

XRD 스펙트럼의 비음독립성분분석을 통한 혼합물 구성비 결정

유한민^{1*} · 전치혁² · 이혜선² · 홍재화³

¹포항공과대학교 정보통신대학원, ²산업경영공학과,

³포스코기술연구소 계측연구그룹

(2007. 9. 5. 접수. 2007. 10. 19. 승인)

Determination of mixing ratios in a mixture via non-negative independent component analysis using XRD spectrum

Hanmin You^{1*}, Chi-Hyuck Jun², Hyeseon Lee² and Jae-Hwa Hong³

¹Graduate School for information Technology, POSTECH, Pohang 790-784, Korea

²Department of Industrial and Management Engineering, POSTECH, Pohang 790-784, Korea

³Instrumentation Research Group, Technical Research Laboratory, POSCO, Pohang 790-785, Korea

(Received September 5, 2007; Accepted October 19, 2007)

요약: X 선 회절법은 물질 내부에 원자의 배열 상태를 연구하는데 널리 사용되는 실험 방법으로써 넓은 응용 범위를 가지고 있다. 특히 분말 X 선 회절법은 비파괴적으로 다양한 형태의 시료에 대한 측정이 가능하기 때문에 결정의 배향성, 결정의 크기, 결정 내부의 응력 측정 등에 널리 이용되고 있다. 분말 X 선 회절 방법을 이용하여 혼합물의 성분을 정량적으로 측정하기 위해서는 시료를 구성하고 있는 source 스펙트럼을 도출하고 혼합된 시료의 XRD 스펙트럼에 회귀식을 적합시켜 혼합물 구성비를 얻는 방법이 제안된 바 있다. 그러나 구성성분의 특성상 스펙트럼의 피이크가 폭이 좁고 민감한 경우에는 노이즈의 영향을 받아 도출된 source 스펙트럼이 원래의 순수성분의 형태와 달리 나타날 수 있다. 특히 순수성분 시료를 구할 수 없거나 측정이 불가능한 경우 혼합물 구성비 측정에 어려움이 있다. 본 연구에서는 노이즈 간섭에 의한 source 스펙트럼 도출의 문제를 해결하는 방안으로 비음독립성분분석을 이용하여 혼합된 미지시료로부터 순수한 성분에 해당하는 스펙트럼을 분리해내는 방법을 제안하고자 한다.

Abstract: X-ray diffraction method has been widely used for qualitative and quantitative analysis of a mixture of materials since every crystalline material gives a unique X-ray diffraction pattern independently of others, with the intensity of each pattern proportional to that material's concentration in a mixture. For determination of mixing ratios, extracting source spectra correctly is important and crucial. Based on the source spectra extracted, a regression model with non-negativity constraint is applied for determining mixing ratios. In some mixtures, however, X-ray diffraction spectrum has sharp and narrow peaks, which may result in partial negative source spectrum from independent component analysis. We propose several procedures of extracting non-negative source spectra and determining mixing ratios. The proposed method is validated with experimental data on powder mixtures.

Key words: X-ray diffraction, nonnegative independent component analysis, nonnegative least squares

★ Corresponding author

Phone : +82-(0)54-279-2854 Fax : +82-(0)54-279-2870

E-mail : inha99@postech.ac.kr

1. 서 론

X 선 회절법(X-Ray Diffraction: XRD)은 X 선이 시료 내에서 회절하는 현상을 이용하여 물질의 결정구조를 밝혀내는데 적용되는 분광기술이다. 다양한 형태의 시료에 사용되며 시료의 회절강도 패턴으로부터 미지시료내의 물질을 확인할 수 있고, 혹은 이미 확인된 물질의 분자구조를 분석할 수 있다.

