in the system increases, the PRESS and s^2 continue to rise slightly. This can be explained that this technique can give the high quality of prediction, but the accuracy will be decreased when the number of component in the system increases.

4.1.2 Effect of Random Error or Noise in Simulated Spectra

In actual measurement there is random error or noise in the data. The random noise is caused from the measurement of the spectra and the determination of the concentration of component. The analyses were performed on the simulated spectra of the mixture with added noise in order to demonstrate the effect of the random noise in infrared spectra on the accuracy of factor analysis.

The random noises were added to the simulated spectra of the mixture of components 1, 2 and 3 at volume fraction 0.5 : 0.1 : 0.4. The simulated spectra without added noise and the simulated spectra with varying quantities of added noise were shown in Figure 4.3 and the results obtained by factor analysis were shown in Table 4.5.
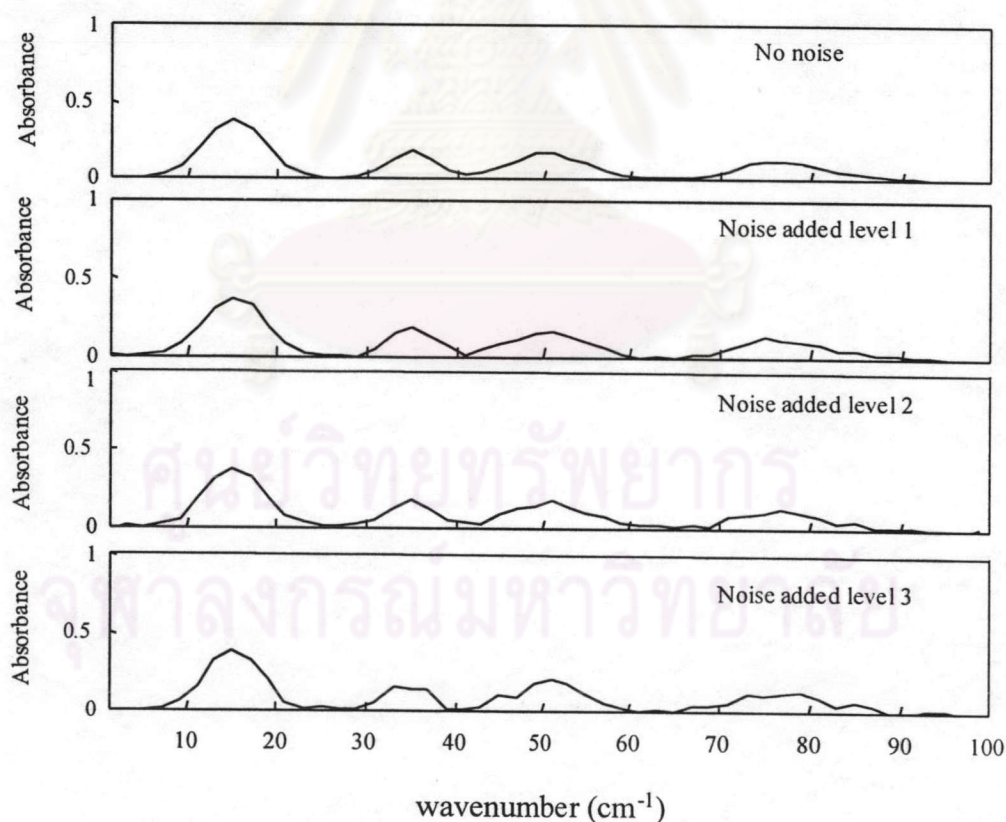

Figure 4.3   Simulated spectra of 0.5 : 0.1 : 0.4 mixture of
components 1, 2, and 3 with added noise.

Table 4.5 Relative error of the results from factor analysis when applied to simulated spectrum of 0.5 : 0.1 : 0.4 mixture of components 1, 2, and 3 with added noise.

Noise Level	Component 1 0.5000	Component 2 0.1000	Component 3 0.4000	Average % relative error
Noise Free	0.4971	0.0994	0.3996	0.40
1	0.4937	0.0984	0.4035	1.22
2	0.4900	0.0946	0.4154	3.75
3	0.4885	0.0894	0.4221	6.06

The results from Table 4.5 show that the average relative error is only 0.41 % when noise-free data were employed. When the noise was added to the original data at level 1 and 2, the average relative error increases to 1.22 % and 3.75 %, respectively. The result show that the quantitative result obtained by factor analysis with high noise does not highly differs from the original. The average relative error from the data with added noise level 3 is 6.07 %, this shows the limitation of factor analysis with the high level of noise. But this condition rarely occurs in the real measurement.

