(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2019-0090578
(43) 공개일자 2019년08월02일

(51) 국제특허분류(Int. Cl.)
G01S 3/80 (2006.01) *G10L 21/02* (2006.01)
H04R 3/00 (2006.01)
(52) CPC특허분류
G01S 3/8006 (2013.01)
G10L 21/02 (2013.01)
(21) 출원번호 10-2018-0009446
(22) 출원일자 2018년01월25일
심사청구일자 2018년01월25일

(71) 출원인
서강대학교산학협력단
서울특별시 마포구 백범로 35 (신수동, 서강대학교)
(72) 발명자
박형민
서울특별시 강남구 삼성로 212 3동 1002호 (대치동, 은마아파트)
이란
서울특별시 마포구 염리동 128-62 제일타운 302호
(74) 대리인
이지연

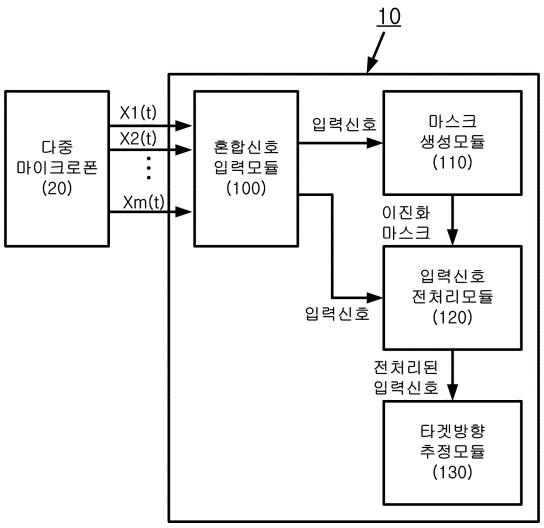
전체 청구항 수 : 총 11 항

(54) 발명의 명칭 **분산도 마스크를 이용한 음원 국지화 방법 및 음원 국지화 장치**

(57) 요 약

본 발명은 분산도를 기반으로 한 마스크를 전술한 기술적 과제를 달성하기 위한 본 발명의 제1 특징에 따른 음원 국지화 방법은, (a) 적어도 둘 이상의 마이크를 포함하는 다중 마이크로폰을 통해 타겟 음원신호와 잡음 및 반향 신호가 혼합된 신호를 입력받는 단계; (b) 상기 입력된 신호들을 이용하여 타겟 음원과 잡음원에 대한 정보인 일관성 대 분산도의 비(CDR)을 이용하여 분산도를 기반으로 한 이진화 마스크를 생성하는 단계; (c) 상기 생성된 이진화 마스크를 이용하여 다중 마이크로폰으로의 입력 신호를 전처리하는 단계; (d) 상기 전처리된 입력 신호에 대하여 GCC-PHAT 또는 SRP-PHAT와 같은 사전 설정된 알고리즘을 수행하여 타겟 음원에 대한 방향을 추정하는 단계; 를 구비한다.

대 표 도 - 도6



(52) CPC특허분류
H04R 3/005 (2013.01)

이 발명을 지원한 국가연구개발사업
과제고유번호 2017R1A2B4009964
부처명 과학기술정보통신부
연구관리전문기관 한국연구재단
연구사업명 기초연구사업(중견연구)
연구과제명 시청각 정보에 대한 멀티모달 딥러닝 기반의 강인한 연속음성인식 기술 개발
기 여 율 1/1
주관기관 서강대학교 산학협력단
연구기간 2017.03.01 ~ 2020.02.29

명 세 서

청구범위

청구항 1

(a) 적어도 둘 이상의 마이크를 포함하는 다중 마이크로폰을 통해 타겟 음원신호와 잡음 및 반향 신호가 혼합된 신호를 입력받는 단계;

(b) 상기 입력된 신호들을 이용하여 타겟 음원과 잡음원에 대한 정보를 반영하는 분산도를 기반으로 하는 마스크를 생성하는 단계;

(c) 상기 생성된 마스크를 이용하여 다중 마이크로폰으로의 입력 신호를 전처리하는 단계;

(d) 상기 전처리된 입력 신호에 대하여 사전 설정된 알고리즘을 수행하여 타겟 음원에 대한 방향을 추정하는 단계;

를 구비하는 것을 특징으로 하는 음원 국지화 방법.

청구항 2

제1항에 있어서. 상기 (b) 마스크를 생성하는 단계는,

입력 신호의 각 주파수 프레임(f) 및 시간 프레임(l)에 대한 일관성 대 분산도의 비(Coherence to Diffuseness ratio) (CDR(l,f)을 계산하고,

상기 CDR(l,f) 을 이용하여 분산도(D(l,f))를 계산하고,

상기 분산도(D(l,f))를 이용하여 아래의 수학식에 따라 마스크(MASk)값을 설정하여 이진화 마스크(M)을 생성하는 것을 특징으로 하는 음원 국지화 방법.

$$\mathbf{MASK} = \begin{cases} \mathbf{1 : D} \leq \boldsymbol{threshold} \\ \mathbf{0 : D} > \boldsymbol{threshold} \end{cases}$$

청구항 3

제2항에 있어서, 상기 (c) 입력 신호를 전처리하는 단계는,

이진화 마스크를 이용하여 입력 신호를 이진화시키는 것을 특징으로 하는 음원 국지화 방법.

청구항 4

제1항에 있어서, 상기 (d) 단계의 사전 설정된 알고리즘은 일반화된 교차 상관(Generalized Cross-correlation;'GCC')값을 기반으로 한 음원 국지화 방법이거나 Steered Response Power('SRP')를 기반으로 한 음원 국지화 방법인 것을 특징으로 하는 음원 국지화 방법.

청구항 5

제4항에 있어서, 상기 알고리즘들은 각 주파수의 신호들에 대하여 아래의 수학식에 따른 가중치(Ψ(ω))를 부가하는 Phase Transform('PHAT") 방식을 적용하는 것을 특징으로 하는 음원 국지화 방법.

$$\Psi_{kl}(\omega) = \frac{1}{|X_k(\omega)X_l^*(\omega)|}$$

여기서, k, l 은 마이크의 번호이며, ω=2πf 이며, $X_k(\omega)$는 k 번째 마이크의 입력 신호에 대한 푸리에 변환값

이며, $X_l^*(\omega)$는 l번째 마이크의 입력 신호에 대한 푸리에 변환값의 켤레값임.

청구항 6

제2항에 있어서, 각 주파수 프레임(f) 및 시간 프레임(l)에 대한 CDR(l,f)은 잡음(n), 타겟 음원 신호(s) 및 잡음과 타겟 음원 신호의 혼합 신호(x)에 대한 일관성들을 이용하여 아래의 수학식에 따라 추정되는 것을 특징으로 하는 음원 국지화 방법.

$$CDR(l,f) = \frac{\Gamma_n(f) - \Gamma_x(l,f)}{\Gamma_x(l,f) - \Gamma_s(f)}$$

여기서, $\Gamma_n(f)$는 잡음(n)에 대한 일관성이며, $\Gamma_s(f)$는 타겟 음원 신호(s)에 대한 일관성이며, $\Gamma_x(l,f)$는 잡음과 타겟 음원 신호의 혼합 신호(x)에 대한 일관성임.

