



(19) 대한민국특허청(KR)  
(12) 공개특허공보(A)

(11) 공개번호 10-2017-0134467  
(43) 공개일자 2017년12월06일

- (51) 국제특허분류(Int. Cl.)  
*G10L 19/008* (2014.01) *G10L 19/02* (2006.01)  
*G10L 21/0272* (2013.01) *H03M 1/12* (2006.01)
- (52) CPC특허분류  
*G10L 19/008* (2013.01)  
*G10L 19/02* (2013.01)
- (21) 출원번호 10-2017-7028242
- (22) 출원일자(국제) 2016년03월10일  
심사청구일자 없음
- (85) 번역문제출일자 2017년10월02일
- (86) 국제출원번호 PCT/EP2016/055135
- (87) 국제공개번호 WO 2016/162165  
국제공개일자 2016년10월13일
- (30) 우선권주장  
15305536.3 2015년04월10일  
유럽특허청(EPO)(EP)  
(뒷면에 계속)

- (71) 출원인  
톰슨 라이센싱  
프랑스 92130 이씨레탈리노 루 잔다르크 1-5
- (72) 발명자  
빌렌 카그다스  
프랑스 35576 세송 세비네 아브뉴 테 샹 블랑 975  
자끄 데 샹 블랑 씨에스 176 16 테크니컬러 알 앤  
드 디 프랑스  
오제로프 알렉세이  
프랑스 35576 세송 세비네 아브뉴 테 샹 블랑 975  
자끄 데 샹 블랑 씨에스 176 16 테크니컬러 알 앤  
드 디 프랑스  
페레즈 파트리끄  
프랑스 35576 세송 세비네 아브뉴 테 샹 블랑 975  
자끄 데 샹 블랑 씨에스 176 16 테크니컬러 알 앤  
드 디 프랑스
- (74) 대리인  
특허법인코리아나

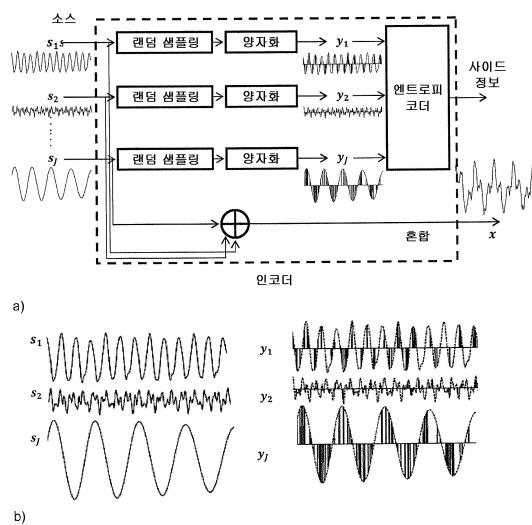
전체 청구항 수 : 총 13 항

(54) 발명의 명칭 다수의 오디오 신호들을 인코딩하기 위한 방법 및 디바이스, 그리고 개선된 분리로 다수의 오디오 신호들의 혼합을 디코딩하기 위한 방법 및 디바이스

### (57) 요약

다수의 오디오 신호들을 인코딩하는 방법은 다수의 오디오 신호들의 각각을 랜덤 샘플링 및 양자화하는 단계, 및 상기 다수의 오디오 신호들의 혼합으로부터 다수의 오디오 신호들을 디코딩 및 분리하기 위하여 이용될 수 있는 사이드 정보로서 샘플링 및 양자화된 다수의 오디오 신호들을 인코딩하는 단계를 포함한다. 다수의 오디오 신호들의 혼합을 디코딩하기 위한 방법은 다수의 오디오 신호들의 각각의 오디오 신호의 양자화된 샘플들을 포함하는 사이드 정보를 디코딩 및 역다중화하는 단계, 임의의 데이터 소스로부터, 상기 다수의 오디오 신호들의 혼합을 수신 또는 취출하는 단계, 및 상기 다수의 오디오 신호들을 근사화하는 다수의 추정된 오디오 신호들을 생성하는 단계를 포함하고, 다수의 오디오 신호들의 각각의 오디오 신호의 상기 양자화된 샘플들이 이용된다.

### 대 표 도 - 도2



(52) CPC특허분류

*G10L 21/0272* (2013.01)

*H03M 1/128* (2013.01)

(30) 우선권주장

15306144.5 2015년07월10일

유럽특허청(EPO)(EP)

15306425.8 2015년09월16일

유럽특허청(EPO)(EP)

---

## 명세서

### 청구범위

#### 청구항 1

다수의 시간 도메인 오디오 신호들을 인코딩하는 방법으로서,

- 상기 다수의 시간 도메인 오디오 신호들의 각각을 랜덤 샘플링 및 양자화하는 단계; 및
- 상기 다수의 시간 도메인 오디오 신호들의 혼합으로부터 상기 다수의 시간 도메인 오디오 신호들을 디코딩 및 분리하기 위하여 이용될 수 있는 사이드 정보로서 샘플링 및 양자화된 상기 다수의 시간 도메인 오디오 신호들을 인코딩하는 단계를 포함하는, 다수의 시간 도메인 오디오 신호들을 인코딩하는 방법.

#### 청구항 2

제 1 항에 있어서,

상기 랜덤 샘플링은 미리 정의된 의사 랜덤 패턴을 이용하는, 다수의 시간 도메인 오디오 신호들을 인코딩하는 방법.

#### 청구항 3

제 1 항 또는 제 2 항에 있어서,

상기 다수의 시간 도메인 오디오 신호의 혼합은 이것이 도착함에 따라 점진적으로 인코딩되는, 다수의 시간 도메인 오디오 신호들을 인코딩하는 방법.

#### 청구항 4

제 1 항 내지 제 3 항 중 어느 한 항에 있어서,

어떤 소스가 어떤 시간 기간에서 무음인지를 결정하는 단계, 및 결정된 정보를 상기 사이드 정보에서 인코딩하는 단계를 더 포함하는, 다수의 시간 도메인 오디오 신호들을 인코딩하는 방법.

#### 청구항 5

다수의 오디오 신호들의 혼합을 디코딩하기 위한 방법으로서,

- 상기 다수의 오디오 신호들의 각각의 오디오 신호의 양자화된 시간 도메인 샘플들을 포함하는 사이드 정보를 디코딩 및 역다중화하는 단계;
- 저장 또는 임의의 데이터 소스로부터, 상기 다수의 오디오 신호들의 혼합을 수신 또는 취출하는 단계; 및
- 상기 다수의 오디오 신호들을 근사화하는 다수의 추정된 오디오 신호들을 생성하는 단계로서, 상기 다수의 오디오 신호들의 각각의 오디오 신호의 상기 양자화된 샘플들이 이용되는, 상기 다수의 추정된 오디오 신호들을 생성하는 단계를 포함하는, 다수의 오디오 신호들의 혼합을 디코딩하기 위한 방법.

#### 청구항 6

제 5 항에 있어서,

상기 다수의 추정된 오디오 신호들을 생성하는 단계는

- 랜덤 비음수 값들로부터 분산 텐서  $V$  를 계산하는 단계;
- 상기 다수의 오디오 신호들의 양자화된 샘플들의 소스 전력 스펙트럼의 조건부 기대치를 계산하는 단계로서, 추정된 소스 전력 스펙트럼  $P(f, n, j)$  이 획득되고 상기 다수의 오디오 신호들의 분산 텐서  $V$  및 복소 단시간 푸리에 변환 (STFT) 계수가 이용되는, 상기 조건부 기대치를 계산하는 단계;

- 상기 추정된 소스 전력 스펙트럼  $\mathbf{P}(f, n, j)$  으로부터 상기 분산 텐서  $V$  를 반복적으로 재산출하는 단계;

- 결과적인 분산 텐서  $V$  로부터 STFT 계수들  $\hat{\mathbf{S}}$  의 어레이를 계산하는 단계; 및

- 상기 STFT 계수들  $\hat{\mathbf{S}}$  의 어레이를 상기 시간 도메인으로 변환하는 단계로서, 상기 다수의 추정된 오디오 신호들이 획득되는, 상기 시간 도메인으로 변환하는 단계를 포함하는, 다수의 오디오 신호들의 혼합을 디코딩하기 위한 방법.

#### 청구항 7

제 5 항 또는 제 6 항에 있어서,

상기 다수의 오디오 신호들 중 적어도 하나에 대한 오디오 인페인팅 단계를 더 포함하는, 다수의 오디오 신호들의 혼합을 디코딩하기 위한 방법.

#### 청구항 8

제 5 항 내지 제 7 항 중 어느 한 항에 있어서,

상기 사이드 정보는 어떤 오디오 소스가 어느 시간 기간에서 무음인지를 정의하는 정보를 더 포함하고, 상기 분산 텐서  $V$ 를 정의하는 행렬  $H$  및  $Q$ 를 자동으로 결정하는 단계를 더 포함하는, 다수의 오디오 신호들의 혼합을 디코딩하기 위한 방법.

#### 청구항 9

다수의 오디오 신호들을 인코딩하기 위한 장치로서,

프로세서, 및 명령들을 저장하는 메모리를 포함하고,

상기 명령들은, 실행될 때, 상기 장치로 하여금,

- 상기 다수의 시간 도메인 오디오 신호들의 각각을 랜덤 샘플링 및 양자화하는 단계; 및

- 상기 다수의 오디오 신호들의 혼합으로부터 상기 다수의 시간 도메인 오디오 신호들을 디코딩 및 분리하기 위하여 이용될 수 있는 사이드 정보로서 샘플링 및 양자화된 상기 다수의 시간 도메인 오디오 신호들을 인코딩하는 단계를 포함하는 다수의 시간 도메인 오디오 신호들을 인코딩하는 방법을 수행하게 하는, 다수의 오디오 신호들을 인코딩하기 위한 장치.

#### 청구항 10

제 9 항에 있어서,

상기 랜덤 샘플링은 미리 정의된 의사 랜덤 패턴을 이용하는, 다수의 오디오 신호들을 인코딩하기 위한 장치.