기존에는 분말 X 선 회절 방법을 이용하여 혼합물의 성분을 정량적으로 측정하기 위해 시료를 구성하고 있는 순수성분에 대한 각각의 XRD 스펙트럼을 측정하고 혼합된 시료의 XRD 스펙트럼을 측정하여 이를 비교하는 방법이 일반적으로 사용되었다. 하지만 이 방법은 시료를 구성하는 순수한 물질을 구할 수 없거나 측정이 불가능 한 경우 사용 할 수 없을 뿐 아니라 이것들을 측정, 비교하는 시간도 많이 걸렸다.

혼합물 시료로부터 서로 독립인 성분을 분리해내는 것을 미지성분분리(blind source separation)라고 하는데, Lin and Liu¹의 자동화 스펙트럼 분리 기법, Windig and Meuzelaar²의 분산 다이어그램 기법, Schostack and Malinowski³의 반복 중요집합 요인분석등이 미지성분분리의 초기연구에 해당된다. 최근에는 독립성분 분석기법(independent component analysis)이 시그널 데이터로부터 독립성분을 분리하는데 가장 많이 사용되는 미지성분분리 기법으로서 패턴인식, 영상의학, 컴퓨터단층촬영 등에 적용되고 있다.⁴

Lee *et al.*⁵은 독립성분분석을 이용하여 혼합물 데이터로부터 서로 독립인 순수성분을 분리하고, 비음최소자승법을 사용하여 혼합물 구성비를 예측하는 절차를 제안하였다.

하지만 독립성분분석기법의 특성상 도출된 측정종류나 형태에 따라 스펙트럼 피크들의 폭이 민감하고 좁게 나타나는 경우 성분간 노이즈가 섞일수 있어서 분리된 독립성분들이 순수성분일때와 다소 다른 형태로 나타날수 있고, 그 중 하나가 독립성분 중 부분적으로 음수의 형태가 나오는 경우이다.

본 연구에서는 Hyvärinen and Oja⁶ 이 제시한 Fast ICA 알고리즘을 기반으로 도출된 독립성분이 부분적으로 음수인 경우 다음과 같이 두가지의 혼합물 구성비 결정 절차를 제안한다. 첫째, 데이터 전처리와 독립성분도출후 처리에 의한 혼합물 구성비 예측 방법과, 둘째, 독립성분분석 알고리즘 수행시 비음(non-negative) 제약식을 주는 방법이다.

이 절차는 혼합 비율이 사전에 알려진 임의의 비율

로 섞은 혼합 분말들의 엑스선 회절강도값에 적용하여 예측 정확도를 구하고, 제안방법의 타당성 평가를 위해 동일한 데이터를 부분 최소자승법(partial least square method; PLS)에 적용한 결과와 비교하였다.

2. 이 론

본 연구에서는 혼합물의 구성성분 비율을 구하기 위해 Hyvärinen and Oja⁶이 제시한 Fast ICA 알고리즘을 부분적으로 수정한 비음독립성분분석기법을 제안하였으며 다른 연구에 비해 상대적으로 간단하면서도 알고리즘의 수렴속도가 빠르다는 장점이 있다. 이와 관련하여 우선 비음독립성분분석에 대해 간략히 소개하면 다음과 같다.

비음독립성분분석은 주어진 혼합 데이터에 대해 source(S)가 비음이고 근저성(well-grounded)이라는 조건하에 서로 독립인 source를 추출하는 통계적 기법이다.⁷ 즉, Fast ICA 알고리즘에 적용했을 때 발생하는 독립성분의 음수부분을 처리하면서 동시에 비정규성이 최대가 되는 source를 분리하는 것이다. 만약 실수 값을 가지는 p 개의 미지 독립성분 (S_1, S_2, \dots, S_p)이 있다고 가정하는 경우 개별 독립성분에 대한 두가지 조건을 수식으로 표현하면 다음과 같다.⁸

$$\text{Non-negative: } \Pr(S_i < 0) = 0, \quad (i = 1, 2, \dots, p)$$

Well-grounded:

$$\Pr(S_i < \delta) > 0, \quad \forall \delta (> 0), (i = 1, 2, \dots, p)$$

본 연구에서 제안하는 비음독립성분분석 알고리즘은 기존방법에 비해 간단하면서도 실용적인 점에 목적을 두고 두 가지 관점에서 접근하였다. 첫번째는 전처리와 후처리 과정을 통해 데이터 변환을 함으로써 독립성분의 비음조건을 적용하는 방법(제안방법 I)이고, 두번째는 Fast ICA 알고리즘내의 반복과정에서 비음의 제약식을 주어 비음의 독립성분을 추출하는 방법이다(제안방법 II).