4.1.3 Effect of Peak Overlap in Simulated Spectra

In a real system, multicomponent infrared spectra always have a peak overlap that makes the conventional analytical method difficult. Can factor analysis still be used for this case? This part had studied the effect of peak overlap in infrared spectra with the ability of factor analysis.

The components 6, 7, and 8 were generated, as the structures shown in Figure 4.4. They do not differ so much in peak width and peak position since they will be used to demonstrate the case where peak overlap presents in the mixture spectra. Figure 4.5 illustrates the example of the mixture at volume fraction 0.2 : 0.2 : 0.6.
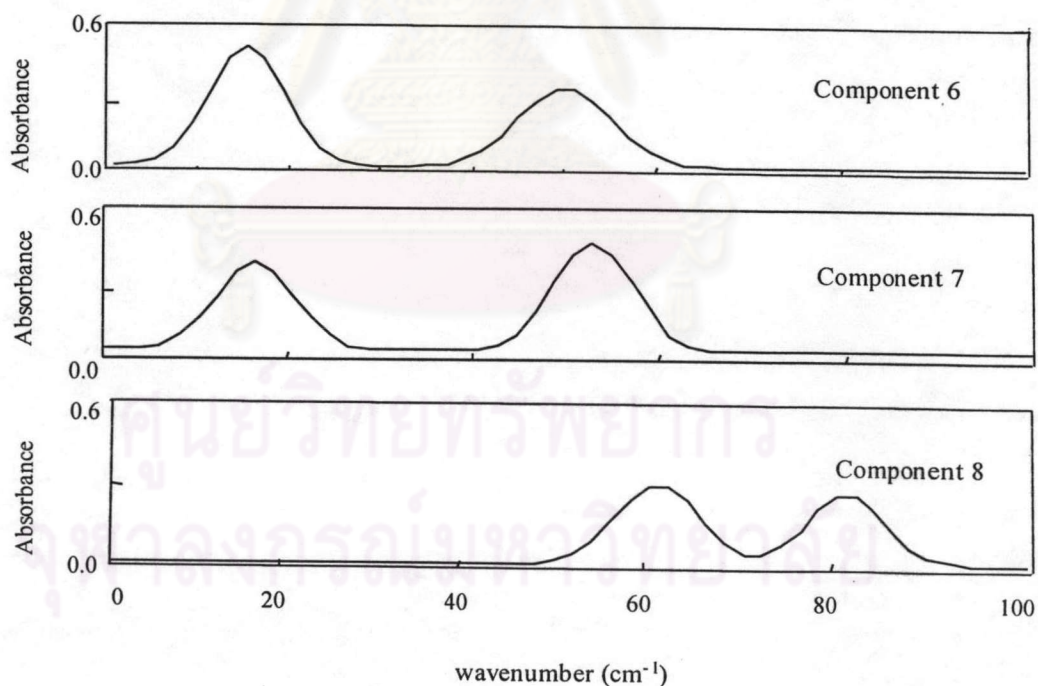