청구항 7

제2항에 있어서, 상기 분산도(D(*l,f*))는 아래의 수학식에 따라 계산되는 것을 특징으로 하는 음원 국지화 방법.

$$\boldsymbol{D(l,f) = \frac{1}{CDR(l,f)+1}}$$

$$\boldsymbol{0 \le D \le 1}$$

청구항 8

다중 마이크로폰과 연결되어 다중 마이크로폰으로부터 타겟 음원 신호와 잡음 및 반향 신호가 혼합된 신호를 입력받는 혼합 신호 입력 모듈;

상기 혼합 신호 입력 모듈로부터 제공된 혼합 신호를 이용하여 분산도를 기반으로 한 이진화 마스크를 생성하여 출력하는 마스크 생성 모듈;

상기 마스크 생성 모듈로부터 이진화 마스크를 제공받고, 상기 혼함 신호 입력 모듈로부터 제공된 혼합 신호에 대하여 이진화 마스크를 적용하여 혼합 신호를 전처리하여 출력하는 입력 신호 전처리 모듈;

상기 입력 신호 전처리 모듈로부터 전처리된 혼합 신호를 제공받고, 상기 혼합 신호에 대하여 사전 설정된 국지화 알고리을 수행하여 타겟 음원에 대한 방향을 추정하여 출력하는 타겟 방향 추정 모듈;

을 구비하여, 다중 마이크로폰으로부터 입력되는 신호들을 이용하여 타겟 음원의 방향을 추정하는 것을 특징으로 하는 음원 국지화 장치.

청구항 9

제8항에 있어서, 상기 마스크 생성 모듈은,

상기 혼합 신호 입력 모듈로부터 제공된 혼합 신호의 각 주파수 프레임(*f*) 및 시간 프레임(*l*)에 대한 일관성 대 분산도의 비(Coherence to Diffuseness ratio) (CDR(*l,f*)을 계산하고,

상기 CDR(*l,f*) 을 이용하여 분산도(D(*l,f*))를 계산하고,

상기 분산도(D(*l,f*))를 이용하여 아래의 수학식에 따라 마스크(MASk)값을 설정하여 이진화 마스크(M)을 생성하는 것을 특징으로 하는 음원 국지화 장치.

$$\mathbf{MASK} = \begin{cases} \mathbf{1 : D \le \mathit{threshold}} \\ \mathbf{0 : D > \mathit{threshold}} \end{cases}$$

청구항 10

제8항에 있어서, 상기 타겟 방향 추정 모듈의 사전 설정된 국지화 알고리즘은 일반화된 교차 상관(Generalized Cross-correlation;'GCC')값을 기반으로 한 음원 국지화 방법이거나 Steered Response Power('SRP')를 기반으로 한 음원 국지화 방법인 것을 특징으로 하는 음원 국지화 장치.

청구항 11

제9항에 있어서, 각 주파수 프레임(f) 및 시간 프레임(l)에 대한 CDR(l,f)은 잡음(n), 타겟 음원 신호(s) 및 잡음과 타겟 음원 신호의 혼합 신호(x)에 대한 일관성들을 이용하여 아래의 수학식에 따라 추정되는 것을 특징으로 하는 음원 국지화 장치.

$$CDR(l,f) = \frac{\Gamma_n(f) - \Gamma_x(l,f)}{\Gamma_x(l,f) - \Gamma_s(f)}$$

여기서, $\Gamma_n(f)$는 잡음(n)에 대한 일관성이며, $\Gamma_s(f)$는 타겟 음원 신호(s)에 대한 일관성이며, $\Gamma_x(l,f)$는 잡음과 타겟 음원 신호의 혼합 신호(x)에 대한 일관성임.

발명의 설명

기 술 분 야

[0001] 본 발명은 음원 국지화 방법에 관한 것으로서, 더욱 구체적으로는 타겟 음원 신호가 잡음 및 반향 환경에서 다중 마이크로폰으로 입력된 혼합 신호에 대하여, 일관성(Coherence) 대 분산도(Diffuseness) 파워비인 CDR(Coherence to Diffuseness ratio)을 이용하여 만든 분산도 마스크를 적용하여 교차 상관 기법을 기반으로 하여 타겟 음원의 방향을 추정함으로써, 반향 및 분산 잡음에 강인한 음원 국지화 방법 및 음원 국지화 장치에 관한 것이다.

배 경 기 술

[0002] 최근 인공지능이 급속도로 발전함에 따라 인공지능 음성인식에 대한 관심도가 높아지고 있다. 이에 따라 인공지능 음성인식 스피커가 대중화되면서, 원거리 및 잡음이 있는 환경에서는 음성인식의 성능이 저하되는 문제점이 발생되고 있다. 따라서, 반향 및 잡음이 존재하는 환경에서 정확한 성능을 보장하는 음원 국지화 기술의 개발이 요구되고 있다.

[0003] 다중 마이크로폰을 사용할 때에는 음원이 마이크로 입력될 때의 음원의 방향에 따라 마이크마다 음원이 도달하는 시간이 다르게 된다. 따라서 이를 이용해 각 마이크 별 신호의 도달 시간의 차를 추정하여 음원의 방향을 유추할 수 있다.

[0004] 음원의 위치를 탐지하기 위한 음원 국지화 알고리즘으로 교차 상관 기반의 음원 국지화 방법이 가장 널리 사용되고 있다. 교차 상관 기반의 음원 국지화 방법은 안정적인 성능과 다중 음원 방향 추정이 가능하다는 장점을 갖지만, 잡음이나 반향이 존재하는 환경에서 전 방향으로 교차 상관 값이 커지게 되어 음원 국지화 성능이 크게 악화되는 단점이 있다. 즉, 대부분의 교차 상관 기반 방법들은 주변 잡음원으로 인한 가산 잡음 왜곡이 존재할 때나, 반향으로 인해 직선 경로 성분이 반사되어 원신호와 혼합되어 입력되는 경우에는 그 정확도가 매우 떨어지게 된다. 특히, 반향 성분들에 의한 왜곡이 존재할 때는 이전 시간에 존재했던 직접 경로 성분이 공간에서 반사됨에 따라 지연, 감쇄되어 현재 신호와 혼합되기 때문에, 직접 경로 성분과 상관도가 높아 잡음제거와 다른 별도의 알고리즘이 요구된다. 이러한 환경에서의 성능을 향상시키기 위한 기술들이 개발되어 왔지만 대부분 제한된 성능 향상을 보였고, 정확도 향상을 위한 긴 데이터와 많은 계산량이 요구되고 있는 실정이다

[0005] 따라서, 잡음과 반향이 존재하는 환경에서의 음성인식 성능을 개선할 수 있는 안정적이고 정확한 음원 국지화 기술이 요구된다.

[0006] 이하, 본 발명에 따른 음원 국지화 방법을 설명하기에 앞서, 기존의 음원 국지화 알고리즘들에 대해 설명한다.

[0007] < 일반화된 교차 상관 방법 (Generalized Cross Correaltion) >

[0008] 도 1은 구좌표계에서의 음원으로부터 나온 음성 신호가 각도별 다중 마이크로폰에 도착하는 시간 지연을 설명하기 위하여 도시한 그래프이다. 도 1을 참조하면, 원거리에서 (r_S, ϕ_S, θ_S) 방향에 타겟 음원이 존재할 때, 잡음 환경에서 두 개의 마이크로 들어오는 신호 중 두번째 마이크의 신호인 $x_2(t)$는 수학식 1과 같다.

수학식 1

$$x_2(t) = \alpha_2(r_s, \varnothing_s, \theta_s)s[t - f_{12}(r_s, \varnothing_s, \theta_s)] + v_m(t)$$

[0009]

[0010] 여기서, $x_m(t)$는 마이크로 들어온 입력 신호이며, $s(t)$ 및 $v_m(t)$는 각각 원신호와 가산잡음을 나타내며, α_m은 원 신호의 감쇠도이다. f_{12}는 음원의 위치에 따른 첫번째 마이크와 두번째 마이크 사이의 상대적 시간 지연을 의미한다. f_{12}는 수학식 2로 나타낼 수 있다.