3. 실험방법

3.1. 시료

본 연구에서 사용된 시료는 혼합분말로써 철의 제조 과정에서 촉매제로 사용되는 몰드플럭스의 주요 구성성분인 SiO_2 , TiO_2 , MgO 로 구성되어 있다. 각각의 분말시료는 일본 Junsei 사의 extra pure 급을 사용하였다. 본 실험에서는 3개의 독립성분 SiO_2 , TiO_2 ,

MgO을 실험자가 임의의 구성비율로 조합하여 10개의 혼합분말 시료를 만들었는데 시료별 실제 혼합비율은 Table 1에 표시하였다.

참고로 본 실험에서는 SiO₂, TiO₂, MgO에 대해 각각 100%로 구성된 순수한 시료를 추가로 형성하였으며, 이에 대해 XRD를 적용해서 얻은 표준 독립성분 스펙트럼은 Fig. 1과 같으며 각 독립성분은 일정한 각도에서 고유의 피크를 가지고 있는 것을 확인할 수 있다(SiO₂: 26.5°, TiO₂: 25.5°, MgO: 43°). 참고로 Fig. 1은 2θ에서 70°까지 0.05° 간격의 1000관측차원(x축)에서 측정된 엑스선 회절강도값(y축)이므로 혼합물 데이터는 1000×10 행렬이 된다.

XRD 스펙트럼은 Bruker 사의 D4 모델을 이용하여 측정하였다. X 선원으로는 구리 음극을 사용한 봉입형 X-ray tube를 이용하였으며 검출기는 Scintillation counter를 사용하였고 측정 조건은 Table 2와 같다. 본 연구에서 사용한 10개의 혼합분말 시료에 대해 X선 회절 강도값을 측정하면 Fig. 2와 같은 스펙트럼을 얻는다.

3.2. 분석조건

XRD분석기기의 개별 분석조건에 대해 설명하면 다음과 같다. 첫번째, Rotation은 시료의 분당 회전수로써 클수록 엑스선을 균등하게 찍는다. 두번째, Power

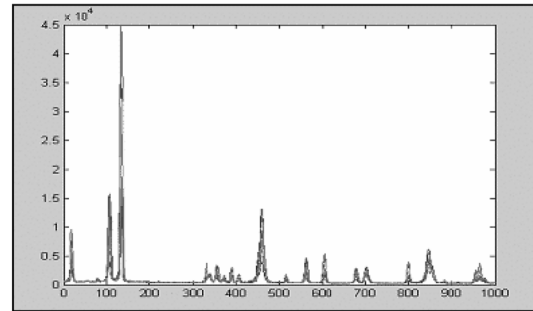


Fig. 2. Mixing powder spectra of 10 samples.

는 전압과 전류를 통해 조정할 수 있으며 클수록 정확도가 높은 스펙트럼을 얻을 수 있지만 기기의 성능상 한계가 있다. 세번째, 회절각도범위는 시료의 일부분만 엑스선을 찍어도 그 시료의 특징을 알 수 있으므로 특정 범위안에 해당하는 각도를 지정할 수 있다. 네번째, Increment는 정해진 각도를 초당 몇도씩 엑스레이를 찍는지를 나타내는 값으로 작을수록 스펙트럼의 형태가 매끄러워지고 관측차원(x축)이 커진다. 다섯번째, Scanspeed는 정해진 각도를 스캔하는 속도로써 클수록 시료의 더 세밀한 측정이 가능하다. 여섯번째, Anti, Scatterity Slit, Div Scatterity Slit, Detector Slit는 엑스레이 튜브와 검출기에 설치된 슬릿들로서 이것들의 크기를 조정해서 X선에서 나오는 빛의 양을