Figure 4.4   Simulated spectra of component 6, 7, and 8.
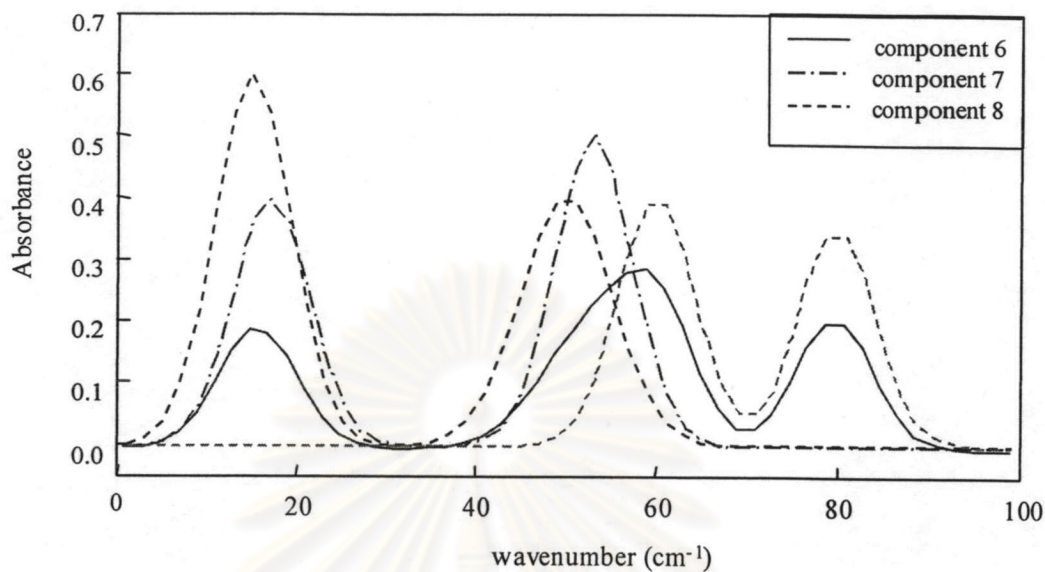
Figure 4.5   Simulated spectra of the mixture of components 6, 7,
and 8 at volume fraction 0.2 : 0.2 : 0.6.

From Figure 4.5, there are the peak overlaps between components 6 and 7 at 15-17 cm^{-1} and between components 6, 7, and 8 at 50-60 cm^{-1}. To study the ability of factor analysis when handling this condition, five synthetic spectra of the components 6, 7 and 8 mixtures at five volume fractions that presented the peak overlap were tested. After analyzing the data by factor analysis program, the results were reported in Table 4.6.

Table 4.6   Volume fractions of pure component obtained from factor analysis with spectral overlap.

No	Testing Sample Volume fraction Comp.1: Comp.2 : Comp.3	Results obtained by FA Component 1	Component 2	Component 3	Average % Relative Error
1	0.8 : 0.1 : 0.1	0.7990	0.1015	0.0995	0.71
2	0.2 : 0.3 : 0.5	0.1987	0.3030	0.4983	0.43
3	0.6 : 0.2 : 0.2	0.5984	0.2024	0.1992	0.62
4	0.3 : 0.3 : 0.4	0.2985	0.3031	0.3984	0.63
5	0.2 : 0.4 : 0.4	0.1985	0.4036	0.3978	0.73

The result obtained from factor analysis shown in Table 4.6 indicates that when factor analysis is used, the peak-overlap problem is minimized. Because factor analysis monitors the relative intensity of each reference at a particular wavelength, while the conventional method uses the shape of peaks for the analysis.

4.1.4   Effect of Nonlinearity in Simulated Spectra

Since factor analysis is applicable under the Beer-Lambert assumption the linear relationship between the absorbance and the concentration of the sample is assumed. However, nearly all-actual data from the measurement contain some nonlinearlities. The nonlinearity originate from the interactions among various components in the system and the non-ideal behavior in the instrument. These cause the nonlinearities such as peak shift and band broadening. In order to make the data as realistic as possible, the nonlinearlity was added to the data. The 5% and 20% nonlinear effects were added to the data of component 1 by deducing every absorbance value by Equations [4.1] and [4.2].