수학식 2

$$f_{12}(r_s, \varnothing_s, \theta_s) = \frac{1}{c}[d_{s,1}(r_s, \varnothing_s, \theta_s) - d_{s,2}(r_s, \varnothing_s, \theta_s)]$$

[0011]

[0012] 여기서, c는 소리의 속도이며 d는 음원(s)과 각 마이크 사이의 거리이다.

[0013] 여기서 첫번째 마이크와 두번째 마이크 간에 발생할 수 있는 샘플 지연(τ)은 수학식 3과 같다. 수학식 3에서 round 의 의미는 추정한 값을 측정가능한 샘플 지연 단위로 환산하는 의미에서 사용된다.

수학식 3

$$\tau = \text{round}[f_{12}(r_s, \varnothing_s, \theta_s)]$$

[0014]

[0015] 수학식 1을 참조하여 다중 마이크에 들어오는 입력 신호를 살펴 보면, 기준 마이크에 입력된 신호가 두번째 마이크에서 감쇄되고 지연되어 들어오는 형태임을 알 수 있다. 따라서, 두 마이크 사이의 시간 지연(τ)에 해당하는 지연 보상이 이루어졌을 때, 두 마이크 사이의 교차 상관값이 최대를 보일 것이라고 예상할 수 있다. 도 2는 지연 보상이 되었을 때의 교차 상관값을 설명하기 위한 그림이다.

[0016] 두 마이크의 신호 $x_1(t)$와 $x_2(t)$ 사이의 교차 상관값은 수학식 4와 같고, 이를 푸리에 변환하여 주파수 영역으로 변환하면 수학식 5와 같다.

수학식 4

$$c_{12}(\tau) = \int_{-\infty}^{+\infty} x_1(t)x_2(t+\tau)dt$$

[0017]

수학식 5

[0018] $$C_{12}(\omega) = \int_{-\infty}^{+\infty} c_{12}(\tau) e^{-j\omega\tau} d\tau$$

[0019] 수학식 4와 5를 정리하고, 푸리에 변환에서의 컨볼루션 특징을 적용하면 $\omega=2\pi f$ 일 때 수학식 6과 같이 정리할 수 있다.

수학식 6

[0020] $$C_{12}(\omega) = X_1(\omega)X'_2(\omega)$$

[0021] 여기서, $X_1(\omega)$는 $x_1(t)$의 푸리에 변환식이며, $X_2'(\omega)$는 $x_2(t)$의 푸리에 변환식의 켤레 복소이다.

[0022] 이를 다시 역푸리에 변환함으로써, 수학식 7과 같이, 원래의 마이크 입력에서 τ만큼의 시간 지연 보상이 이루어진 교차 상관값에 대한 식을 구할 수 있다.

수학식 7

[0023] $$c_{12}(\tau) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} X_1(\omega)X'_2(\omega)e^{j\omega\tau} d\omega$$

[0024] 수학식 7의 교차 상관값을 바탕으로 하여, $x_1(t)$와 $x_2(t)$가 각각 필터 $H_1(\omega)$와 $H_2(\omega)$를 통과한 신호에 대해 교차 상관값을 정의하면 수학식 8 내지 수학식 10과 같이 나타낼 수 있다.

수학식 8

[0025] $$R_{12}(\tau) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} (H_1(\omega)X_1(\omega))(H_2(\omega)X_2(\omega))'e^{j\omega\tau} d\omega$$

수학식 9

[0026] $$R_{12}(\tau) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} H_1(\omega)H'_2(\omega)X_1(\omega)X'_2(\omega)e^{j\omega\tau} d\omega$$

수학식 10

[0027] $$R_{12}(\tau) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \Psi_{12}(\omega)X_1(\omega)X'_2(\omega)e^{j\omega\tau}d\omega$$

[0028] 여기서, $\Psi_{12}(\omega)$는 가중치 함수가 된다.

[0029] 따라서, 최종적인 음원의 방향은 수학식 11과 같이, 수학식 10의 일반화된 교차 상관값인 $R_{12}(\tau)$ 를 최대화하는 τ를 추정함으로써 추정할 수 있게 된다.

수학식 11

[0030] $$\hat{\tau} = \arg\max_{\tau \in D} R_{12}(\tau)$$

[0031] < 가중치 교차 상관 방법(weighted GCC) >

[0032] 일반적으로 신호를 취득할 때에는 주파수의 겹침 문제를 해결하기 위해 가장 낮은 주파수 영역의 파장보다 작은 값을 갖도록 두 마이크 사이의 거리를 가깝게 둔다. 낮은 주파수 영역에는 대부분의 음성 에너지가 분포 하는데 이러한 낮은 주파수 값에 영향을 많이 받게 될수록 교차 상관 값의 피크 폭은 그 폭의 너비가 더 넓어지는 경향을 띤다. 그 결과 평평해진 교차 상관 값에 기반한 음원 국지화는 낮은 성능을 보이게 된다. 따라서 이러한 문제를 보완하기 위하여 주파수에 따라 다른 가중치를 주어 교차 상관 값을 구한다. 일반화된 교차 상관의 경우도 두 필터 값 $H_1(\omega)$와 $H_2(\omega)$에 의해 주파수 성분에 가중치가 더해진 형태로 볼 수 있는데, $\Psi = H_1 H_2'$ 가 바로 가중치를 반영하는 부분이라고 할 수 있다.

[0033] GCC에서 사용하는 가장 간단한 가중치 함수 이외에도 교차 상관 값을 주파수 성분 별로 보완하는 여러 가중치 함수들이 존재하는데, 이 중, 가장 많이 쓰는 가중치 함수는 phase transform(PHAT)이다. 이 방식은 간단한 방식으로 억제가 필요한 부분의 교차 상관 값의 크기를 효과적으로 줄일 수 있으며, 각 주파수 대역에서 각 위상에 대해 동일한 가중치를 제공한다. 이러한 특성에 따라 PHAT 기반의 교차 상관 값은 이론적으로는 델타 함수만큼이나 뾰족한 첨도를 갖게 된다. 따라서 반향이 있는 환경에서 음원을 찾아야 할 때나 다중 음원을 찾아야 되는 경우 유리하게 작용된다.

[0034] PHAT 가중치 함수는 $\frac{1}{|G_{X_1X_2}(f)|}$ 이며, 여기서 $G_{X_iX_j}(f) = X_i(f)X_j^H(f)$ 이고, i와 j는 마이크로폰 번호를 가리킨다. GCC-PHAT는 GCC에 비해 다른 방향 대비 음원 방향에 대해 높은 교차 상관값을 가지며 피크에서 비교적 뾰족한 첨도를 보인다.

[0035] 반향 성분에 강인한 PHAT 의 가중치를 적용한 GCC-PHAT는 실질적으로 음원의 파워가 작을 때 잡음의 파워의 영향이 커짐에 따라 추정 오차가 커지게 되므로 잡음의 존재에 민감하게 되는 단점이 있다.

[0036] < Steered Response Power 빔포밍 >

[0037] Steered response power 기반의 음원 국지화 방법은 음원이 존재하지 않는 방향에 비해 음원이 존재하는 방향에서 더 큰 에너지를 갖는다는 가정을 이용하여, 마이크로 입사 가능한 모든 방향에 대한 delay-and-sum beamformer 출력을 계산하여 출력 power가 가장 높은 방향을 찾아 음원 국지화를 수행한다.