#### 청구항 11

다수의 오디오 신호들의 혼합을 디코딩하기 위한 장치로서,

프로세서, 및 명령들을 저장하는 메모리를 포함하고,

상기 명령들은, 실행될 때, 상기 장치로 하여금

- 상기 다수의 오디오 신호들 각각의 오디오 신호의 양자화된 시간 도메인 샘플들을 포함하는 사이드 정보를 디코딩 및 역다중화하는 단계;

- 저장 또는 입출력 데이터 소스로부터, 상기 다수의 오디오 신호들의 혼합을 수신 또는 취출하는 단계; 및

- 상기 다수의 오디오 신호들을 근사화하는 다수의 추정된 오디오 신호들을 생성하는 단계로서, 상기 다수의 오디오 신호들의 각각의 오디오 신호의 상기 양자화된 샘플들이 이용되는, 상기 다수의 추정된 오디오 신호들을 생성하는 단계를 포함하는 다수의 오디오 신호들을 디코딩하는 방법을 수행하게 하는, 다수의 오디오 신호들의

혼합을 디코딩하기 위한 장치.

## 청구항 12

제 11 항에 있어서,

상기 다수의 추정된 오디오 신호들을 생성하는 단계는

- 랜덤 비음수 값들로부터 분산 텐서  $V$  를 계산하는 단계;
- 상기 다수의 오디오 신호들의 양자화된 샘플들의 상기 소스 전력 스펙트럼의 조건부 기대치를 계산하는 단계로서, 추정된 소스 전력 스펙트럼  $\mathbf{P}(f, n, j)$  이 획득되고 상기 다수의 오디오 신호들의 분산 텐서  $V$  및 복소 단기 푸리에 변환 (STFT) 계수가 이용되는, 상기 조건부 기대치를 계산하는 단계;
- 상기 추정된 소스 전력 스펙트럼  $\mathbf{P}(f, n, j)$  으로부터 상기 분산 텐서  $V$  를 반복적으로 재산출하는 단계;
- 결과적인 분산 텐서  $V$  로부터 STFT 계수들  $\hat{\mathbf{S}}$  의 어레이를 계산하는 단계; 및
- STFT 계수들  $\hat{\mathbf{S}}$  의 어레이를 상기 시간 도메인으로 변환하는 단계로서, 상기 다수의 추정된 오디오 신호들이 획득되는, 상기 시간 도메인으로 변환하는 단계를 포함하는, 다수의 오디오 신호들의 혼합을 디코딩하기 위한 장치.

## 청구항 13

제 11 항 또는 제 12 항에 있어서,

상기 다수의 시간 도메인 오디오 신호들 중 적어도 하나에 대한 오디오 인페인팅 단계를 더 포함하는, 다수의 오디오 신호들의 혼합을 디코딩하기 위한 장치.

### 발명의 설명

#### 기술 분야

[0001] 이 발명은 다수의 오디오 신호들을 인코딩하기 위한 방법 및 디바이스, 및 다수의 오디오 신호들의 개선된 분리로 다수의 오디오 신호들의 혼합을 디코딩하기 위한 방법 및 디바이스에 관한 것이다.

#### 배경 기술

[0002] 오디오 소스 분리의 문제는 개별적인 소스들 (예컨대, 스피치, 음악 악기들, 잡음 등) 을 그들의 혼합 (*mixture*) 들로부터 추정하는 것에 있다. 오디오의 맥락에서, 혼합은 단일 또는 다수의 마이크로폰들에 의한 다수의 소스들의 레코딩 (*recording*) 을 의미한다. 오디오 신호들에 대한 통지된 소스 분리 (*informed source separation; ISS*) 는, 소스들에 대한 일부 정보가 이용가능하다고 하면, 소스들의 혼합으로부터 개별적인 오디오 소스들을 추출하는 문제로서 간주될 수 있다. ISS 는 또한, 이 소스들의 혼합이 인코딩 및 디코딩 스테이지들의 둘 모두에 대해 알려져 있다면, 오디오 객체들 (소스들) 의 압축 [6], 즉, 멀티소스 오디오를 인코딩하는 것에 관한 것이다. 이 문제들의 둘 모두는 상호연결된다. 그들은 폭넓은 범위의 애플리케이션들에 대해 중요하다.

[0003] 알려진 해결책들 (예컨대, [3], [4], [5]) 는 원래의 소스들이 인코딩 스테이지 동안에 이용가능하다는 가정에 의존한다. 사이드 정보 (*side-information*) 가 계산되고, 혼합과 함께 송신되고, 둘 모두는 소스들을 복원하기 위하여 디코딩 스테이지에서 프로세싱된다. 몇몇 ISS 방법들이 알려져 있지만, 모든 이 접근법들에서, 인코딩 스테이지는 디코딩 스테이지보다 더 복잡하고 계산적으로 비용이 많이 듦다. 그러므로, 이 접근법들은 인코딩을 수행하는 플랫폼이 인코더에 의해 요구된 계산적 복잡성을 다룰 수 없는 경우에는 바람직하지 않다. 최종적으로, 알려진 복잡한 인코더들은 일부 애플리케이션들을 위하여 매우 중요한 온라인 인코딩, 즉, 신호가 도착함에 따라 신호를 점진적으로 인코딩하는 것을 위하여 이용가능하지 않다.

#### 발명의 내용

- [0004] 상기를 고려하면, ISS 문제들의 둘 모두에 대한 완전히 자동적이고 효율적인 해결책을 가지는 것이 매우 바람직하다. 특히, 인코더가 디코더보다 상당히 더 적은 프로세싱을 요구하는 해결책이 바람직할 것이다.
- [0005] 본 발명은 프로세싱 부하의 대부분을 인코더 측으로부터 디코더 측으로 시프트하는 단순한 인코딩 방식을 제공한다. 사이드 정보를 생성하기 위한 제안된 단순한 방법은 낮은 복잡성의 인코딩뿐만 아니라, 디코더에서의 효율적인 복원도 가능하게 한다. 최종적으로, (배치 인코딩 (batch encoding) 으로 칭해지는) 인코딩 동안에 전체 신호가 알려질 필요가 있는 일부 현존하는 효율적인 방법들과 대조적으로, 제안된 인코딩 방식은 온라인 인코딩, 즉, 신호가 이것이 도착함에 따라 점진적으로 인코딩되는 것을 허용한다.
- [0006] 인코더는 랜덤 패턴 (random pattern) 으로 오디오 소스들로부터 랜덤 샘플 (random sample) 들을 취한다. 하나의 실시형태에서, 그것은 미리 정의된 의사 랜덤 패턴 (pseudo-random pattern) 이다. 샘플링된 값들은 미리 정의된 양자화기에 의해 양자화되고, 결과적인 양자화된 샘플들은 사이드 정보를 생성하기 위하여 엔트로피 코더 (entropy coder) 에 의해 연접 (concatenate) 되고 무손실로 압축된다. 그혼합은 또한, 인코딩 측에서 생성될 수 있거나, 또는 그것은 디코딩 측에서 다른 방법들을 통해 이미 이용가능하다.
- [0007] 디코더는 사이드 정보로부터 양자화된 샘플들을 먼저 복원하고, 그 다음으로, 양자화된 샘플들 및 혼합이 주어질 경우, 혼합 내에서 확률적으로 가장 가능성 있는 소스들을 추정한다.
- [0008] 하나의 실시형태에서, 본 원리들은 청구항 1 에서 개시된 바와 같은, 다수의 오디오 신호들을 인코딩하기 위한 방법에 관한 것이다. 하나의 실시형태에서, 본 원리들은 청구항 3 에서 개시된 바와 같은, 다수의 오디오 신호의 혼합을 디코딩하기 위한 방법에 관한 것이다.
- [0009] 하나의 실시형태에서, 본 원리들은 이하에서 설명된 바와 같은 인코딩 방법의 각각의 단계에 대하여 하나씩, 복수의 별도의 하드웨어 컴포넌트들을 포함하는 인코딩 디바이스에 관한 것이다. 하나의 실시형태에서, 본 원리들은 이하에서 설명된 바와 같은 디코딩 방법의 각각의 단계에 대하여 하나씩, 복수의 별도의 하드웨어 컴포넌트들을 포함하는 디코딩 디바이스에 관한 것이다.
- [0010] 하나의 실시형태에서, 본 원리들은 컴퓨터로 하여금, 이하에서 설명된 바와 같은 단계들을 포함하는 인코딩 방법을 수행하게 하기 위한 실행가능한 명령들을 가지는 컴퓨터 판독가능 매체에 관한 것이다. 하나의 실시형태에서, 본 원리들은 컴퓨터로 하여금, 이하에서 설명된 바와 같은 단계들을 포함하는 디코딩 방법을 수행하게 하기 위한 실행가능한 명령들을 가지는 컴퓨터 판독가능 매체에 관한 것이다.
- [0011] 하나의 실시형태에서, 본 원리들은 적어도 하나의 하드웨어 컴포넌트, 예컨대, 하드웨어 프로세서, 및 적어도 하나의 소프트웨어 컴포넌트를 유형적으로 구체화하는 비일시적 유형의 컴퓨터 판독가능 저장 매체를 포함하는, 오디오 소스들을 분리하기 위한 인코딩 디바이스에 관한 것이고, 적어도 하나의 하드웨어 프로세서 상에서 실행될 때, 소프트웨어 컴포넌트는 이하에서 설명된 바와 같은 인코딩 방법의 단계들을 일으킨다. 하나의 실시형태에서, 본 원리들은 적어도 하나의 하드웨어 컴포넌트, 예컨대, 하드웨어 프로세서, 및 적어도 하나의 소프트웨어 컴포넌트를 유형적으로 구체화하는 비일시적 유형의 컴퓨터 판독가능 저장 매체를 포함하는, 오디오 소스들을 분리하기 위한 인코딩 디바이스에 관한 것이고, 적어도 하나의 하드웨어 프로세서 상에서 실행될 때, 소프트웨어 컴포넌트는 이하에서 설명된 바와 같은 디코딩 방법의 단계들을 일으킨다.
- [0012] 본 원리들의 추가의 목적들, 특징들, 및 장점들은 동반된 도면들과 관련하여 취해질 때, 다음의 설명 및 첨부된 청구항들의 고려로부터 분명해질 것이다.