Table 1. Actual mixing ratios of samples

	1	2	3	4	5	6	7	8	9	10
SiO ₂	0.456	0.256	0.233	0.449	0.603	0.154	0.213	0.183	0.366	0.301
TiO ₂	0.114	0.361	0.360	0.355	0.141	0.490	0.627	0.487	0.214	0.387
MgO	0.430	0.374	0.408	0.196	0.256	0.356	0.160	0.330	0.420	0.313

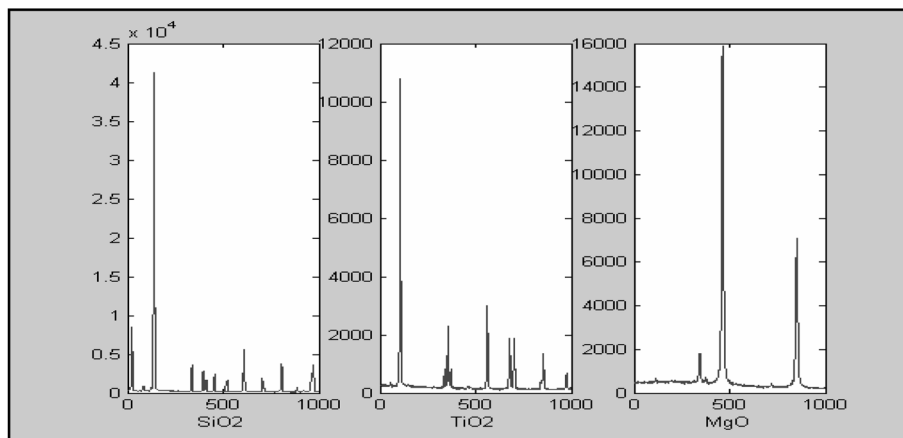


Fig. 1. XRD spectra of pure SiO₂, TiO₂, MgO.

Table 2. XRD tuning conditions

X-ray parameters	
Rotation (r.p.m)	30
Power (Kv, mA)	40, 40
Degree of diffraction (°C)	20~70
Increment (sec/step)	0.05
Scanspeed (degree/step)	0.5
Anti, Div, Detector	2 mm, 2 mm, 0.2 mm

조절할 수 있다.

4. 실험결과 및 분석

4.1. 기존방법 적용결과

실험에서 얻어진 혼합물 데이터를 Lee *et al.*⁵ 이 제안한 절차에 적용시키면 Fig. 3과 같은 세개의 독립성분 스펙트럼을 얻을 수 있다.

하지만 Fast ICA 알고리즘을 사용하는 경우 Fig. 1에 비해 Fig. 3의 점선사각형안에 표시된 것처럼 본래 존재하지 않는 음수부분이 발생하는 문제가 생기는데 그 이유는 개별 독립성분내에 존재하는 성분들 간의 분산이 커서 추정과정에서 상호간섭이 발생하기 때문이다.

본 논문에서는 구성성분 비율의 전반적인 예측정확도를 비교하는 척도로써 평균절대오차(mean absolute deviation: MAD)를 사용하였고 그 식은 다음과 같다.

$$MAD = \frac{1}{3 \times 10} \sum_{j=1}^{10} \sum_{i=1}^3 |A_{ji} - \hat{A}_{ji}|$$

여기서 A_{ji} 는 j 번째 시료에 해당하는 i 번째 독립성분의

Table 3. MAD of previous method

	SiO ₂	TiO ₂	MgO	average
Existing	0.145	0.083	0.066	0.098

실제 성분비율, \hat{A}_{ji} 는 예측 성분비율을 의미한다. 시료별 실제 구성성분비율과 기존방법을 이용한 시료별 예측 구성성분비율을 비교하여 계산한 평균절대오차는 Table 3에 표시하였다.