[0038] SRP 빔포밍은 다수 개(M)의 마이크에 대하여 입력 신호를 이용하며, 각 입력 신호는 수학식 12와 같이 나타낼 수 있다.

수학식 12

$$x_m(t) = s(t)*h(\vec{d}_s,t),\ \ m = 1,2,\ldots M$$

[0039]

[0040] 여기서, s(t)는 음원 신호이며, $h(\vec{d}_s,t)$ 는 음원 위치에 따른 impulse response이다.

[0041] 도 3은 종래의 delay sum beamforming 방법을 이용한 Steered Response Power 알고리즘을 설명하기 위하여 도시한 그림이다. 도 3을 참조하면, 각각의 마이크는 서로 다른 위치를 가지므로, 마이크에 입력되는 신호는 시간차를 가지게 된다. 이때, Steered Response Power 알고리즘은 delay-and-sum beamformer는 M개의 각 마이크로 입력되는 신호에 대하여 적당한 steering delay δm 를 주고 모든 신호를 더함으로서 enhanced된 신호를 얻는 방법이다.

[0042] 도 3에서는 신호들이 일정한 delay를 거친 후에 모두 더해지는 것을 보여주고 있다. 출력(Output)은 결과적으로 신호의 방향이 음원의 위치에 집중되기 때문에 증가된다. Delay-and-sum의 output은 수학식 13 및 수학식 14와 같이 나타낼 수 있다.

수학식 13

$$y(t,\delta_1,\delta_2,\ldots,\delta_M) = \sum_{m=1}^{M} x_m(t-\delta_m)$$

[0043]

수학식 14

$$\delta_m = \tau_m - \tau_0$$

[0044]

[0045] 이때, steering delay δm 은 특정 전파 지연 τ_m과 상수 τ_0의 차로 표현된다.

[0046] 수학식 12를 사용하여 수학식 13을 정리하면 수학식 15와 같이 되며, 이를 푸리에 변환하면 수학식 16과 같다.

수학식 15

$$y(t,\delta_1,\delta_2,\ldots,\delta_M) = s(t)*\sum_{m=1}^{M} h(\vec{d}_s, t-\tau_m+\tau_0)$$

[0047]

수학식 16

$$Y(\omega,\delta_1,\delta_2,\ldots,\delta_M) = \sum_{m=1}^{M} X_m(\omega)e^{-jw\delta_m}$$

[0048]

[0049] 여기서, $X_m(\omega)$는 $x_m(t)$의 푸리에 변환식이다.

[0050] 위에서의 각 마이크에 대한 steering delay δ_1, ''', δ_m 를 조정함으로써 특정 방향에 대한 beamformer 출력을 얻을 수 있다. 따라서 각 마이크에 대한 steering delay δ_1, ''', δ_m 를 마이크로 음원이 입사 가능한 모든 방향에 맞춰 조정하고, 그에 따른 beamformer 출력 power를 구함으로써 steered response power를 구할 수 있게 된다. 이를 주파수 축에서의 식으로 나타내면 수학식 17과 같다.

수학식 17

$$P(\delta_1,\delta_2,\ldots,\delta_M)=\int_{-\infty}^{\infty}Y(\omega,\delta_1,\ldots,\delta_M)\,Y^*(\omega,\delta_1,\ldots,\delta_M)d\omega$$

$$Y^*(\omega,\delta_1,\ldots,\delta_M)\ :\ \text{conjugate of}\ \ Y(\omega,\delta_1,\ldots,\delta_M)$$

[0051]

[0052] 수학식 17을 수학식 16을 통해 나타내면 수학식 18과 같다.

수학식 18

$$P(\delta_1,\ldots,\delta_M)=\int_{-\infty}^{\infty}\Big(\sum_{k=1}^{M}X_k(\omega)e^{-j\omega\delta_k}\Big)\Big(\sum_{l=1}^{M}X_l(\omega)e^{-j\omega\delta_l}\Big)d\omega$$

$$=\int_{-\infty}^{\infty}\sum_{k=1}^{M}\sum_{l=1}^{M}(X_k(\omega)X_l^*(\omega))e^{j\omega(\delta_l-\delta_k)}d\omega$$

[0053]

[0054] 수학식 14에 의해 수학식 19 및 수학식 20과 같은 관계를 갖는다.

수학식 19

$$\delta_l-\delta_k=\tau_l-\tau_k$$

[0055]

수학식 20

$$P(\delta_1,\ldots,\delta_M)=\int_{-\infty}^{\infty}\sum_{k=1}^{M}\sum_{l=1}^{M}(X_k(\omega)X_l^*(\omega))e^{j\omega(\tau_l-\tau_k)}d\omega$$

[0056]

[0057] 따라서, $P(\delta_1,\ ''',\ \delta_M)$ 는 모든 마이크 쌍에 대한 steered response power를 δ_1, ''', δ_M 에 따라 구하게 되며, 그것이 최대가 되는 steering delay δ_1, ''', δ_M 를 구하여 방향을 찾게 된다. 일반적으로 마이크 배열 중심으로부터 가능한 입사 방향으로의 azimuth(θ), elevation(ϕ)의 쌍 (θ, ϕ)에 대한 steering delay δ_1, ''', δ_N를 계산하고, 모든 (θ, ϕ) 에 해당하는 steered response power를 계산하여 최대가 되는 (θ, ϕ) 를 구하게 된다.

[0058] 위 수학식 20에서 결과값은 신호의 세기가 큰 주파수 성분에 큰 영향을 받기 때문에, 주파수별 가중치를 동일하

게 주기 위해서 Phase Transform(PHAT)은 각 주파수에 수학식 21과 같은 추가적인 가중치를 적용한다.

수학식 21

$$\Psi_{kl}(\omega) = \frac{1}{|X_k(\omega)X_l^*(\omega)|}$$

[0059]

[0060] 가중치 함수를 적용하고 time delay 를 $\tau_l - \tau_k = \tau_{lk}$ 로 표현하면 수학식 22와 같다.

수학식 22

$$P(\delta_1,\ldots,\delta_M) = \sum_{k=1}^{M}\sum_{l=1}^{M}\int_{-\infty}^{\infty}\frac{1}{|X_k(\omega)X_l^*(\omega)|}(X_k(\omega)X_l^*(\omega))e^{j\omega\tau_{lk}}d\omega$$

[0061]

[0062] 따라서, 최종적으로 steering delay를 수학식 23과 같이 얻어 음원의 방향을 찾을 수 있게 된다.

수학식 23

$$\hat{\delta}_1,\ldots,\hat{\delta}_M = \arg\max_{\delta_1,\ldots,\delta_M} P(\delta_1,\ldots,\delta_M)$$

[0063]

[0064] 그리고 마이크 한 쌍에 대한 localization 방법으로 일반적인 Generalized Cross Correlation(GCC)는 수학식 24와 같다.

수학식 24

$$R_{kl}(\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\Psi_{kl}(\omega)(X_k(\omega)X_k^*(\omega))e^{j\omega\tau}d\omega$$

[0065]

[0066] GCC는 두 마이크로 입사되는 신호의 cross-correlation을 이용하며, 이 값이 최대가 되는 time delay τ를 구함으로써 신호의 방향을 찾게 된다. SRP에 대한 수학식 22에서의 적분은 수학식 22에 2π 상수배를 한 값으로 대체할 수 있으며, 결과적으로 SRP는 모든 마이크 쌍에 대한 GCC를 누적한 것이 된다.