$$A_{5\%nonlinearity} \quad = \quad A - 0.05\ A^2 \qquad [4.1]$$

$$A_{20\%\ nonlinearity} \quad = \quad A - 0.20\ A^2 \qquad [4.2]$$

where $A_{nonlinearity}$ is the new value of the absorbance with the nonlinearity and A is the original absorbance. The difference in the response between a perfectly linear, 5%, and 20% nonlinearlity were illustrated in Figure 4.6.
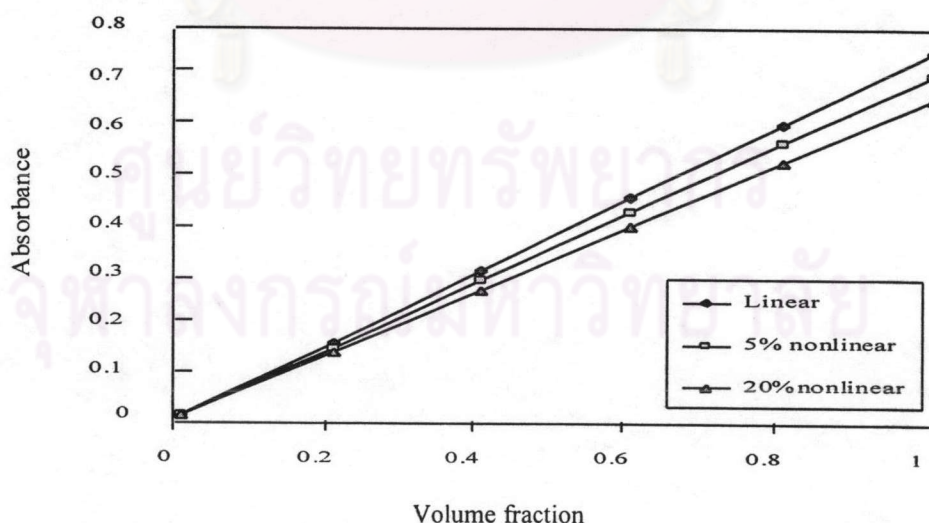

Figure 4.6   The difference in the response between perflectly

linear, 5 % , and 20% nonlinerity added.

According to Figure 4.7, the absorbance with nonlinearity is significantly different from the original data at the approximate volume fraction 0.6. The nonlinearlity data of component 1 was mixed with components 2 and 3 at volume fractions 0.6, 0.7, and 0.8.

Table 4.7   Results of perfectly linear and nonlinearity in simulated spectra

Mixture	Component	Volume fraction	Result		
			Perfectly linear	5% nonlinearlity	20% nonlinearlity
1	1	0.8000	0.7989	0.7992	0.7990
	2	0.1000	0.0996	0.0995	0.0995
	3	0.1000	0.1015	0.1013	0.1014
2	1	0.7000	0.6991	0.6994	0.6992
	2	0.1000	0.1014	0.1012	0.1013
	3	0.2000	0.1996	0.1994	0.1995
3	1	0.6000	0.5973	0.5978	0.5976
	2	0.3000	0.3033	0.3029	0.3031
	3	0.1000	0.0994	0.0993	0.0993
4	1	0.3000	0.3031	0.3027	0.3029
	2	0.3000	0.2985	0.2989	0.2987
	3	0.4000	0.3984	0.3983	0.3984
5	1	0.2000	0.1986	0.1990	0.1988
	2	0.4000	0.4036	0.4032	0.4034
	3	0.4000	0.3978	0.3978	0.3978
PRESS			0.0919	0.1016	0.1117
s^2			0.0064	0.0065	0.0067

The results from Table 4.7 show that the nonlinearity reduces the accuracy of factor analysis, but do not highly differ from the original data.

4.1.5 The Improvement of Prediction Quality in Factor Analysis

According to the three rules of the number of calibration sample in calibration set, when the higher accuracy of the result was required, the number of calibration sample will be increased. Table 4.8 presents the performance of factor analysis with three-component system using the different number of calibration sample.

Table 4.8 Comparison between true value and results obtained by factor analysis of three-component system using the different number of calibration sample.

Mixture	Component	Volume fraction	Results by using number of calibration samples		
			9 (Rule of 3)	15 (Rule of 5)	30 (Rule of10)
1	1	0.5000	0.4964	0.4971	0.4988
	2	0.4000	0.4055	0.4035	0.4007
	3	0.1000	0.0981	0.0994	0.9995
2	1	0.1000	0.0997	0.1007	0.0988
	2	0.1000	0.0997	0.0992	0.1001
	3	0.8000	0.8027	0.8007	0.8002
3	1	0.4000	0.3990	0.3993	0.3998
	2	0.1000	0.1000	0.1007	0.1007
	3	0.5000	0.5010	0.5000	0.5005
4	1	0.2000	0.1984	0.1989	0.1995
	2	0.3000	0.3030	0.3025	0.3025
	3	0.5000	0.4986	0.4986	0.4900
5	1	0.5000	0.4993	0.4978	0.4990
	2	0.3000	0.3043	0.3029	0.3011
	3	0.2000	0.1985	0.1993	0.1995
PRESS			0.10171	0.10169	0.10164
s^2			0.00645	0.00650	0.00648

The results from Table 4.8 demonstrate that the prediction quality of factor analysis increases when the number of calibration samples increases. In another word, the greater number of samples in calibration set gives the better result. In real measurement the rule of number of calibration sample in calibration set often be compromised. Because the more accuracy requires the additional time and budget. It is nice to have at least an adequate number of calibration samples in calibration set that can provide the acceptable error.