### 도면의 간단한 설명

- [0013] 예시적인 실시형태들은 이하에서 도시하는 첨부 도면들을 참조하여 설명되고,  
도 1 은 인코더 및 디코더를 포함하는 송신 및/또는 저장 시스템의 구조이고;  
도 2 는 예시적인 인코더의 단순화된 구조이고;  
도 3 은 예시적인 디코더의 단순화된 구조이고; 그리고  
도 4 는 CS-ISS와 전통적 ISS 사이의 성능 비교이다.

### 발명을 실시하기 위한 구체적인 내용

- [0014] 도 1 은 인코더 및 디코더를 포함하는 송신 및/또는 저장 시스템의 구조를 도시한다. 원래의 사운드 소스들

$s_1, s_2, \dots, s_J$  은 혼합  $x$  및 사이드 정보를 제공하는 인코더로 입력된다. 디코더는 사운드를 복원하기 위하여 혼합  $x$  및 사이드 정보를 이용하고, 여기서, 일부 정보는 순실된 것으로 가정되고; 그러므로, 디코더는 사운드 소스들을 추정할 필요가 있고, 추정된 사운드 소스들  $\tilde{s}_1, \tilde{s}_2, \dots, \tilde{s}_J$  을 제공한다.

[0015] 원래의 소스들  $s_1, s_2, \dots, s_J$  은 인코더에서 이용가능하고, 사이드 정보를 생성하기 위하여 인코더에 의해 프로세싱되는 것으로 가정된다. 혼합은 또한, 인코더에 의해 생성될 수 있거나, 또는 그것은 디코더에서 다른 수단에 의해 이용가능할 수 있다. 예를 들어, 인터넷 상에서 이용가능한 알려진 오디오 트랙에 대하여, 개별적인 소스들로부터 생성된 사이드 정보는 예컨대, 오디오 트랙의 저작권자들 또는 다른 사람들에 의해 저장될 수 있다. 본원에서 설명된 하나의 문제는, 단일 채널 오디오 소스들이 단일 마이크로폰들로 레코딩되며, 이들이 혼합을 형성하기 위하여 함께 추가된다는 것이. 다른 구성들, 예컨대, 다수의 마이크로폰들을 이용한 멀티채널 오디오 또는 레코딩들은 설명된 방법들을 간단한 방식으로 확장함으로써 용이하게 다루어질 수 있다.

[0016] 상기 설명된 설정 내에서 여기에서 고려되는 하나의 기술적 문제는: 사이드 정보를 생성하기 위한 인코더를 가질 때, 원래의 소스들  $s_1, s_2, \dots, s_J$  에 가능한 한 근접한 소스들  $\tilde{s}_1, \tilde{s}_2, \dots, \tilde{s}_J$  을 추정할 수 있는 디코더를 설계하는 것에 있다. 디코더는 추정된 소스들의 주어진 품질에 대한 사이드 정보의 필요한 크기를 최소화하기 위하여, 사이드 정보 및 알려진 혼합  $x$  을 효율적인 방식으로 이용해야 한다. 디코더는 혼합 및 그것이 소스들을 이용하여 어떻게 형성되는지의 둘 모두를 알고 있는 것으로 가정된다.

[0017] 그러므로, 발명은 2 개의 부분들: 인코더 및 디코더를 포함한다.

[0018] 도 2 a) 는 예시적인 인코더의 단순화된 구조를 도시한다. 인코더는 계산적으로 단순한 것으로 설계된다. 그것은 오디오 소스들로부터 랜덤 샘플들을 취한다. 하나의 실시형태에서, 그것은 미리 정의된 의사 랜덤 패턴을 이용한다. 또 다른 실시형태에서, 그것은 임의의 랜덤 패턴을 이용한다. 샘플링된 값들은 (미리 정의된) 양자화기에 의해 양자화되고, 결과적인 양자화된 샘플들  $y_1, y_2, \dots, y_J$  은 사이드 정보를 생성하기 위하여 엔트로피 코더 (예컨대, 허프만 코더 (Huffman coder) 또는 산술 코더 (arithmetic coder)) 에 의해 연접 (concatenate) 되고 무손실로 압축된다. 혼합은 또한, 디코딩 측에서 이미 이용가능하지 않을 경우에 생성된다.

[0019] 도 2 b) 는 인코더 내에서의 확대된 예시적인 신호들을 도시한다. 혼합 신호  $x$  는 상이한 소스 신호들  $s_1, s_2, \dots, s_J$  을 오버레이 (overlay) 하거나 혼합함으로써 획득된다. 소스 신호들  $s_1, s_2, \dots, s_J$  의 각각은 또한, 랜덤 샘플링 단위들로 랜덤 샘플링되고, 샘플들은 양자화된 샘플들  $y_1, y_2, \dots, y_J$  을 획득하기 위하여 하나 이상의 양자화기들 (이 실시형태에서는, 각각의 신호에 대하여 하나의 양자화기) 에서 양자화된다. 양자화된 샘플들은 사이드 정보로서 이용되도록 인코딩된다. 다른 실시형태들에서, 샘플링 및 양자화의 시퀀스 순서는 바뀌어질 수도 있다는 것에 주목한다.

[0020] 도 3 은 예시적인 디코더의 단순화된 구조를 도시한다. 디코더는 사이드 정보로부터 양자화된 샘플들  $y_1, y_2, \dots, y_J$  을 먼저 복원한다. 다음으로, 그것은 관측된 샘플들  $y_1, y_2, \dots, y_J$  및 혼합  $x$  이 주어질 경우에, 그리고 알려진 구조들 및 소스들 사이의 상관들을 활용하여, 확률적으로 가장 가능성 있는 소스들  $\tilde{s}_1, \tilde{s}_2, \dots, \tilde{s}_J$  을 추정한다.

[0021] 인코더의 가능한 구현예들은 매우 단순하다. 디코더의 하나의 가능한 구현예는 다음의 2 개의 가정들에 기초하여 동작한다:

[0022] (1) 소스들은 윈도우 크기  $F$  및 윈도우들의 수  $N$  을 갖는 단시간 푸리에 변환 (Short-Time Fourier Transform; STFT) 도메인에서 공동으로 가우시안 (Gaussian) 분포되어 있다.

[0023] (2) 가우시안 분포의 분산 텐서 (variance tensor)  $V \in \mathbb{R}_+^{F \times N \times J}$  는 이하와 같도록, 랭크  $K$  의 낮은 랭크의 비음수 텐서 분해 (Non-Negative Tensor Decomposition; NTF) 를 가진다.

$$\mathbf{V}(f, n, j) = \sum_{k=1}^K \mathbf{H}(n, k) \mathbf{W}(f, k) \mathbf{Q}(j, k), \quad \mathbf{H} \in \mathbb{R}_+^{N \times K}, \mathbf{W} \in \mathbb{R}_+^{F \times K}, \mathbf{Q} \in \mathbb{R}_+^{J \times K}$$

[0024]

o] 2 개의 가정들을 따르면, 디코더의 동작은 다음의 단계들로 요약될 수 있다:

- [0026] 1. 랜덤 비음수 값들로 행렬들  $\mathbf{H} \in \mathbb{R}_+^{N \times K}, \mathbf{W} \in \mathbb{R}_+^{F \times K}, \mathbf{Q} \in \mathbb{R}_+^{J \times K}$  을 초기화하고, 분산 텐서  $\mathbf{V} \in \mathbb{R}_+^{F \times N \times J}$  를 다음과 같이 계산한다:

$$\mathbf{V}(f, n, j) = \sum_{k=1}^K \mathbf{H}(n, k) \mathbf{W}(f, k) \mathbf{Q}(j, k)$$

[0027]

2. 수렴 또는 최대 수의 반복들에 도달할 때까지, 하기를 반복한다:

2.1 다음과 같도록, 소스 파워 스펙트럼 (source power spectrum) 들의 조건부 기대치들을 계산한다.

$$\mathbf{P}(f, n, j) = \mathbb{E}\{\mathbf{S}(f, n, j)^2 | \mathbf{x}, \mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_J, \mathbf{V}\}$$

[0031] 식중,  $\mathbf{S} \in \mathbb{C}^{F \times N \times J}$  은 소스들의 STFT 복소 계수 (complex coefficient) 들의 어레이이다. 이 조건부 기대치 계산에 대한 더 많은 세부사항들은 아래에 제공된다.

[0032] 2.2 다음과 같도록, 추정된 소스 파워 스펙트럼들  $\mathbf{P}(f, n, j)$  의 3-분산 텐서와, NTF 모델 근사화  $\mathbf{V}(f, n, j)$  의 3-분산 텐서 사이의 IS 발산 [15] 을 최소화하는 증배 업데이트 (multiplicative update; MU) 규칙들을 이용하여 NTF 모델 파라미터들  $\mathbf{H} \in \mathbb{R}_+^{N \times K}, \mathbf{W} \in \mathbb{R}_+^{F \times K}, \mathbf{Q} \in \mathbb{R}_+^{J \times K}$  을 재추정한다:

$$\mathbf{Q}(j, k) \leftarrow \mathbf{Q}(j, k) \left( \frac{\sum_{f,n} \mathbf{W}(f, k) \mathbf{H}(n, k) \mathbf{P}(f, n, j) \mathbf{V}(f, n, j)^{-2}}{\sum_{f,n} \mathbf{W}(f, k) \mathbf{H}(n, k) \mathbf{V}(f, n, j)^{-1}} \right)$$

$$\mathbf{W}(f, k) \leftarrow \mathbf{W}(f, k) \left( \frac{\sum_{j,n} \mathbf{Q}(j, k) \mathbf{H}(n, k) \mathbf{P}(f, n, j) \mathbf{V}(f, n, j)^{-2}}{\sum_{j,n} \mathbf{Q}(j, k) \mathbf{H}(n, k) \mathbf{V}(f, n, j)^{-1}} \right)$$

$$\mathbf{H}(n, k) \leftarrow \mathbf{H}(n, k) \left( \frac{\sum_{f,j} \mathbf{W}(f, k) \mathbf{Q}(j, k) \mathbf{P}(f, n, j) \mathbf{V}(f, n, j)^{-2}}{\sum_{f,j} \mathbf{W}(f, k) \mathbf{Q}(j, k) \mathbf{V}(f, n, j)^{-1}} \right)$$

[0033]

이 업데이트들은 여러번 반복적으로 반복될 수 있다.