선행기술문헌

특허문헌

[0067] (특허문헌 0001) 한국등록특허공보 제 10-1269236호

(특허문헌 0002) 한국등록특허공보 제 10-1073632호

(특허문헌 0003) 한국등록특허공보 제 10-0612616호

발명의 내용

해결하려는 과제

[0068] 전술한 문제점을 해결하기 위한 본 발명의 목적은 분산도 마스크를 이용하여 반향 및 잡음에 강인한 음원 국지화 방법 및 장치를 제공하는 것이다.

과제의 해결 수단

[0069] 전술한 기술적 과제를 달성하기 위한 본 발명의 제1 특징에 따른 음원 국지화 방법은, (a) 적어도 둘 이상의 마이크를 포함하는 다중 마이크로폰을 통해 타겟 음원신호와 잡음 및 반향 신호가 혼합된 신호를 입력받는 단계; (b) 상기 입력된 신호들을 이용하여 타겟 음원과 잡음원에 대한 정보를 반영하는 분산도를 기반으로 하는 마스크를 생성하는 단계; (c) 상기 생성된 마스크를 이용하여 다중 마이크로폰으로의 입력 신호를 전처리하는 단계; (d) 상기 전처리된 입력 신호에 대하여 사전 설정된 알고리즘을 수행하여 타겟 음원에 대한 방향을 추정하는 단계; 를 구비한다.

[0070] 전술한 제1 특징에 따른 음원 국지화 방법에 있어서, 상기 (b) 마스크를 생성하는 단계는, 입력 신호의 각 주파수 프레임(f) 및 시간 프레임(l)에 대한 일관성 대 분산도의 비(Coherence to Diffuseness ratio) (CDR(l,f)을 계산하고, 상기 CDR(l,f) 을 이용하여 분산도(D(l,f))를 계산하고, 상기 분산도(D(l,f))를 이용하여 아래의 수학식에 따라 마스크(MASk)값을 설정하여 이진화 마스크(M)을 생성하는 것이 바람직하다.

$$\mathbf{MASK} = \begin{cases} \mathbf{1 : D \leq \mathit{threshold}} \\ \mathbf{0 : D > \mathit{threshold}} \end{cases}$$

[0071]

[0072] 전술한 제1 특징에 따른 음원 국지화 방법에 있어서, 상기 (c) 입력 신호를 전처리하는 단계는, 이진화 마스크를 이용하여 입력 신호를 이진화시키는 것이 바람직하다.

[0073] 전술한 제1 특징에 따른 음원 국지화 방법에 있어서, 상기 (d) 단계의 사전 설정된 알고리즘은 일반화된 교차상관(Generalized Cross-correlation;'GCC')값을 기반으로 한 음원 국지화 방법이거나 Steered Response Power('SRP')를 기반으로 한 음원 국지화 방법인 것이 바람직하며,

[0074] 상기 알고리즘들은 각 주파수의 신호들에 대하여 아래의 수학식에 따른 가중치($\Psi(\omega)$)를 부가하는 Phase Transform('PHAT") 방식을 적용하는 것이 더욱 바람직하다.

$$\Psi_{kl}(\omega) = \frac{1}{|X_k(\omega)X_l^*(\omega)|}$$

[0075]

[0076] 여기서, k, l 은 마이크의 번호이며, $\omega=2\pi f$ 이며, $X_k(\omega)$는 k 번째 마이크의 입력 신호에 대한 푸리에 변환값이며, $X_l^*(\omega)$는 l번째 마이크의 입력 신호에 대한 푸리에 변환값의 켤레값임.

[0077] 전술한 제1 특징에 따른 음원 국지화 방법에 있어서, 각 주파수 프레임(f) 및 시간 프레임(l)에 대한 CDR(l,f)은 잡음(n), 타겟 음원 신호(s) 및 잡음과 타겟 음원 신호의 혼합 신호(x)에 대한 일관성들을 이용하여 아래의 수학식에 따라 추정되는 것이 바람직하다.

$$CDR(l,f) = \frac{\Gamma_n(f) - \Gamma_x(l,f)}{\Gamma_x(l,f) - \Gamma_s(f)}$$

[0078]

[0079] 여기서, $\Gamma_n(f)$는 잡음(n)에 대한 일관성이며, $\Gamma_s(f)$는 타겟 음원 신호(s)에 대한 일관성이며, $\Gamma_x(l,f)$는 잡음과 타겟 음원 신호의 혼합 신호(x)에 대한 일관성임.

[0080] 전술한 제1 특징에 따른 음원 국지화 방법에 있어서, 상기 분산도(D(l,f))는 아래의 수학식에 따라 계산되는 것이 바람직하다.

$$D(l,f) = \frac{1}{CDR(l,f)+1}$$

$$0 \le D \le 1$$

[0081]

[0082] 본 발명의 제2 특징에 따른 음원 국지화 장치는, 다중 마이크로폰과 연결되어 다중 마이크로폰으로부터 타겟 음원 신호와 잡음 및 반향 신호가 혼합된 신호를 입력받는 혼합 신호 입력 모듈; 상기 혼합 신호 입력 모듈로부터 제공된 혼합 신호를 이용하여 분산도를 기반으로 한 이진화 마스크를 생성하여 출력하는 마스크 생성 모듈; 상기 마스크 생성 모듈로부터 이진화 마스크를 제공받고, 상기 혼함 신호 입력 모듈로부터 제공된 혼합 신호에 대하여 이진화 마스크를 적용하여 혼합 신호를 전처리하여 출력하는 입력 신호 전처리 모듈; 상기 입력 신호 전처리 모듈로부터 전처리된 혼합 신호를 제공받고, 상기 혼합 신호에 대하여 사전 설정된 국지화 알고리을 수행하여 타겟 음원에 대한 방향을 추정하여 출력하는 타겟 방향 추정 모듈; 을 구비하여, 다중 마이크로폰으로부터 입력되는 신호들을 이용하여 타겟 음원의 방향을 추정한다.

[0083] 전술한 제2 특징에 따른 음원 국지화 장치에 있어서, 상기 마스크 생성 모듈은, 상기 혼합 신호 입력 모듈로부터 제공된 혼합 신호의 각 주파수 프레임(f) 및 시간 프레임(l)에 대한 일관성 대 분산도의 비(Coherence to Diffuseness ratio) (CDR(l,f)을 계산하고, 상기 CDR(l,f) 을 이용하여 분산도(D(l,f))를 계산하고, 상기 분산도(D(l,f))를 이용하여 아래의 수학식에 따라 마스크(MASk)값을 설정하여 이진화 마스크(M)을 생성하는 것이 바람직하다.

$$\mathbf{MASK} = \begin{cases} \mathbf{1 : D} \le \boldsymbol{threshold} \\ \mathbf{0 : D} > \boldsymbol{threshold} \end{cases}$$

[0084]

[0085] 전술한 제2 특징에 따른 음원 국지화 장치에 있어서, 상기 타겟 방향 추정 모듈의 사전 설정된 국지화 알고리즘은 일반화된 교차 상관(Generalized Cross-correlation;'GCC')값을 기반으로 한 음원 국지화 방법이거나 Steered Response Power('SRP')를 기반으로 한 음원 국지화 방법인 것이 바람직하다.