4.2   Experimental Section

4.2.1 Determination of *o*-, *m*-, and *p*-xylene in Ternary Xylene Mixture

Factor analysis was applied to the real mixtures of *o*-, *m*-, and *p*-xylene. The ternary xylene mixtures were obtained by mixing the three pure xylene isomers at designed proportions. The individual spectrum of pure xylene isomers and twenty spectra of ternary xylene mixture were shown in Figures 4.7 and 4.8, respectively.
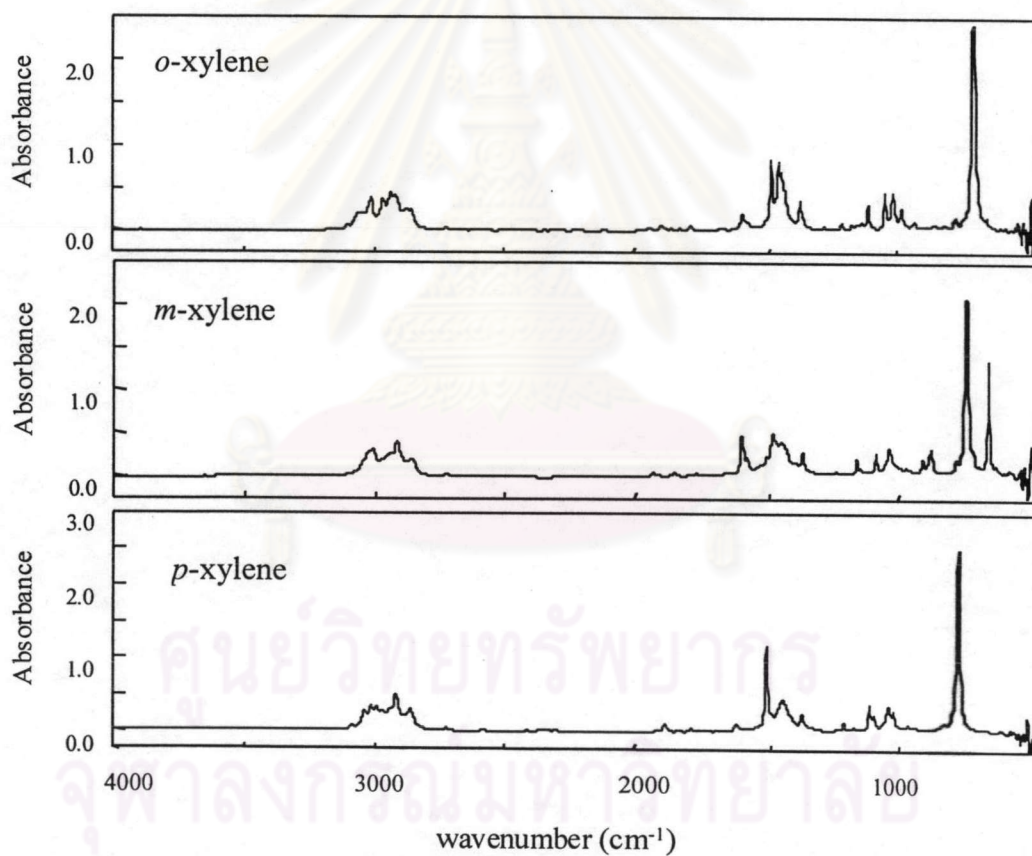