[0035] 3. STFT 계수들  $\mathbf{S} \in \mathbb{C}^{F \times N \times J}$  의 어레이를 이하와 같은 사후 평균 (posterior mean) 으로서 계산하고

$$\tilde{\mathbf{S}}(f, n, j) = \mathbb{E}\{\mathbf{S}(f, n, j) | \mathbf{x}, \mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_J, \mathbf{V}\}$$

[0036]

[0037] 추정된 소스들  $\tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2, \dots, \tilde{\mathbf{s}}_J$  을 복원하기 위하여 시간 도메인으로 다시 변환한다. 이 사후 평균에 대한 더 많은 세부사항들은 이하에서 제공된다.

[0038] 다음은 상기 산출들에 대한 일부 수학적인 기초들을 설명한다.

[0039] 텐서는 더 높은 차원의 행렬로서 볼 수 있는 데이터 구조이다. 행렬은 2 차원인 반면, 텐서는 N 차원일 수 있다. 본 경우에 있어서, V 는 (입방체 (cube) 와 같은) 3 차원 텐서이다. 그것은 소스들의 공동 가우시안 분포의 공분산 행렬 (covariance matrix) 을 나타낸다.

- [0040] 행렬은 각각이 낮은 랭크 모델에서 2 개의 벡터들을 승산함으로써 형성되는 몇 개의 랭크-1 행렬들의 합으로서 나타내어질 수 있다. 본 경우에 있어서, 텐서는 K 랭크 1 텐서들의 합으로서 유사하게 나타내어지고, 여기서, 랭크 1 텐서는 3 개의 벡터들, 예컨대,  $h_i$ ,  $q_i$ , 및  $w_i$  을 승산함으로써 형성된다. 이 벡터들은 행렬들  $H$ ,  $Q$ , 및  $W$  를 형성하기 위하여 함께 놓여진다. K 랭크 1 텐서들에 대한 벡터들의 K 세트들이 있다. 본 질적으로, 텐서는 K 컴포넌트들에 의해 나타내어지고, 행렬들  $H$ ,  $Q$ , 및  $W$  는 컴포넌트들이 각각 상이한 프레임들, STFT 의 상이한 주파수들, 및 상이한 소스들을 따라 어떻게 분포되는지를 나타낸다.
- [0041] 행렬들에서의 낮은 랭크 모델과 유사하게, 작은 K 가 오디오 데이터, 예컨대, 음악과 같은 데이터의 특성을 더 양호하게 정의하기 때문에, K 는 작게 유지된다. 이 때문에, V 가 낮은 랭크의 텐서이어야 한다는 정보를 이용함으로써 신호의 알려지지 않은 특성을 추측하는 것이 가능하다. 이것은 미지수 (unknown) 들의 수를 감소시키고 데이터의 상이한 부분들 사이의 상호관련성을 정의한다.
- [0042] 상기 설명된 반복 알고리즘의 단계들은 다음과 같이 설명될 수 있다. 먼저, 행렬들  $H$ ,  $Q$ , 및  $W$  와, 이에 따라,  $V$  를 초기화한다.
- [0043]  $V$  가 주어질 경우, 신호의 확률 분포가 알려진다. 그리고 신호들의 관측된 부분을 보면 (신호들은 오직 부분적으로 관측됨), 예컨대, 위너 필터링 (Wiener filtering) 에 의해 STFT 계수들  $\hat{S}$  을 추정하는 것이 가능하다. 이것은 신호의 사후 평균이다. 또한, 이하에서 이용될 신호의 사후 공분산 (posterior covariance) 이 또한 계산된다. 이 단계는 신호의 각각의 윈도우에 대하여 독립적으로 수행되고, 그것은 병렬화 가능하다. 이것은 예상 단계 또는 E-단계로 칭해진다.
- [0044] 일단 사후 평균 및 공분산이 계산되면, 이것들은 사후 파워 스펙트럼들  $p$  를 계산하기 위하여 이용된다. 이것은 더 이전의 모델 파라미터들, 즉,  $H$ ,  $Q$ , 및  $W$  를 업데이트하기 위하여 필요하게 된다. 더 양호한 추정치에 도달하기 위하여 이 단계를 한 번을 초과하여 반복하는 것이 유리할 수도 있다 (예컨대, 2 내지 10 회). 이것은 최대화 단계 또는 M-단계로 칭해진다.
- [0045] 일단 모델 파라미터들  $H$ ,  $Q$ , 및  $W$  가 업데이트되면, (STFT 계수들  $\hat{S}$  을 추정하는 것으로부터) 모든 단계들은 실시형태들에서, 어떤 수렴에 도달될 때까지 반복될 수 있다. 수렴에 도달된 후, 실시형태에서, STFT 계수들  $\hat{S}$  의 사후 평균은 오디오 신호를 최종적인 결과로서 획득하기 위하여 시간 도메인으로 변환된다.
- [0046] 발명의 하나의 장점은 그것이 그 혼합으로부터의 다수의 오디오 소스 신호들의 개선된 복원을 허용한다는 것이다. 이것은 강력한 디바이스들에 대한 필요성 없이 멀티소스 오디오 레코딩의 효율적인 저장 및 송신을 가능하게 한다. 이동 전화들 또는 태블릿들은 심한 배터리 고갈 또는 프로세서 사용 없이 오디오 트랙의 다수의 소스들에 관한 정보를 압축하기 위하여 용이하게 이용될 수 있다.
- [0047] 추가의 장점은 개별적인 소스들에 대해 압축된 정보가 그것들이 필요하게 될 경우에만 디코딩되므로, 소스들을 인코딩 및 디코딩하기 위한 계산적 자원들이 더 효율적으로 사용된다는 것이다. 음악 제작과 같은 일부 애플리케이션들에서, 개별적인 소스들에 대한 정보는 항상 인코딩되고 저장되지만, 그것이 항상 필요하고 나중에 액세스되는 것은 아니다. 그러므로, 매 인코딩된 오디오 스트림에 대해 높은 복잡성의 프로세싱을 수행하는 고가의 인코더와 대조적으로, 낮은 복잡성의 인코더 및 높은 복잡성의 디코더를 갖는 시스템은 개별적인 소스들이 더 이후에 실제로 필요하게 되는 그 오디오 스트림들에 대해서만 프로세싱 파워를 사용하는 이익을 가진다.
- [0048] 발명에 의해 제공된 제 3 장점은 새롭고 더 양호한 디코딩 방법들에 대한 적응가능성이다. 데이터 내의 상관들을 활용하는 새롭고 개선된 방법이 탐색될 때, 디코딩하기 위한 새로운 방법이 고안될 수 있고 ( $\mathbf{x}, \mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_J$  이 주어질 경우에,  $\tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2, \dots, \tilde{\mathbf{s}}_J$  을 추정하기 위한 더 양호한 방법), 소스들을 재인코딩 할 필요 없이 더 양호한 품질로 더 오래된 인코딩된 비트스트림들을 디코딩하는 것이 가능하다. 반면에 전통적인 인코딩-디코딩 패러다임들에서는, 데이터 내의 상관들을 활용하는 개선된 방법이 새로운 인코딩 방법으로 이어질 때, 새로운 접근법의 장점을 활용하기 위하여 소스들을 디코딩 및 재인코딩할 필요가 있다. 또한, 이미 인코딩된 비트스트림을 재인코딩하는 프로세스는 원래의 소스들에 대하여 추가의 에러들을 도입하는 것으로 알려져 있다.
- [0049] 발명의 제 4 장점은 소스들을 온라인 방식으로 인코딩하기 위한 가능성이 있고, 즉, 소스들은 그것들이 인코더에

도착함에 따라 인코딩되고, 전체 스트림의 이용가능성은 인코딩을 위해 필요하지 않다.

[0050] 발명의 제 5 장점은 분리된 오디오 소스 신호들에서의 캡들이 바로잡아질 수 있다는 것이고, 이것은 오디오 인페인팅 (audio inpainting) 으로서 알려져 있다. 이에 따라, 발명은 다음에서 설명된 바와 같이, 공동 오디오 인페인팅 및 소스 분리를 허용한다.

[0051] 본원에서 개시된 접근법은 분산된 소스 코딩 (distributed source coding) [9] 및 특히, 분산된 비디오 코딩 (distributed video coding) [10] 패러다임들에 의해 영감 받았고, 여기서, 목표는 또한, 복잡성을 인코더로부터 디코더로 시프트하기 위한 것이다. 접근법은 압축 센싱/샘플링 원리들 [11 - 13] 에 의존하는데, 소스들이 오디오 소스들이 희소 (sparse) 한 기반 (basis) 과 비일관적 (incoherent) 인 기반 [13] 의 벡터들의 랜덤으로 선택된 서브세트에 의해 걸쳐 이어진 선형 서브공간 상에서 투영되기 때문이다. 개시된 접근법은 압축 샘플링 기반 ISS (*compressive sampling-based ISS; CS-ISS*) 로 칭해질 수 있다.

[0052] 더욱 구체적으로, 소스들의 시간적 샘플들의 서브세트의 단순한 랜덤 선택, 그 다음으로, 균일한 양자화 및 엔트로피 인코더에 의해 소스들을 인코딩하는 것이 제안된다. 하나의 실시형태에서, 이것은 디코더로 송신되는 유일한 사이드 정보이다.