Table 3의 결과 중 SiO₂가 상대적으로 큰 이유는 Fig. 3과 같이 26.5°에서 나타나는 SiO₂의 피크가 독립성분간의 간섭으로 인해 동일한 각도에서 MgO의 독립성분의 음수부분으로 잘못 추출되어서 정보량이 감소되었기 때문이다. 또한, TiO₂, MgO도 본래 존재하지 않는 음수부분이 발생함으로 인해 예측 정확도가 떨어진 것으로 예상할 수 있다.

4.2. 제안방법 I 적용결과

혼합물 구성비를 결정하기 위한 제안방법 I은 두가지 방법을 제안한다. 첫번째는 전처리 과정에서 로그변환한 데이터를 기존방법에 적용하여 혼합물 구성비를 구하는 방법이고, 두번째는 로그변환한 데이터를 기존방법에 적용해서 도출된 독립성분을 지수연산을 해준 후 이를 최소자승법을 사용하여 혼합물 구성비를 구하는 방법이다. 제안방법 I을 적용해서 얻은 평균절대오차는 Table 4에 표시하였다.

Table 4의 결과를 살펴보면 Table 3의 평균절대오차보다 낮은 것을 알 수 있다. 그 이유는 본 논문에서 사용한 혼합분말 데이터의 경우 개별 독립성분의 성분간의 분산이 매우 크므로 전처리 과정에서 데이터에 로그변환을 해줌으로써 개별 독립성분의 성분간

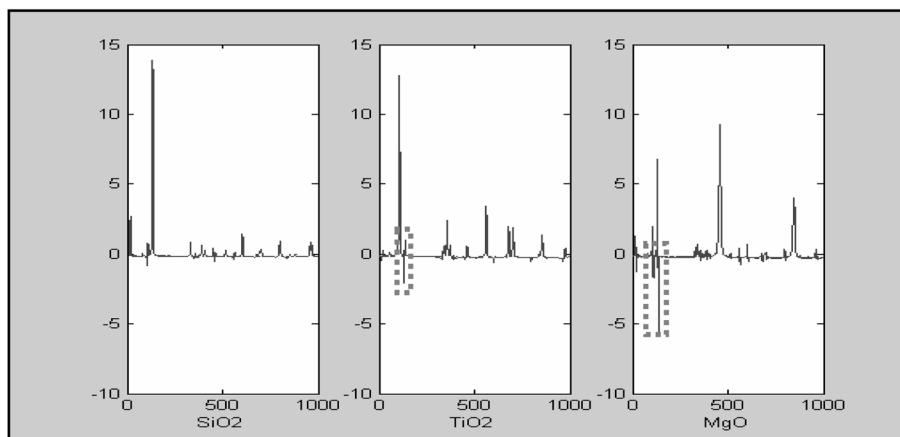


Fig. 3. Source spectra extracted by the previous method.

Table 4. MAD of proposed method I

	SiO ₂	TiO ₂	MgO	average
method I-1	0.054	0.070	0.041	0.055
method I-2	0.031	0.062	0.037	0.043

분산을 줄이는 동시에 독립성분 전체의 분산을 줄이는 효과를 얻을 수 있기 때문이다. 이로 인해 개별 혼합데이터 내에 포함된 구성성분들의 특징이 명확해져서 기존방법을 이용한 독립성분의 추출이 용이해졌다. 또한 방법 I-2의 경우 예상한대로 후처리 과정에서 지수연산을 한 독립성분을 최소자승법에 적용함으로써 방법 I-1보다 예측정확도가 높아졌다.

4.3. 제안방법 II 적용결과

혼합물의 구성성분비율을 구하기 위한 제안방법 II는 비음독립성분분석을 적용하여 독립성분을 도출하고, 이 시그널의 노이즈를 제거한 부분만 추출하여 비음최소제곱법에 적용하는 방법으로써 알고리즘 내에서 비음 처리하는 방식에 따라 세가지 방법을 적용하였으며 식은 다음과 같다.