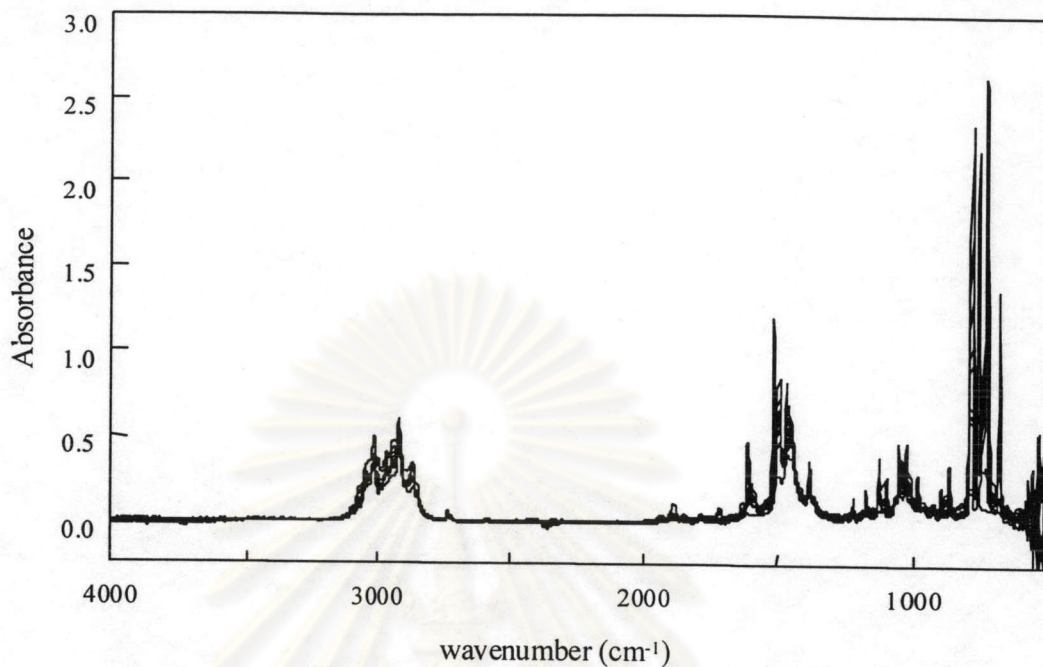Figure 4.7   Infrared spectra of *o*-, *m*-, and *p*-xylene.

Figure 4.8   Twenty infrared spectra of ternary xylene mixtures.

The experimentally obtained spectra were loaded to factor analysis program. Twenty spectra were divided into two sets. The first set called calibration set, which contained twelve spectra of the mixture at various proportions and three pure xylene spectra. The other set called testing set, which contained the five spectra for testing. The results of five samples obtained by factor analysis were shown in Table 4.9.

Table. 4.9   Results of five testing samples obtained from factor analysis program.

Component	Mixture No.				
	1	2	3	4	5
o-xylene					
True value	0.5000	0.5000	0.4000	0.3333	0.0000
Found by EFA	0.4939	0.5160	0.3852	0.3238	0.0000
Difference.	0.0061	0.0160	0.0148	0.0005	0.0000
p-xylene					
True value	0.2000	0.0000	0.1000	0.3333	0.3333
Found by EFA	0.2073	0.0000	0.1117	0.3651	0.3331
Difference.	0.0073	0.0000	0.0117	0.0082	0.0002
m-xylene					
True value	0.3000	0.5000	0.5000	0.3333	0.6666
Found by EFA	0.2988	0.4840	0.5031	0.3110	0.6669
Difference	0.0012	0.0160	0.0031	0.0223	0.0003
PRESS	0.1030				
s^2	0.0064				

The results from Table 4.9 show that there is some errors when evaluating the factor analysis in ternary xylene mixture. Actually, there is the error from experiment but the other is should be caused by the calculation process of factor analysis. This could be improved if the nature of physical property of xylene that affects with the xylene spectra was studied.