[0053] 시간 도메인에서의 샘플링의 장점은 2 배인 것에 주목한다. 첫 째, 그것은 임의의 변환된 도메인에서의 샘플링보다 더 빠르다. 둘 째, 시간적 기반은 오디오 신호들이 희소한 단시간 푸리에 변환 (STFT) 프레임과 충분히 비일관적이고, 그것은 STFT 계수들의 낮은 랭크의 NTF 표현과 훨씬 더 비일관적이다. 측정 및 이전 정보 도메인들의 비일관성은 소스들의 복원을 위하여 필수적이라는 것이 압축 센싱 이론에서 나타나 있다 [13].

[0054] 양자화된 소스 샘플들 및 혼합으로부터 디코더에서 소스들을 복원하기 위하여, 모델 기반 압축 센싱 [14] 과 일치하는 모델 기반 접근법을 이용하는 것이 제안된다. 특히, 하나의 실시형태에서, 소스 스펙트로그램 (source spectrogram) 들의 이타쿠라-사이토 (Itakura-Saito; IS) 비음수 텐서 인수화 (nonnegative tensor factorization; NTF) 모델이 [4, 5] 에서와 같이 이용된다. 그 가우시안 확률적 공식 [15] 때문에, 이 모델은 소스 샘플들의 송신된 양자화된 부분 및 혼합으로부터 최대 우도 (maximum-likelihood; ML) 의 의미에서 추정될 수도 있다. 모델을 추정하기 위하여, 증배 업데이트 (MU) 규칙들 [15] 에 기초한 새로운 일반화된 기대치 최대화 (generalized expectation-maximization; GEM) 알고리즘 [16] 이 이용될 수 있다. 추정된 모델 및 모든 다른 관측들이 주어질 경우, 소스들은 위너 필터링 [17] 에 의해 추정될 수 있다.

#### CS-ISS 프레임워크의 개관

[0055] 제안된 CS-ISS 인코더/디코더의 전체적인 구조는 위에서 이미 설명된 바와 같이, 도 2 에서 도시되어 있다. 인코더는 미리 정의된 랜덤화 패턴을 이용하여 희망하는 레이트로 소스들을 랜덤으로 서브샘플링하고, 이 샘플들을 양자화한다. 다음으로, 양자화된 샘플들은 최종적인 인코딩된 비트스트림을 형성하기 위하여 엔트로피 인코더로 압축될 단일 스트림에서 순서화된다. 랜덤 샘플링 패턴 (또는 랜덤 패턴을 생성하는 시드 (seed)) 은 인코더 및 디코더의 둘 모두에 의해 알려져 있고, 그러므로, 하나의 실시형태에서 송신될 필요가 없다. 또 다른 실시형태에서, 랜덤 샘플링 패턴, 또는 랜덤 패턴을 생성하는 시드는 디코더로 송신된다. 오디오 혼합은 또한, 디코더에 의해 알려진 것으로 가정된다. 디코더는 소스들의 양자화된 샘플들을 취출하기 위한 엔트로피 디코딩과, 그 다음으로, 이하에서 상세하게 논의될 CS-ISS 디코딩을 수행한다.

[0056] 제안된 CS-ISS 프레임워크는 전통적인 ISS 에 비해 몇몇 장점들을 가지고, 이것들은 다음과 같이 요약될 수 있다:

[0057] 제 1 장점은 도 2 에서의 단순한 인코더가 예컨대, 저전력 디바이스들에서, 필요한 바와 같이 낮은 복잡성의 인코딩을 위하여 이용될 수 있다는 것이다. 낮은 복잡성의 인코딩 방식은 또한, 인코딩이 자주 이용되지만, 오직 소수의 인코딩된 스트림들이 디코딩될 필요가 있는 애플리케이션들에 대하여 유리하다. 이러한 애플리케이션의 예는, 각각의 제작된 음악의 소스들이 미래의 이용을 위하여 유지되지만, 거의 필요하지 않은 스튜디오에서의 음악 제작이다. 이 때문에, 프로세싱 파워 및 프로세싱 시간의 측면에서의 상당한 절감들이 CS-ISS 로 가능하다.

[0058] 제 2 장점은 시간 도메인에서 (그리고 변환된 도메인에서가 아님) 샘플링을 수행하는 것이 단순한 샘플링 방식 뿐만 아니라, 필요할 때, 온라인 방식으로 인코딩을 수행하기 위한 가능성을 제공하는 것이며, 이는 다른 방법들 [4, 5] 에 대해 항상 간단한 것은 아니다. 또한, 독립적인 인코딩 방식은 디코딩 효율을 양보하지 않으면서, 분산된 방식으로 소스들을 인코딩 할 가능성을 가능하게 한다.

[0060] 제 3 장점은 인코딩 단계가 디코딩 단계에 대한 임의의 가정을 없이 수행된다는 것이다. 그러므로, 이 실시 형태에서 제안된 것과는 다른 디코더들을 이용하는 것이 가능하다. 이것은 더 양호하게 수행하는 디코더가 설계될 때, 인코딩된 소스들이 재인코딩의 필요 없이 개선된 디코딩으로부터 직접으로 이익을 얻을 수 있다는 의미에서, 전통적인 ISS [2 - 5] 에 비해 현저한 장점을 제공한다. 이것은 인코더에서 이용된 랜덤 샘플링에 의해 가능해진다. 압축 센싱 이론은 랜덤 샘플링 방식이 많은 수의 도메인들과의 비일관성을 제공하여, 데이터에 대한 상이한 이전 정보에 의존하는 효율적인 디코더들을 설계하는 것이 가능하게 된다는 것을 보여준다.

[0061] CS-ISS 디코더

[0062]  $\Omega''$  을 갖는 랜덤 샘플들의 지원을 표시하여, 소스  $j \in \llbracket 1, J \rrbracket$  가 시간 인덱스들  $t \in \Omega_j'' \subseteq \llbracket 1, T \rrbracket$ 에서 샘플링된다고 한다. 엔트로피 디코딩 스테이지 후에, CS-ISS 디코더는 소스들의 양자화된 샘플들의 서브세트  $y_{jt}''(\Omega_j''), j \in \llbracket 1, J \rrbracket$  를 가지고, 여기서, 양자화된 샘플들은 다음과 같이 정의되고,

$$y_{jt}'' = s_{jt}'' + b_{jt}'' \quad (1)$$

[0063] [0064] 식중,  $s_{jt}''$  은 참 (true) 소스 신호를 표시하고,  $b_{jt}''$  은 양자화 잡음이다.

[0065] 본원에서, 시간 도메인 신호들은 2 개의 프라임 (prime) 들을 갖는 문자, 예컨대, x'' 에 의해 나타내어지는 한편, 프레임화되고 원도우화된 시간 도메인 신호들은 하나의 프라임을 갖는 문자, 예컨대, x' 에 의해 나타내어지고, 복소 값인 단시간 푸리에 변환 (STFT) 계수들은 프라임을 갖지 않는 문자, 예컨대, x 에 의해 나타내어진다는 것에 주목한다.

[0066] 혼합은 이하와 같도록, 원래의 소스들의 합인 것으로 가정된다.

$$x_t'' = \sum_{j=1}^J s_{jt}'' , \quad t \in \llbracket 1, T \rrbracket, j \in \llbracket 1, J \rrbracket \quad (2)$$

[0067] 혼합은 디코더에서 알려진 것으로 가정된다. 혼합은 본원에서 잡음이 없고 양자화 없는 것으로 가정된다는 것에 주목한다. 그러나, 개시된 알고리즘은 마찬가지로, 혼합에서 잡음을 포함하도록 용이하게 확장될 수 있다.

[0068] STFT 계수들을 계산하기 위하여, 혼합 및 소스들은 윈도우 길이 M 및 총 N 개의 윈도우들을 갖는 윈도우화된 시간 도메인으로 먼저 변환된다.  $y'_{jmn}, s'_{jmn}$ , 및  $x'_{mn}$  에 의해 나타내어진 결과적인 계수들은  $j = 1, \dots, J, n = 1, \dots, N$ , 및  $m = 1, \dots, M$ 에 대하여 (양자화된 소스 샘플들의 경우에는 적절한 서브세트  $\Omega_{jn}'$ 에서의 오직 m에 대하여) 윈도우화된 시간 도메인에서 양자화된 소스들, 원래의 소스들, 및 혼합을 각각 나타낸다.

[0069] 소스들  $s_{jfn}$  및 혼합  $x_{fn}$ 의 STFT 계수들은 단위 푸리에 변환  $U \in \mathbb{C}^{F \times M}$  ( $F=M$ ) 을 윈도우화된 시간 도메인 대응부들의 각각의 윈도우에 적용함으로써 계산된다. 예를 들어,  $[x_{1n}, \dots, x_{Fn}]^T = U[x'_{1n}, \dots, x'_{Mn}]^T$  이다.

[0070] 소스들은 정규 분포  $(s_{jfn} \sim N_c(0, v_{jfn}))$  를 갖는 STFT 도메인에서 모델링되고, 여기서, 분산 텐서 V =  $[v_{jfn}]_{j,f,n}$  는 다음의 낮은 랭크의 NTF 구조 [18]를 가진다:

$$v_{jfn} = \sum_{k=1}^K q_{jk} w_{fk} h_{nk}, K < \max(J, F, N) \quad (3)$$

[0071] 모델은  $\Theta = \{Q, W, H\}$ 에 의해 파라미터화되며 여기서  $Q = [q_{jk}] \in \mathbb{R}_+^{J \times K}, W = [w_{fk}] \in \mathbb{R}_+^{F \times K}$ , 및  $H = [h_{nk}] \in \mathbb{R}_+^{N \times K}$  이다.