Method II-1: If $w^T x \leq 0, w^T x(i,j) = -w^T x(i,j)$

Method II-2: If $w^T x \leq 0, wx(i) = \left| \frac{w^T x(i,j)}{\min w^T x(i,j)} \right|$

Method II-3: If $w^T x \leq 0, w^T x(i,j) = 0$

한 예로 제안방법 II-1의 경우 비음독립성분분석을 적용한 후 노이즈를 제거할 때 시그널이 일정수준 (none, 0.1, 0.5, 1, 2, 3, 4, 5)보다 큰 범위에 대하여 회귀분석을 실시한 결과는 Fig. 4과 같다. 이를 통해 절사점이 1인 점에서 평균절대오차가 최소가 되는 최적값이 존재하는 것을 알 수 있으며 이 점에서 세가지 독립성분간의 균형이 맞는 결과가 도출되었다. 이를 통해 최적 절사점으로 갈수록 평균절대오차가 가장 작은 성분의 방향으로 다른 성분들도 수렴해간다는 것을 알 수 있다.

제안방법 II를 적용해서 얻은 평균절대오차는 Table 5에 표시하였다. Table 5를 통해 방법 II-1이 다른 방법에 비해 상대적으로 평균절대오차가 작은 것을 알 수 있다. 그 이유는 방법 II-2, 3 경우 비정규성이 최대가 되는 독립성분을 도출하는 과정에서 비음처리를 위해 독립성분의 음수부분을 제거하고 이로 인해 정보의 손실을 가져온 반면 방법 II-1의 경우 음수부분

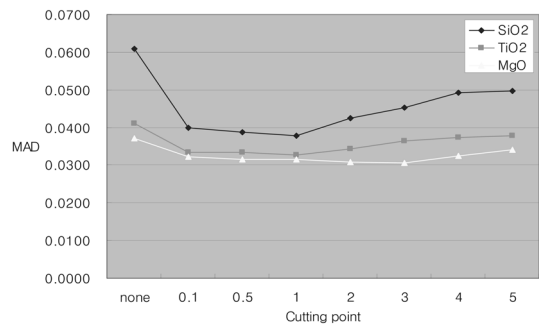


Fig. 4. MAD vs cutting value.

Table 5. MAD of proposed method II

	SiO ₂	TiO ₂	MgO	average	Cutting value
methodII-1	0.038	0.032	0.031	0.034	1
methodII-2	0.044	0.035	0.043	0.040	1
methodII-3	0.041	0.051	0.044	0.045	0.5

을 양수로 변환시킴으로써 기존의 정보를 보존하기 때문이다.

4.4. 종합 분석결과

이제까지 분석한 결과를 Fig. 5으로 나타내었다. 이를 통해 제안방법 II-1이 독립성분간의 오차의 균형이 맞으므로 안정적인 방법이라고 할 수 있으며, 전체적으로 가장 좋은 성능을 나타내는 것을 알 수 있다. 제안방법 II-1이 제안방법 I-2에 비해 결과가 좋은 이유는 단순한 데이터의 전처리나 후처리의 효과보다 알고리즘내의 비정규성을 최대화시키는 반복 과정에서 비음처리를 한것의 효과가 크기 때문이다.

4.5. 제안방법의 타당성 검증

혼합분말 데이터는 Table 5와 같이 구성성분의 시

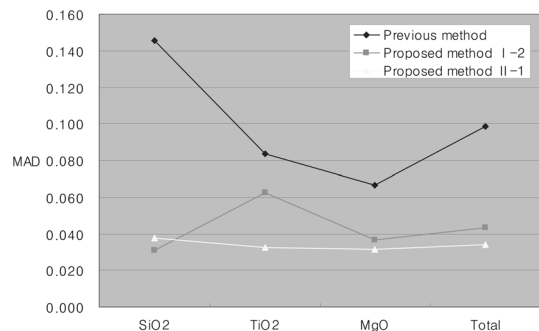


Fig. 5. MAD comparison of previous and proposed methods.