The full spectra of mixed xylene contained various spectral regions that present the different results. The full spectrum of 1:1:1 of o-, m-, and p-xylene was studied as the example. The figure of 1:1:1 volume fraction of o-, m-, and p-xylene spectrum was shown in Figure 4.9.
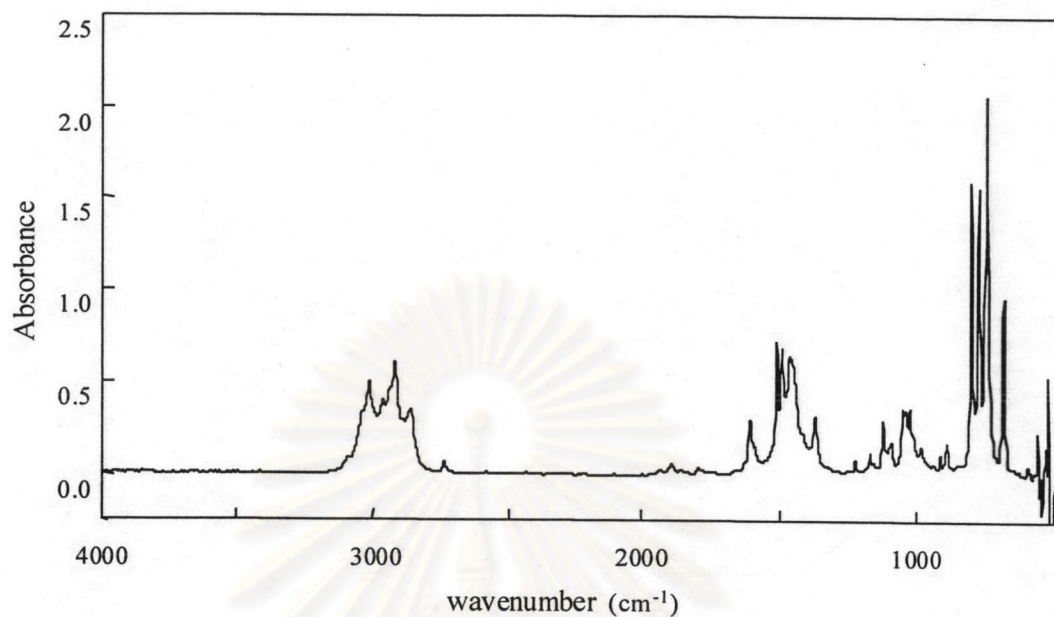
Figure 4.9 Infrared spectrum of 1:1:1 *o*-, *m*-, and *p*-xylene mixture.

The full spectrum of mixed xylene was separated into four regions and the data of these regions were employed for the calculation in factor analysis program. The result obtained from the program was shown in Table 4.10.

Table 4.10   Results of the real 1:1:1 mixture of three xylene isomers obtained by
applying factor analysis to various spectral regions.

Spectral region (cm^{-1})	Volume fraction obtained by factor analysis		
	o-xylene 0.3333	*p*-xylene 0.3333	*m*-xylene 0.3333
2800-3100	0.9529	0.001	0.0470
1750-1950	0.1188	0.4776	0.4036
850-1250	0.1788	0.4974	0.3240
600-900	0.3318	0.3453	0.3229
500-4000	0.3040	0.3797	0.3162

The results from Table 4.10 illustrate that the analysis in the region of 600-900 cm^{-1} gives the more accurate results than the other regions. This region is corresponding to the in-plane and out-of-plane C-H bending vibrations of the aromatic ring. There exist the characteristic peaks of *p*-xylene at 795 cm^{-1}, *m*-xylene at 768 cm^{-1}, and *o*-xylene at 739 cm^{-1} in this region. As described in chapter 1, the conventional infrared analysis uses this band for the analysis. However, this band presents the deviation from Beer-Lambert law and causes the conventional method inapplicable. The results obtained from factor analysis show that it could solve this problem. But the previous study[16], which investigated the mixture of 1:1:1 of *o*-, *m*-, and *p*-xylene by applying least-square method pointed out that this region give the bad result caused from their narrow band-shape and must apply the baseline fitting and first derivative to the sample spectra before performed least-squares analysis.

The combination region 1750 –1950 cm^{-1} shows bad result because this region contains low intensity and there may be the effect of poor baseline correction. For the region of C-H stretching 2800-3100 cm^{-1}, this region gives the

poorest results. From the previous study[18], it was described that this band is the most spectral variation in mid-infrared region. This shows the limitation of factor analysis when applied to the data that presents the great deviation from Beer-Lambert law.

The result from the region of 850-1250 cm^{-1} is shows the poor accuracy. Haleness and Easterling[18] pointed out that this band contains the strong and narrow band shape caused from the great molecular perturbation of the bending variation of the aryl hydrogen. This can be explained that factor analysis unable to evaluated the highly strong and narrow band shape.