[0073] 본 원리들의 실시형태에 따르면, 소스 신호들은 알고리즘 1에서 간략하게 설명되는 일반화된 기대치 최대화 알고리즘으로 복원된다. 알고리즘은 예상 단계에서의 위너 필터링을 통해 주어진 모델  $\Theta$  을 이용하여 관측들로부터 소스들 및 소스 통계들을 추정하고, 그 다음으로, 최대화 단계에서의 사후 소스 통계들을 이용하여 모델을 업데이트한다. 알고리즘의 각각의 단계에 대한 세부사항들은 아래에 주어진다.

---

### 알고리즘 1 NTF 모델을 이용한 CS-ISS 디코딩을 위한 GEM 알고리즘

---

- 1: **프로시저**  $(\mathbf{x}', \{\mathbf{y}'_j\}_1^J, \{\Omega'_j\}_1^J, K)$  를 CS-ISS 디코딩
  - 2:     비음의  $\mathbf{Q}, \mathbf{W}, \mathbf{H}$  를 랜덤하게 초기화
  - 3:     **반복**
  - 4:          $\mathbf{Q}, \mathbf{W}, \mathbf{H}, \mathbf{x}', \{\mathbf{y}'_j\}_1^J, \{\Omega'_j\}_1^J$  이 주어지면  $\hat{\mathbf{s}}$  (소스)  
 $\hat{\mathbf{P}}$  (사후 전력 스펙트럼)를 추정  $\triangleright$  E-단계, 섹션 3.1 참조
  - 5:          $\hat{\mathbf{P}}$ 이 주어지면  $\mathbf{Q}, \mathbf{W}, \mathbf{H}$  를 업데이트  $\triangleright$  M-단계, 섹션 3.2 참조
  - 6:         수렴 기준 충족될 때까지
  - 7: **프로시저 종료**
- 

[0074]

[0075]

소스들의 추정

[0076]

모든 기저 분포들은 가우시안이고 소스들과 관측들 사이의 모든 관련성들은 선형이므로, 모델 파라미터들  $\mathbf{Q}, \mathbf{W}, \mathbf{H}$  에 의해 (3)에서 정의된 공분산 텐서  $\mathbf{V}$  가 주어질 경우, 소스들은 위너 필터 [17]를 통해 최소 평균 제곱 에러 (minimum mean square error; MMSE)의 의미에서 추정될 수도 있다.

[0077]

$n$  번째 프레임에 대한 관측된 데이터 벡터  $\bar{o}'_n$  를

$$\bar{o}'_n \triangleq [\bar{y}'_{1n}^T, \dots, \bar{y}'_{jn}^T, \bar{x}'_n^T]^T \quad \text{로서} \quad \text{정의하며,} \quad \text{여기서} \quad \bar{x}'_n \triangleq [x'_{1n}, \dots, x'_{Mn}]^T \quad \text{및}$$

$$\bar{y}'_{jn} \triangleq [y'_{mn}, m \in \Omega'_{jn}]^T \quad \text{이다.}$$

[0079]

대응하는 관측된 데이터  $\bar{o}'_n$  및 NTF 모델  $\Theta$  이 주어질 경우, 각각의 소스 프레임  $s_{jn}$  의 사후 분포는  $s_{jn} | \bar{o}'_n; \Theta \sim N_c(\hat{s}_{jn}, \hat{\Sigma}_{s_{jn}s_{jn}})$  으로 기재될 수 있고, 여기서  $\hat{s}_{jn}$  및  $\hat{\Sigma}_{s_{jn}s_{jn}}$  은 각각 사후 평균 및 사후 공분산 행렬이다. 그것들의 각각은 다음과 같이 위너 필터링에 의해 계산될 수 있으며

$$\hat{s}_{jn} = \Sigma_{\bar{o}'_n s_{jn}}^H \Sigma_{\bar{o}'_n \bar{o}'_n}^{-1} \bar{o}'_n, \quad (4)$$

$$\hat{\Sigma}_{s_{jn}s_{jn}} = \Sigma_{s_{jn}s_{jn}} - \Sigma_{\bar{o}'_n s_{jn}}^H \Sigma_{\bar{o}'_n \bar{o}'_n}^{-1} \Sigma_{\bar{o}'_n s_{jn}}, \quad (5)$$

[0080]

[0081] 정의들이 주어질 경우,

$$\Sigma_{\bar{o}'_n \bar{o}'_n} = \begin{bmatrix} \Sigma_{\bar{y}'_{1n} \bar{y}'_{1n}} & \cdots & \mathbf{0} & \Sigma_{\bar{x}'_n \bar{y}'_{1n}}^H \\ \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \cdots & \Sigma_{\bar{y}'_{Jn} \bar{y}'_{Jn}} & \Sigma_{\bar{x}'_n \bar{y}'_{Jn}}^H \\ \Sigma_{\bar{x}'_n \bar{y}'_{1n}} & \cdots & \Sigma_{\bar{x}'_n \bar{y}'_{Jn}} & \Sigma_{\bar{x}'_n \bar{x}'_n} \end{bmatrix}, \quad (6)$$

$$\Sigma_{\bar{o}'_n s_{jn}} = \left[ \mathbf{0}_{S_1 \times F}^T, \Sigma_{\bar{y}'_{jn} s_{jn}}^T, \mathbf{0}_{S_2 \times F}^T, \Sigma_{\bar{x}'_n s_{jn}}^T \right]^T, \\ S_1 \triangleq \sum_{j=1}^{j-1} |\Omega'_{jn}|, \quad S_2 \triangleq \sum_{j=j+1}^J |\Omega'_{jn}|, \quad (7)$$

$$\Sigma_{s_{jn} s_{jn}} = \text{diag} \left( [v_{jfn}]_f \right), \quad (8)$$

$$\Sigma_{\bar{y}'_{jn} \bar{y}'_{jn}} = \mathbf{U}(\Omega'_{jn})^H \text{diag} \left( [v_{jfn}]_f \right) \mathbf{U}(\Omega'_{jn}), \quad (9)$$

$$\Sigma_{\bar{y}'_{jn} s_{jn}} = \mathbf{U}(\Omega'_{jn})^H \text{diag} \left( [v_{jfn}]_f \right), \quad (10)$$

$$\Sigma_{\bar{x}'_n s_{jn}} = \mathbf{U}^H \text{diag} \left( [v_{jfn}]_f \right), \quad (11)$$

$$\Sigma_{\bar{x}'_n \bar{x}'_n} = \mathbf{U}^H \text{diag} \left( \left[ \sum_j v_{jfn} \right]_f \right) \mathbf{U}, \quad (12)$$

$$\Sigma_{\bar{x}'_n \bar{y}'_{jn}} = \mathbf{U}^H \text{diag} \left( [v_{jfn}]_f \right) \mathbf{U}(\Omega'_{jn}), \quad (13)$$

[0082] 여기서,  $\mathbf{U}(\Omega'_{jn})$  은  $\Omega'_{jn}$  에서의 인덱스를 갖는 U로부터의 열들의  $F \times |\Omega'_{jn}|$  행렬이다. 그러므로, 이

하에서 설명된 바와 같은 NTF 모델을 업데이트하기 위하여 이용될 사후 파워 스펙트럼들  $\hat{P} = [\hat{p}_{jfn}]$  은 다음과 같이 계산될 수 있다

$$\hat{p}_{jfn} = \mathbb{E} [|s_{jfn}|^2 | \bar{o}'_n; \theta] = |\hat{s}_{jfn}|^2 + \hat{\Sigma}_{s_{jn} s_{jn}}(f, f). \quad (14)$$

[0085] 모델의 업데이트

[0086] NTF 모델 파라미터들은  $D_{IS}(\tilde{P} || \mathbf{V}) = \sum_{j,f,n} d_{IS}(\tilde{p}_{jfn} || v_{jfn})$  으로서 정의된 NTF 모델 근사화 V의 3-분산 텐서와 추정된 소스 파워 스펙트럼들  $\tilde{P}$  의 3-분산 텐서 사이의 IS 발산 [15] 을 최소화하는 증배 업데이트 (MU) 규칙들을 이용하여 재추정될 수 있고, 여기서,

$$d_{IS}(x || y) = \frac{x}{y} - \log \left( \frac{x}{y} \right) - 1 \quad \text{은 IS 발산이고; } \tilde{p}_{jfn} \text{ 및 } v_{jfn} \text{ 은 (14) 및 (3) 에 의해 특정된다.}$$

그 결과로서, Q, W, H 는 [18] 에서 제시된 MU 규칙들로 업데이트될 수 있다. 이 MU 규칙들은 모델 추정을 개선시키기 위하여 몇 회 반복될 수 있다.

[0088] 또한, NTF/NMF 모델을 이용하는 소스 분리 애플리케이션들에서는, 개별적인 소스들에 대한 일부 이전의 정보를 가지는 것이 종종 필요하다. 이 정보는 소스들로부터의 일부 샘플들, 또는 소스가 시간의 어느 순간에 "비활성" 인지에 대한 지식일 수 있다. 그러나, 이러한 정보가 시행 (enforce) 되어야 할 때, 알고리즘들이 각

각의 소스가 얼마나 많은 컴포넌트들로 구성되는지를 사전 정의할 필요가 있는 것은 항상 그래 왔다. 이것 은,  $\mathbf{Q}$  및  $\mathbf{H}$ 의 어떤 부분들이 제로로 설정되고 각각의 컴포넌트는 특정 소스에 배정 모델 파라미터들  $\mathbf{W} \in \mathbb{R}_+^{M \times K}$ ,  $\mathbf{H} \in \mathbb{R}_+^{N \times K}$ ,  $\mathbf{Q} \in \mathbb{R}_+^{J \times K}$  을 초기화함으로써 종종 시행된다. 하나의 실시형태에서, 모델의 계산은 컴포넌트들 K의 총 수가 주어질 경우, 각각의 소스가 수동이 아니라 자동으로 컴포넌트들에 배정되도록 수정된다. 이것은, STFT 도메인 모델 파라미터들을 통하는 것이 아니라, (제로들의 시간 도메인 샘플들을 가지기 위한 제약을 갖는) 시간 도메인 샘플들을 통해 소스들의 "무음 (silence)" 을 시행함으로써, 그리고 그것들이 자동으로 조절되도록 모델 파라미터들에 대한 초기 조건들을 완화함으로써 달성된다. (Q에 의해 정의된) 소스 컴포넌트 분포 상에서 희소한 구조를 시행하기 위한 추가의 수정은 또한, 상기 증배 업데이트 수학식들을 약간 수정함으로써 가능하다. 이것은 컴포넌트들로의 소스들의 자동 배정으로 귀착된다.