Table 6. MAD of PLS method

	SiO ₂		TiO ₂		MgO		average	
	Training	Test	Training	Test	Training	Test	Training	Test
PLS	0.004	0.041	0.004	0.047	0.004	0.037	0.004	0.042

료별 실제 혼합비율을 알고 있고 학습표본이 존재하므로 이를 이용한 지도학습기법이 사용 가능하다. 그 중 부분최소자승법은 독립변수와 종속변수의 분산을 동시에 고려하여 모델의 설명력을 높이는 회귀분석방법으로, 특히 종속변수의 개수가 독립변수에 비해 작은 경우 유용한 방법이다. 혼합분말 데이터의 경우 주어진 종속변수가 10개이고 독립변수가 1000개인 행렬로써 일반적으로 쓰이는 다변량 회귀분석을 적용하기는 불가능하다. 따라서 본 연구에서는 부분최소자승법을 이용하여 구성성분의 추정 혼합비율을 도출하고 이를 시료별 실제 혼합비율과 비교하여 표준절대오차를 계산하여 제안방법의 결과와 비교하였다.

Table 6의 결과에서 Test 데이터의 오차와 비교해보면 reference 성분비율이 존재하는 경우의 예측모델링 성능과 유사함을 볼 수 있고, 방법 II-1의 경우는 오차가 더 작은 것을 알 수 있다. 따라서 제안방법은 시료별 실제 혼합비율을 알지 못하는 조건에서도 혼합물 구성비를 추정하는데 신뢰할 수 있는 방법이라고 할 수 있다.

5. 결 론

본 연구에서는 비음독립성분분석을 이용하여 독립성분이 양수라는 제약조건하에 다변량 데이터로부터 서로 독립인 성분을 분리한 다음, 비음최소자승법을 사용하여 혼합물 구성비를 예측하는 방법을 제안하였다. 제안방법의 타당성을 검증하기 위해서 동일한 데이터를 부분최소자승법에 적용하여 정확도를 비교한 결과 Fast ICA 알고리즘내에서 비정규성을 최대화시키기 위한 반복과정마다 비음처리를 하는 방법의 결과가 더 좋은 것을 확인하였다.

분석결과를 통해 분말 X 선 회절 방법을 이용하여

혼합물 구성비를 정량적으로 측정하기 위해 제안된 절차는 예측 정확도 측면에서 안정된 방법이다. 본 제안방법은 습식방법에 의한 측정이 어려운 경우나 혼합 성분간 상호간섭 혹은 노이즈로 인해 source 도출에 어려움이 있는 경우 적용할 수 있는 유용한 방법이라고 하겠다.

감사의 글

본 연구는 포스코 기술연구소의 연구개발사업 지원에 의해 이루어진 것이며, 이에 감사 드립니다.

참고문헌

1. C. H. Lin, S. C. Liu, *Journal of Chinese Chemical Society*, **25**, 167-177 (1978).
2. W. Windig, H. L. C. Meuzelaar, *Analytical Chemistry*, **56**(13), 2297-2303 (1984).
3. K. J. Schostack, E. R. Malinowski, *Chemometrics and Intelligent Laboratory Systems*, **6**, 21-29 (1989).
4. S. G. Choi, A. Cichocki, H. M. Park, S. Y. Lee, *A Review, Neural Information Processing*, **6**, 1-57 (2005).
5. H. S. Lee, C. H. Jun, J. K. Song, J. W. Hong, *Chemometrics and Intelligent Laboratory Systems*, **87**(1), 107-113 (2007).
6. A. Hyvärinen, E. Oja, *Neural Networks*, **13**(4-5), 411-430 (2000).
7. M. D. Plumbley, *Neural Networks*, **14**(3), 534-543 (2003).
8. Z. J. Yuan, E. Oja, A, *Lecture Notes in Computer Science*, **3195**, 1-8 (2004).