[0086] 전술한 제2 특징에 따른 음원 국지화 장치에 있어서, 각 주파수 프레임(f) 및 시간 프레임(l)에 대한 CDR(l,f)은 잡음(n), 타겟 음원 신호(s) 및 잡음과 타겟 음원 신호의 혼합 신호(x)에 대한 일관성들을 이용하여 아래의 수학식에 따라 추정되는 것이 바람직하다.

$$CDR(l,f) = \frac{\Gamma_n(f) - \Gamma_x(l,f)}{\Gamma_x(l,f) - \Gamma_s(f)}$$

[0087]

[0088] 여기서, $\Gamma_n(f)$는 잡음(n)에 대한 일관성이며, $\Gamma_s(f)$는 타겟 음원 신호(s)에 대한 일관성이며, $\Gamma_x(l,f)$는 잡음과 타겟 음원 신호의 혼합 신호(x)에 대한 일관성임.

발명의 효과

[0089] 본 발명에 따른 음원 국지화 방법에 따라 입력 신호에 대해 CDR을 기반으로 한 이진화 마스크를 적용함으로써, 잡음과 반향이 증가되더라도 그 정확도가 증가되어 기존의 다른 알고리즘들에 비하여 잡음과 반향에 강인한 성능을 보인다.

도면의 간단한 설명

[0090] 도 1은 구좌표계에서의 음원으로부터 나온 음성 신호가 각도별 다중 마이크로폰에 도착하는 시간 지연을 설명하기 위하여 도시한 그래프이다.

도 2는 지연 보상이 되었을 때의 교차 상관값을 설명하기 위한 그림이다.

도 3은 종래의 delay sum beamforming 방법을 이용한 Steered Response Power 알고리즘을 설명하기 위하여 도시한 그림이다.

도 4는 분산 잡음의 구성을 예시적으로 도시한 그림이다.

도 5는 반향 환경에서의 입력 신호의 입력 경로를 예시적으로 도시한 상태도이다.

도 6은 본 발명의 바람직한 실시예에 따른 음원 국지화 방법을 구현한 음원 국지화 장치를 전체적으로 도시한 블록도이다.

도 7의 (a),(b) 및 (c)는 본 발명의 바람직한 실시예에 따른 음원 국지화 방법에 있어서, 입력된 혼합신호, 이에 따라 추정된 CDR, 및 이진화 마스크를 각각 도시한 그래프이다.

도 8은 본 발명의 바람직한 실시예에 따른 음원 국지화 방법의 성능을 종래의 방법과 비교하여 도시한 그래프들이다.

발명을 실시하기 위한 구체적인 내용

[0091] 본 발명에 따른 음원 국지화 방법 및 장치는 방향 및 잡음 환경에서, 타겟 음원과 잡음원에 대한 정보를 반영하는 분산도를 기반으로 한 이진화 마스크를 생성하고, 생성된 이진화 마스크를 이용하여 입력 신호를 변환시킨 후, 변환된 입력신호에 대하여 교차 상관 기법을 기반으로 한 GCC-PHAT 또는 SRP-PHAT을 적용하여 타겟 음원의 방향을 추정하는 것을 특징으로 한다.

[0092] 이하, 본 발명의 바람직한 음원 국지화 방법 및 장치에 대하여 구체적으로 설명한다.

[0093] 수학식 25와 같이, 타겟 음원 신호와 잡음의 파워 비율인 SNR(Signal to Noise ratio)를 정의할 때 타겟 음원의 자기 상관값 $\Phi_s(l,f)$ 과 잡음원의 자기 상관값 $\Phi_n(l,f)$ 의 비로 표현한다. 이 때, 타겟 음원의 자기 상관값 $\Phi_s(l,f)$ 과 잡음원의 자기 상관값 $\Phi_n(l,f)$ 은 잡음과 늦은 반향이 존재하는 환경에 대해 수학식 25와 같이 나타낼 수 있다.

수학식 25

[0094]
$$\mathrm{SNR(l,f)} = \frac{\Phi_s(l,f)}{\Phi_n(l,f)}$$
$$\Phi_{s_1 s_1}(l,f) = \Phi_{s_2 s_2}(l,f) = \Phi_s(l,f)$$
$$\Phi_{n_1 n_1}(l,f) = \Phi_{n_2 n_2}(l,f) = \Phi_n(l,f)$$
Φ : 상관값, $\Phi_{s_1 s_1}$: 타겟 음원의 자기상관, $\Phi_{n_1 n_1}(l,f)$: 잡음원의 자기상관

[0095] 여기서, l은 시간 프레임이며, f는 주파수 빈을 의미한다.

[0096] 다음, 분산도(D)는 수학식 26에 의해 측정된다. 여기서, *Ppw*는 평면파의 페이저, *Pdiff*는 분산 잡음의 페이저를 의미한다. 평면파와 분산 잡음은 각각 일관성에 의해 구분될 수 있는데, 이론적으로는 평면파의 일관성은 1에 해당하고 분산 잡음의 일관성은 0에 해당한다. 그리고, 분산도(D)는 0~1 의 범위의 값으로 나타내어지며, 값이 클수록 분산도가 높고 값이 작을수록 분산도가 낮음을 의미한다.

[0097] 한편, 일관성 대 분산도 비(Coherent to Diffuse power ratio ; 'CDR')는 수학식 27로 나타낼 수 있다. CDR은 SNR에서 타겟 음원을 평면파로 가정하고 잡음을 분산 잡음으로 가정할 때의 경우와 동일하다. 즉, CDR은 일관성이 높은 신호 대비 일관성이 낮은 신호의 비율이라고 볼 수 있다.

수학식 26

[0098]
$$D = \frac{\mathrm{E}[|\mathrm{P_{diff}}|^2]}{\mathrm{E}[|\mathrm{P_{diff}}|^2] + \mathrm{E}\left[\left|\mathrm{P}_{pw}\right|^2\right]}$$

수학식 27

$$CDR = \frac{E\left[|P_{pw}|^2\right]}{E\left[|P_{diff}|^2\right]}$$

$E\left[|P_{pw}|^2\right]$: 평면파의 파워, $E\left[|P_{diff}|^2\right]$: 분산잡음의 파워

[0099]

[0100] 따라서, 음성 신호의 경우에는 일관성이 높은 신호이며, 본 발명의 대상 잡음인 분산 잡음은 일관성이 낮은 신호이므로, CDR은 SNR과 동일하게 사용가능하다. 이때, '일관성'은 수학식 28에 나타난 바와 같이, 두 개의 마이크로 들어오는 입력 신호 $x_1(t)$와 $x_2(t)$가 존재할 때, 이 신호에 대한 교차 상관값 Φ_{x1x2} 을 신호의 자기 상관값들로 정규화한 값으로 정의된다. 이에 따라 일관성은 시간에 따른 음원의 크기 변화에 영향을 받지 않고 일정한 방향으로 들어온다고 가정할 경우 시간에 독립적임을 알 수 있다. 따라서, 일관성은 교차 상관 값에 비하여 타겟 음원과 잡음원에 대해 시간적인 특징을 배제한 공간적인 특징만을 반영할 수 있다. 즉, 타겟 신호와 잡음신호 각각의 공간적인 특징에 따라서 일관성의 식을 정할 수 있게 된다. 따라서, CDR 을 교차 상관 값이 아닌 혼합신호의 일관성, 타겟음원의 일관성, 잡음원의 일관성으로 나타낼 경우 혼합신호를 통해서도 잡음 대비 신호가 우세한 시간 및 주파수 영역을 추정할 수 있다.

수학식 28

$$\Gamma_{x_1x_2}(f) = \frac{\Phi_{x_1x_2}(l,f)}{\sqrt{\Phi_{x_1x_1}(l,f) * \Phi_{x_2x_2}(l,f)}}$$

[0101]

[0102] 수학식 28에 따른 일관성의 정의에 따라, 타겟 음원(s)과 잡음원(n)의 일관성은 수학식 29와 같이 각각 나타낼 수 있다.