4.2.2 Determination of *o*-, *m*-, and *p*-xylene in Real Mixed Xylene.

The study of the ability of factor analysis being used to determine *o*-, *m*-, and *p*-xylene in the actual mixed xylene which unknown proportions of three isomers was performed. In addition, the actual mixed xylene contains three isomers and other components such as toluene, ethyl benzene or other aromatic substance upon the source of mixed xylene. The spectrum of mixed xylene was shown in Figure 4.10.
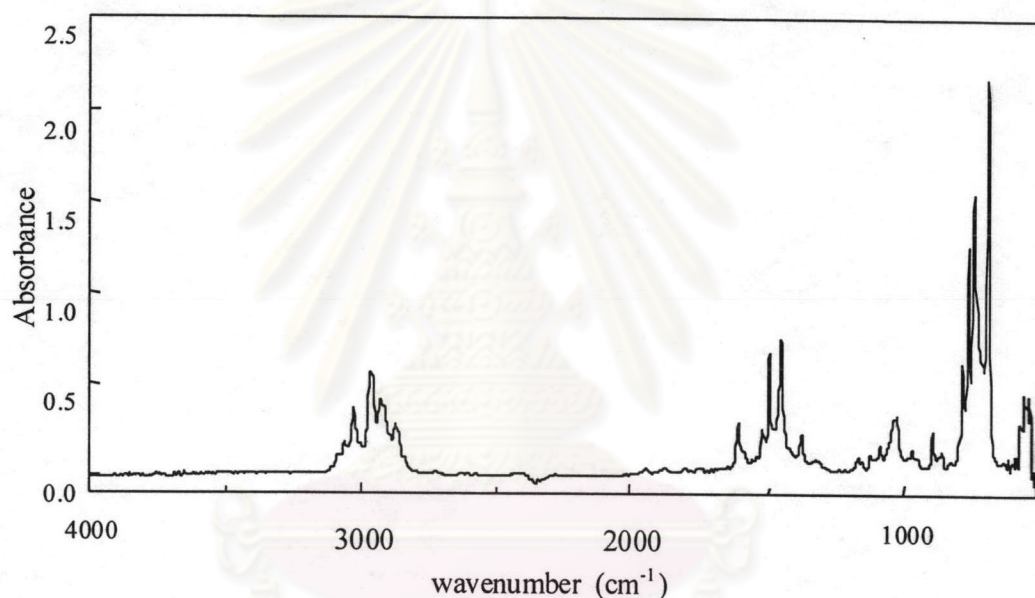
Figure 4.10 Infrared spectrum of unknown mixed xylene.

The unknown mixed xylene was separated into 2 parts, one was sent to the reliable lab for the quantitative analysis by GC and the other one was tested by infrared spectroscopy. The infrared spectrum of unknown mixed xylene was measured and recorded as the testing sample. The calibration set was constructed from fifteen mixed xylene with varying amount of *o*-, *m*-, and *p*- xylene. The results of *o*-, *m*-, and *p*- xylene in the unknown mixed xylene obtained by factor analysis were shown in Table 4.11 compared with the results from GC testing. The detail and reported result of GC testing are shown in Appendix D.

Table 4.11   Results of *o*-, *m*-, and *p*-xylene in mixed xylene obtained from infrared
spectroscopy and GC.

Component	Quantitative Analysis Technique	
	Applied FA to IR spectroscopy (% weight)	Gas Chromatography (% weight)
o-xylene	11.9	10.9
m-xylene	20.9	23.2
p-xylene	10.9	9.9

The results of *o*-, *m*-, and *p*-xylene obtained from the two analytical methods exhibit some differences, however the same trend in the relative proportion of the three isomers is observed. As seen from the results of both techniques, the unknown mixed xylene contains *m*-xylene mostly, then *o*- and *p*-xylene, respectively. When the differences between the results from two techniques were considered, they are approximately 10% different for all data. This result indicated the feasibility of quantitative analysis of mixed xylene by factor analysis from FT-IR spectra.

The differences between the results obtained from two methods should come from many reasons. Number of components in mixed xylene is a common problem in factor analysis. In this experiment, only three isomers of xylene were varied for constructing the calibration samples and other components were assumed to be constant. The error also possibly comes from numerical method of factor analysis when the technique is applied to mixture in which some components are varied. Moreover, there are the variations from unexpected effect in the real mixed xylene. This clearly indicates that the stability of factor analysis depends on several factors. As a consequence, the factor analysis should be studied and tested with the real mixed xylene until the satisfying result is obtained prior to apply with the real measurement.