[0089]

이에 따라, 하나의 실시형태에서, 행렬들 H 및 Q는 소스들의 무음 기간 (silence period) 들의 형태의 사이드 정보  $I_s$  가 존재할 때에 자동으로 결정된다. 사이드 정보  $I_s$  는 어느 소스가 어느 시간 기간들에서 무음인지의 정보를 포함할 수도 있다. 이러한 특정 정보의 존재 시에, NMF 를 사용하기 위한 전통적인 방법은 미리 정의된  $k_i$  컴포넌트들이 각각의 소스에 배정되는 그러한 방법으로 H 및 Q를 초기화하는 것이다. 개선된 해결책은 이러한 초기화에 대한 필요성을 제거하고,  $k_i$  가 미리 알려질 필요가 없도록 H 및 Q를 학습한다. 이것은 1) 시간 도메인 샘플들을 입력으로서 이용하여, STFT 도메인 조작이 의무적이지 않도록 함으로써, 그리고 2) 희소한 구조를 갖도록 행렬 Q를 제약함으로써 가능해진다. 이것은 위에서 설명된 바와 같이, Q에 대한 증배 업데이트 수학식들을 수정함으로서 달성된다.

[0090]

## 결과들

[0091]

본 접근법의 성능을 평가하기 위하여, 16 kHz 에서의 음악 신호의 3 개의 소스들이, 양자화의 상이한 레벨들 (16 비트들, 11 비트들, 6 비트들, 및 1 비트) 및 소스 당 상이한 샘플링 비트레이트들 (0.64, 1.28, 2.56, 5.12, 및 10.24 kbps/소스) 로 제안된 CS-ISS 를 이용하, 인코딩 및 다음으로 디코딩된다. 이 예에서는, 랜덤 샘플링 패턴이 미리 정의되고 인코딩 및 디코딩의 둘 모두 동안에 알려지는 것으로 가정된다. 양자화된 샘플들은, 제로 평균 가우시안 분포 가정을 갖는 산술 인코더를 이용하여 절단 (truncate) 및 압축된다. 디코더 측에서는, 산술 디코더를 따르면, 소스들은 가우시안 윈도우 함수 및 K = 18 에서 고정된 컴포넌트들의 수, 즉, 소스 당 6 개의 컴포넌트들을 갖는 1024 샘플들 (64 ms) 의 반중첩 사인 윈도우 (half-overlapping sine window) 를 이용하여 계산된 STFT 를 갖는 GEM 알고리즘의 50 개의 반복들을 이용하여 양자화된 샘플들로부터 디코딩된다. 재구성된 샘플들의 품질은 [19] 에서 설명된 바와 같은 신호 대 왜곡 비율 (signal to distortion ratio; SDR) 로 측정된다. 결과적인 인코딩된 비트레이트들 및 디코딩된 신호들의 SDR 은 괄호로 인코딩된 샘플들의 백분율과 함께 표 1 에서 제시된다. 예상되는 엔트로피 코딩 스테이지의 가변적인 성능으로 인해, 표 1 에서의 압축된 레이트들은 대응하는 로우 비트레이트 (raw bitrate) 들과는 상이하다는 것에 주목한다.

샘플당 비트	압축 레이트 / SDR(유지된 샘플들 중 %)				
	로우 레이트 (kbps/소스)				
	0.64	1.28	2.56	5.12	10.24
16 비트	0.50 / -1.64 (0.25)	1.00 / 4.28 (0.50)	2.00 / 9.54 (1.00)	4.01 / 16.17 (2.00)	8.00 / 21.87 (4.00)
11 비트	0.43 / 1.30 (0.36)	0.87 / 6.54 (0.73)	1.75 / 13.30 (1.45)	3.50 / 19.47 (2.91)	7.00 / 24.66 (5.82)
6 비트	0.27 / 4.17 (0.67)	0.54 / 7.62 (1.33)	1.08 / 12.09 (2.67)	2.18 / 14.55 (5.33)	4.37 / 16.55 (10.67)
1비트	0.64 / -5.06 (4.00)	1.28 / -2.57 (8.00)	2.56 / 1.08 (16.00)	5.12 / 1.59 (32.00)	10.24 / 1.56 (64.00)

[0092]

표 1: 엔트로피 코딩 전의 상이한 (균일한) 양자화 레벨들 및 상이한 로우 비트레이트들에 대한 (dB 단위의) 대응하는 SDR 을 갖는 CS-ISS 의 엔트로피 코딩 스테이지 후의 (소스 당 kbps 인) 최종적인 비트레이트들. 유지된 샘플들의 백분율은 또한, 괄호로 각각의 경우에 대하여 제공된다. 최상의 레이트-왜곡 절충에 대응하는 결과들은 굵은 글씨체이다.

[0094]

CS-ISS 의 성능은 [4] 에서의 더 복잡한 인코더 및 더 단순한 디코더를 갖는 전형적인 ISS 접근법과 비교된다. ISS 알고리즘은 [5] 에서의 NTF 모델 양자화 및 인코딩과 함께 이용되고, 즉, NTF 계수들은 로그 도메인에서 균일하게 양자화되고, 상이한 NTF 행렬들의 양자화 스텝 크기들은 [5]로부터의 수학식들 (31) 내지 (33) 을 이용하여 계산되고, 인덱스들은 2 상태들 가우시안 혼합 모델 (Gaussian mixture model; GMM) ([5] 의 도 5 참조) 에 기초하여 산술 코더를 이용하여 인코딩된다. 접근법은 상이한 양자화 스텝 크기들 및 NTF 컴포넌트들의 상이한 수들, 즉,  $\Delta = 2^{-2}, 2^{-1.5}, 2^{-1}, \dots, 2^4$  및 K = 4, 6, ..., 30 에 대하여 평가된다. 결과들은 모델 업데이트의 250 개의 반복들로 생성된다. CS-ISS 및 전통적인 ISS 의 둘 모두의 성능은 도 4 에서 도시

되어 있고, ISS 접근법이 고정된 수의 컴포넌트들을 이용하는 우리의 디코더 (인코더는 매우 단순하고 이 값을 계산하지 않음) 와 대조적으로, 최적화된 수의 컴포넌트들 및 양자화를 이용할 수 있더라도, CS-ISS 는 ISS 접근법을 명확하게 능가한다. 성능 차이는 랜덤 샘플링된 시간 도메인 및 낮은 랭크의 NTF 도메인의 비일관성 때문에 CS-ISS 디코더에 의해 달성된 높은 효율로 인한 것이다. 또한, ISS 접근법은 [5]에서 설명된 바와 같은 인코더 구조에서의 충실도의 결여로 인해 10 dB 의 SDR 을 넘어 수행할 수 없다. 시간 제약들로 인해 이 문서에서의 [5]에서 제시된 ISS 알고리즘과 비교하는 것이 가능하지 않았더라도, 결과들은 레이트 왜곡 성능이 유사한 거동을 나타낸다는 것을 나타낸다. 제안된 접근법은 그 낮은 복잡성의 인코더에 의해 뛰어나고, 이 때문에, 더 양호한 레이트 왜곡 성능을 갖는 다른 ISS 접근법들에 대해 여전히 유리할 수 있다는 것이 상기되어야 한다.

[0095] 표 1 및 도 4에서의 CS-ISS 의 성능은 상이한 레벨의 양자화들이 상이한 레이트들에서 바람직할 수도 있다는 것을 나타낸다. 16 비트들도 1 비트 양자화도 양호하게 수행하는 것으로 보이지 않더라도, 성능은 훨씬 더 높은 비트레이트가 이용가능할 때, 16 비트들 양자화가 다른 방식들보다 우수할 수도 있다는 것을 나타낸다.

1 비트와 같은 유사하게 더 대략적인 양자화 (coarser quantization)는 현저히 낮은 비트레이트들을 고려할 때에 유익할 수도 있다. 양자화의 선택은 참조로서 단순한 루프 테이블로 인코더에서 수행될 수 있다. 당업자는 또한, CS-ISS에서의 인코더가 매우 단순하더라도, 제안된 디코더가, 전통적인 ISS 방법들의 인코더들 보다 통상적으로 더 높은, 상당히 높은 복잡성이라는 것에 주목해야 한다. 그러나, 이것은 또한, 예컨대, 그래픽 프로세싱 유닛 (graphical processing unit; GPU)들을 이용하는 병렬 프로세싱을 갖는 제안된 디코더에서 프레임들 사이에서의 위너 필터링의 독립성을 활용함으로써 극복될 수 있다.

[0096] 개시된 해결책은 낮은 랭크의 텐서 구조가 재구성된 신호들의 파워 스펙트로그램에서 나타난다는 사실로 보통 이어진다.

[0097] 동사 "포함한다" 및 그 활용형들의 이용은 청구항에서 기재된 것들 이외의 엘리먼트들 또는 단계들의 존재를 배제하지 않는다는 것이 주목되어야 한다. 또한, 엘리먼트를 선행하는 관사 "a" 또는 "an" 은 복수의 이러한 엘리먼트들의 존재를 배제하지 않는다. 몇몇 "수단"은 하드웨어의 동일한 항목에 의해 나타내어질 수도 있다. 또한, 발명은 각각 및 매 신규한 특징 또는 특징들의 조합에서 존재한다. 본원에서 이용된 바와 같이, "디지털 오디오 신호" 또는 "오디오 신호"는 단지 수학적 추상화를 설명하는 것이 아니라, 그 대신에, 머신 또는 장치에 의해 검출할 수 있는 물리적 매체 내에 구체화되거나 물리적 매체에 의해 반송된 정보를 나타낸다. 이 용어는 레코딩되거나 송신된 신호들을 포함하고, 퍼스 코드 변조 (pulse code modulation; PCM)를 포함하지만, PCM 으로 제한되지 않는 임의의 형태의 인코딩에 의한 전달을 포함하는 것으로 이해되어야 한다.