수학식 29

$$\Gamma_s(f) = \frac{\Phi_{s_1s_2}(l,f)}{\Phi_s(l,f)}, \Gamma_n(f) = \frac{\Phi_{n_1n_2}(l,f)}{\Phi_n(l,f)}$$

[0103]

[0104] 다음, 입력 신호 x(t)에 대한 일관성은 수학식 30과 같이 나타낼 수 있으며, 이를 CDR에 대하여 나타내면 수학식 31 및 수학식 32로 나타낼 수 있으며, 다시 CDR은 각 신호에 대한 일관성을 이용하여 수학식 33과 같이 나타낼 수 있다.

수학식 30

$$\Gamma_x(l,f) = \frac{\Phi_{x_1x_2}(l,f)}{\Phi_x(l,f)} = \frac{\Phi_{s_1s_2}(l,f) + \Phi_{n_1n_2}(l,f)}{\Phi_s(l,f) + \Phi_n(l,f)}$$

$$= \frac{\frac{\Phi_s(l,f)}{\Phi_n(l,f)} * \frac{\Phi_{s_1s_2}(l,f)}{\Phi_s(l,f)} + \frac{\Phi_{n_1n_2}(l,f)}{\Phi_n(l,f)}}{\frac{\Phi_s(l,f)}{\Phi_n(l,f)} + 1}$$

[0105]

수학식 31

$$\Gamma_x(l,f) = \frac{SNR(l,f)\Gamma_s(f) + \Gamma_n(f)}{SNR(l,f) + 1}$$

[0106]

수학식 32

$$\Gamma_x(l,f) = \Gamma_s(f) + \frac{1}{CDR(l,f) + 1}(\Gamma_n(f) - \Gamma_s(f))$$

[0107]

수학식 33

$$CDR(l,f) = \frac{\Gamma_n(f) - \Gamma_x(l,f)}{\Gamma_x(l,f) - \Gamma_s(f)}$$

[0108]

[0109] 다시, 수학식 26에 따른 분산도(D)를 일관성을 이용하여 정리하면 수학식 34과 같이 나타낼 수 있다.

수학식 34

$$D(l,f) = \frac{1}{CDR(l,f) + 1}$$

$$0 \le D \le 1$$

[0110]

[0111] 이하, 본 발명의 바람직한 실시예에 따른 음원 국지화 방법을 적용하여 구현한 음원 국지화 장치에 대하여 구체적으로 설명한다.

[0112] 도 6은 본 발명의 바람직한 실시예에 따른 음원 국지화 방법을 구현한 음원 국지화 장치를 전체적으로 도시한 블록도이다. 도 6을 참조하면, 본 발명에 따른 음원 국지화 장치(10)는, 혼합 신호 입력 모듈(100), 마스크 생

성 모듈(110), 입력 신호 전처리 모듈(120), 타겟 방향 추정 모듈(130)을 구비한다.

[0113] 전술한 구성을 갖는 본 발명에 따른 음원 국지화 장치(10)는 M 개의 마이크들로 이루어진 다중 마이크로폰(20)과 연결되며, 다중 마이크로폰으로부터 입력되는 신호들을 이용하여 타겟 음원의 방향을 추정하게 된다.

[0114] 상기 혼합 신호 입력 모듈(100)은, 상기 다중 마이크로폰과 연결되어 다중 마이크로폰으로부터 타겟 음원 신호와 잡음 및 반향 신호가 혼합된 신호를 입력받게 된다.

[0115] 상기 마스크 생성 모듈(110)은, 상기 혼합 신호 입력 모듈로부터 제공된 혼합 신호를 이용하여 분산도를 기반으로 한 이진화 마스크(M)를 생성하여 출력한다. 상기 마스크 생성 모듈의 동작에 대한 자세한 설명은 후술한다.

[0116] 상기 입력 신호 전처리 모듈(120)은, 상기 마스크 생성 모듈로부터 이진화 마스크를 제공받고, 상기 혼합 신호 입력 모듈로부터 제공된 혼합 신호에 대하여 이진화 마스크를 적용하여 혼합 신호를 전처리하여 출력한다.

[0117] 상기 타겟 방향 추정 모듈(130)은, 상기 입력 신호 전처리 모듈로부터 전처리된 혼합 신호를 제공받고, 상기 혼합 신호에 대하여 GCC 알고리즘 또는 SRP 알고리즘을 이용하거나, 교차 상관값에 대하여 Phase tramsform 가중치 함수를 적용하는 GCC-PHAT 알고리즘이나 SRP-PHAT 알고리즘을 이용하여 타겟 음원의 방향을 추정하여 출력한다.

[0118] 이하, 본 발명에 따른 음원 국지화 방법에서 사용되는 분산도를 이용한 이진화 마스크의 생성 방법에 대하여 구체적으로 설명한다. 한편, 본 발명에 따른 음원 국지화 장치(10)의 마스크 생성 모듈(110)은 아래의 이진화 마스크 생성 방법을 적용하여 구현된 것을 특징으로 한다.

[0119] 본 발명에서는 분산도 값에 따라 타겟 음원이 우세한 시간 및 주파수에서의 방향 추정이 가능하도록 분산도를 기반으로 한 이진화 마스크를 사용한다. 먼저, 타겟 음원, 잡음 및 반향에 대한 일관성에 대하여 구체적인 정의를 설명한다.

[0120] 타겟 음원이 원거리에서 Θ의 도달 방향을 가지고 마이크로폰으로 입력될 경우의 타겟 음원의 일관성(Coherence)은 수학식 35로 나타낼 수 있다.

수학식 35

$$\Gamma_s(f) = \frac{\Phi_{s_1 s_2}(l,f)}{\Phi_s(l,f)} = e^{j2\pi f \Delta t} = e^{jkdsin(\theta)}$$

$\Delta t = \frac{dsin(\theta)}{c}, k = \frac{2\pi f}{c}$, d: 마이크 사이 거리, c : 소리의 속도, θ : 음원의 방향

[0121]

[0122] 다음, 환경적인 배경 잡음을 고려할 때에는 보통 무한의 개수를 가지는 비상관 잡음의 중첩 신호로 가정한다. 그리고 이러한 잡음은 공간적으로 원형 중간에 마이크로폰이 존재할 때에 모든 방향으로부터 입력된다. 즉, 잡음원이 마이크로폰으로부터 멀리 떨어진 모든 방향으로부터 균등하게 분포되어 있어 입력된 신호에서 상호관계성이 없게 되는 것이다. 그리고 이러한 잡음을 바로 분산 잡음 또는 등방성 잡음이라고 일컫는다. 대부분 실험적으로 이러한 분산 잡음을 생성할 때에는 도 4와 같이 많은 수의 비상관 잡음원을 전 방향으로 두어 구성한다. 도 4는 분산 잡음의 구성을 예시적으로 도시한 그림이다.

[0123] 그리고, 분산 잡음에 대한 일관성은 두 마이크로폰에 대하여 수학식 36과 같이 정의된다

수학식 36

$$\Gamma_{\text{diffuse}}(f) = \frac{\Phi_{n_1 n_2}(l,f)}{\Phi_n(l,f)} = \frac{\sin(kd)}{kd} = \frac{\sin\left(2\pi f \frac{d}{c}\right)}{2\pi f \frac{d}{c}}$$

[0124]

[0125] 한편, 반향 신호의 경우는 일반적으로 RIR(Room Impulse Response)와 타겟음원이 컨볼루션되어 입력된다고 가정한다. 이 때, 반향 신호는 도 5와 같이 녹음 환경 내부의 장애물에 부딪혀 여러 반사경로를 통하여 서로 다른 시간차와 감쇄된 크기를 가지고 마이크로폰으로 들어오게 되기 때문에 일반적으로 분산 잡음과 유사한 등방향성의 성질을 띄게 된다. 따라서 늦은 반향신호의 경우는 분산 잡음과 동일하게 취급이 가능하다. 도 5는 반향 환경에서의 입력 신호의 입력 경로를 예시적으로 도시한 상태도이다.