[0098] 그 바람직한 실시형태들에 적용된 것으로서 본 발명의 기본적인 신규한 특징들이 도시되었고, 지적되었지만, 개시된 디바이스들의 형태 및 세부사항들에서의, 설명된 장치 및 방법에서의 다양한 생략들 및 치환들 및 변경들은 본 발명의 사상으로부터 이탈하지 않으면서 당해 분야의 당업자들에 의해 행해질 수도 있다는 것이 이해될 것이다. 동일한 결과들을 달성하기 위하여 실질적으로 동일한 방법으로 실질적으로 동일한 기능을 수행하는 그 엘리먼트들의 모든 조합들은 발명의 범위 내에 있다는 것이 명백히 의도된다. 하나의 설명된 실시형태로부터 또 다른 것으로의 엘리먼트들의 치환들이 또한 완전히 의도되고 고려된다. 설명 및 (적절할 경우에) 청구항들 및 도면들에서 개시된 각각의 특징은 독립적으로, 또는 임의의 적절한 조합으로 제공될 수도 있다. 특징들은 적절한 경우, 하드웨어, 소프트웨어, 또는 이 둘의 조합으로 구현될 수도 있다.

접속들은 적용가능할 경우, 무선 접속들, 또는 유선의, 반드시 직접 또는 전용은 아닌 접속들로서 구현될 수도 있다.

[0099] 인용된 참조문헌들

[0100] [1] E. Vincent, S. Araki, F. J. Theis, G. Nolte, P. Bofill, H. Sawada, A. Ozerov, B. V. Gowreesunker, D. Lutter, 및 N. Q. K. Duong, "The signal separation evaluation campaign (2007-2010): Achievements and remaining challenges", *Signal Processing*, vol. 92, no. 8, pp. 1928-1936, 2012.

[0101] [2] M. Parvaix, L. Girin, 및 J.-M. Brossier, "A watermarkingbased method for informed source separation of audio signals with a single sensor", *IEEE Trans. Audio, Speech, Language Process.*, vol. 18, no. 6, pp. 1464-1475, 2010.

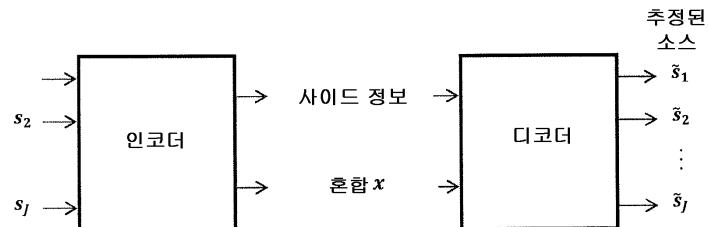
[0102] [3] M. Parvaix 및 L. Girin, "Informed source separation of linear instantaneous under-determined audio mixtures by source index embedding", *IEEE Trans. Audio, Speech, Language Process.*, vol. 19, no. 6, pp.

1721-1733, 2011.

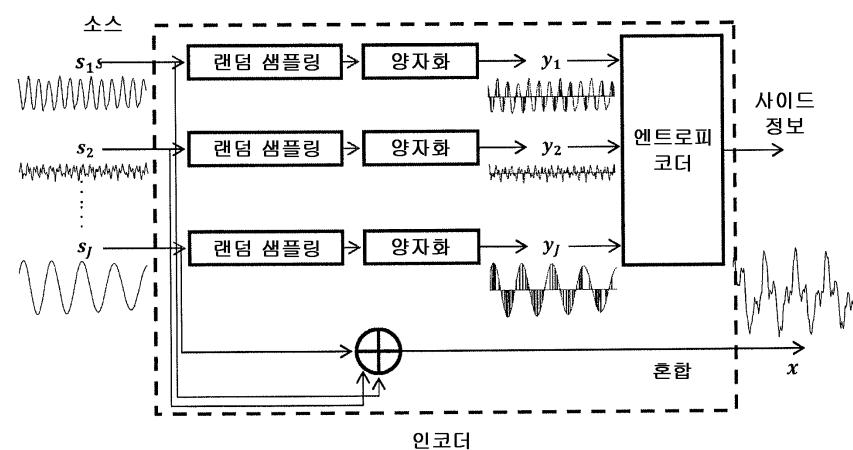
- [0103] [4] A. Liutkus, J. Pine, R. Badeau, L. Girin, 및 G. Richard, "Informed source separation through spectrogram coding and data embedding", *Signal Processing*, vol. 92, no. 8, pp. 1937-1949, 2012.
- [0104] [5] A. Ozerov, A. Liutkus, R. Badeau, 및 G. Richard, "Coding-based informed source separation: Nonnegative tensor factorization approach", *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 8, pp. 1699-1712, Aug. 2013.
- [0105] [6] J. Engdegard, B. Resch, C. Falch, O. Hellmuth, J. Hilpert, **A. Hölzer**, L. Terentiev, J. Breebaart, J. Koppens, E. Schuijers, 및 W. Oomen, "Spatial audio object coding (SAOC) - The upcoming MPEG standard on parametric object based audio coding", (*124th Audio Engineering Society Convention (AES 2008)*, Amsterdam, Netherlands, May 2008 내).
- [0106] [7] A. Ozerov, A. Liutkus, R. Badeau, 및 G. Richard, "Informed source separation: source coding meets source separation", (*IEEE Workshop Applications of Signal Processing to Audio and Acoustics (WASPAA'11)*, New Paltz, New York, USA, Oct. 2011, pp. 257-260 내).
- [0107] [8] S. Kirbiz, A. Ozerov, A. Liutkus, 및 L. Girin, "Perceptual coding-based informed source separation", (*Proc. 22nd European Signal Processing Conference (EUSIPCO)*, 2014, pp. 959-963 내).
- [0108] [9] Z. Xiong, A. D. Liveris, and S. Cheng, "Distributed source coding for sensor networks", *IEEE Signal Processing Magazine*, vol. 21, no. 5, pp. 80-94, September 2004.
- [0109] [10] B. Girod, A. Aaron, S. Rane, 및 D. Rebollo-Monedero, "Distributed video coding", *Proceedings of the IEEE*, vol. 93, no. 1, pp. 71 - 83, January 2005.
- [0110] [11] D. Donoho, "Compressed sensing", *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1289-1306, Apr. 2006.
- [0111] [12] R. G. Baraniuk, "Compressive sensing", *IEEE Signal Processing Mag.*, vol. 24, no. 4, pp. 118-120, July 2007.
- [0112] [13] E. J. Candes 및 M. B. Wakin, "An introduction to compressive sampling", *IEEE Signal Processing Magazine*, vol. 25, pp. 21-30, 2008.
- [0113] [14] R. G. Baraniuk, V. Cevher, M. F. Duarte, 및 C. Hegde, "Model-based compressive sensing", *IEEE Trans. Info. Theory*, vol. 56, no. 4, pp. 1982-2001, Apr. 2010.
- [0114] [15] C. Fevotte, N. Bertin, 및 J.-L. Durrieu, "Nonnegative matrix factorization with the Itakura-Saito divergence. With application to music analysis", *Neural Computation*, vol. 21, no. 3, pp. 793-830, Mar. 2009.
- [0115] [16] A. P. Dempster, N. M. Laird, 및 D. B. Rubin., "Maximum likelihood from incomplete data via the EM algorithm", *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 39, pp. 1-38, 1977.
- [0116] [17] S.M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice Hall, 1993.
- [0117] [18] A. Ozerov, C. Fevotte, R. Blouet, 및 J.-L. Durrieu, "Multichannel nonnegative tensor factorization with structured constraints for user-guided audio source separation", (*IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'11)*, Prague, May 2011, pp. 257-260 내).
- [0118] [19] V. Emiya, E. Vincent, N. Harlander, 및 V. Hohmann, "Subjective and objective quality assessment of audio source separation", *IEEE Trans. Audio, Speech, Language Process.*, vol. 19, no. 7, pp. 2046-2057, 2011.

### 도면

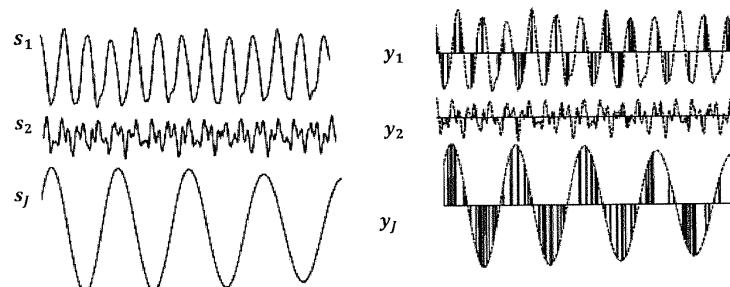
#### 도면1



#### 도면2

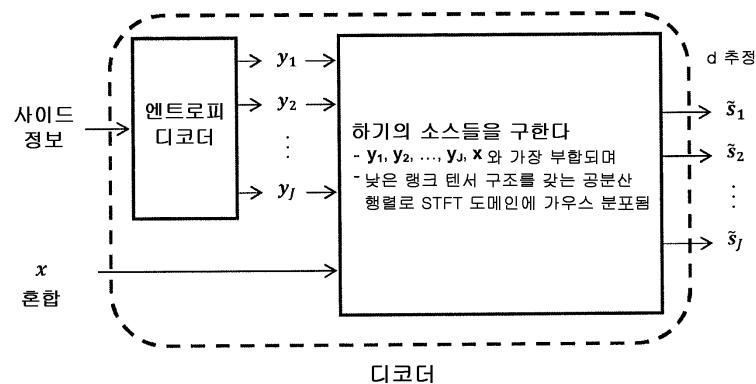


a)



b)

#### 도면3



## 도면4