[0126] 먼저, CDR 값을 구하는 데에 있어 필요한 두 마이크 신호간의 자기 및 교차 상관 값은 상기 각 시간에 대하여 평균적인 값을 구하기 위하여 수학식 37과 같이 순환적으로 구한다. 이 때의 λ는 0과 1 사이의 상수 값이다

수학식 37

$$\widehat{\Phi}_{x_i x_j}(l,f) = \lambda\, \widehat{\Phi}_{x_i x_j}(l-1,f) + (1-\lambda) X_i(l,f) X_j^*(l,f)$$

[0127]

[0128] 또한, 수학식 33으로 나타낸 CDR을 결정하기 위하여, 수학식 38과 같이 타겟 음원의 일관성을 1로 가정하고, 수학식 38을 정리하여 수학식 39로 표현한다. 이를 이용하여 타겟 음원 방향을 모를 때의 CDR을 수학식 40을 통해 구한다. 수학식 40에서, CDR을 찾기 위해 0과 CDR 중에서 최대값을 취하는 이유는 CDR이 음수가 나오는 것을 방지하기 위함이다.

수학식 38

$$|\Gamma_s(f)| = |\Gamma_x(l,f) - (\Gamma_n(f) - \Gamma_x(l,f))CDR(l,f)^{-1}| \triangleq 1$$

[0129]

수학식 39

$$(|\Gamma_x(l,f)|^2 - 1)CDR(l,f)^2 -$$

$$2Re\{\Gamma_x(l,f)(\Gamma_n(f) - \Gamma_x(l,f))^*\}CDR(l,f) + |\Gamma_n(f) - \Gamma_x(l,f)|^2 = 0$$

[0130]

수학식 40

$$\widehat{CDR}_{noDOA}(l,f) = \max\left(0, \frac{\widetilde{\Gamma}_n(f)\mathrm{Re}\{\widetilde{\Gamma}_x(l,f)\} - |\widetilde{\Gamma}_x(l,f)|^2 - \sqrt{\widetilde{\Gamma}_n^2(f)\mathrm{Re}\{\widetilde{\Gamma}_x(l,f)\} - \widetilde{\Gamma}_n^2(f)|\widetilde{\Gamma}_n(f)|^2 + \widetilde{\Gamma}_n^2(f) - 2\widetilde{\Gamma}_n(f)\mathrm{Re}\{\widetilde{\Gamma}_x(l,f)\} + |\widetilde{\Gamma}_x(l,f)|^2}}{|\widetilde{\Gamma}_x(l,f)|^2 - 1}\right)$$

[0131]

[0132] 최종적으로, 분산도(D)의 값은 수학식 41과 같이 정리되고, 앞서 언급된 바와 같이 0부터 1 사이의 값을 갖게 된다.

수학식 41

$$D(l,f) = \frac{1}{1 + \widehat{CDR}_{noDOA}(l,f)}$$

[0133]

[0134] 일반적으로 잡음제거나 반향제거 알고리즘에서는 마스크를 이용하여 잡음원을 제거할 경우에 타겟 음원에 대한 왜곡이 생기기 때문에 음성인식률 성능을 떨어뜨리게 되므로, 잡음제거 또는 반향제거의 성능을 조금 포기하더라도 원 신호의 왜곡이 가지 않는 선에서 마스킹을 진행한다. 하지만, 본 발명에서는 CDR을 통하여 수행하고자 하는 것이 잡음제거나 반향제거가 아닌 잡음과 반향에 대한 정보를 주어 강인한 음원 국지화를 진행하는 것이므로, 타겟 음원의 두 마이크 사이의 위상 차가 보존되는 범위 안에서는 잡음을 최대한 많이 지워주는 것이 타겟 음원에서 더욱 명확한 교차 상관 값을 가질 수 있을 것이라 판단할 수 있다. 따라서, 수학식 42와 같이 선형적으로 값을 가지는 분산도의 값을 임계치(threshold)를 두어 이진화시키게 된다.

수학식 42

$$\mathbf{MASK} = \begin{cases} \mathbf{1 : D \leq threshold} \\ \mathbf{0 : D > threshold} \end{cases}$$

[0135]

[0136] 이때, 임계치(threshold)는 실험을 통하여, 가장 높은 타겟 음원 추정에 대한 정확도를 갖는 값으로 설정되는 것이 바람직하다.

[0137] 전술한 바와 같이, CDR을 이용하여 측정한 분산도를 이용하여 생성한 이진화 마스크를 마이크로폰으로 입력된 입력 신호에 적용하여 GCC-PHAT 또는 SRP-PHAT을 진행함으로써 반향 및 잡음에 강인한 음원 국지화를 구현할 수 있다.

[0138] 도 7의 (a),(b) 및 (c)는 본 발명의 바람직한 실시예에 따른 음원 국지화 방법에 있어서, 입력된 혼합신호, 이에 따라 추정된 CDR, 및 이진화 마스크를 각각 도시한 그래프이다. 도 7을 참조하면, 타겟 신호가 존재하는 구간에서 대부분 우세한 값을 보이면서, 음성 신호가 존재하는 시간 및 주파수를 검출해 내고 있음을 알 수 있다.

[0139] 도 8은 본 발명의 바람직한 실시예에 따른 음원 국지화 방법의 성능을 종래의 방법과 비교하여 도시한 그래프들이다. 도 8의 (a)는 종래의 방법에 따른 GCC-PHAT의 결과 그래프로서 오검출되었던 프레임을 나타내며, 도 8의 (b)는 본 발명에 따라 마스킹 기법을 적용한 GCC-PHAT의 결과를 도시한 그래프이다. 도 8을 참조하면, 종래의 방법에서는 타겟 음원의 방향이 오검출되었으나 본 발명에서는 정확하게 검출됨을 파악할 수 있다.

[0140] 한편, 앞서 설명한 바와 같이, SRP-PHAT 는 2개의 마이크로폰에 적용하는 GCC-PHAT 알고리즘을 멀티 마이크로폰에 적용한 것과 같다. 또한, 수학식적으로도 SRP-PHAT 는 모든 마이크로폰 쌍에 대한 GCC-PHAT의 합이다. 따라서, 마이크를 2개만 사용하는 경우, SRP-PHAT와 GCC-PHAT는 같은 결과를 가지게 된다. 따라서, 본 발명에 따른 국지화 방법은 분산도를 이용하여 생성한 이진화 마스크를 입력신호에 적용한 후, GCC-PHAT 뿐만 아니라 SRP-PHAT를 이용하여 타겟 음원의 방향을 추정할 수 있다.

[0141] 이상에서 본 발명에 대하여 그 바람직한 실시예를 중심으로 설명하였으나, 이는 단지 예시일 뿐 본 발명을 한정하는 것이 아니며, 본 발명이 속하는 분야의 통상의 지식을 가진 자라면 본 발명의 본질적인 특성을 벗어나지 않는 범위에서 이상에 예시되지 않은 여러 가지의 변형과 응용이 가능함을 알 수 있을 것이다. 그리고, 이러한 변형과 응용에 관계된 차이점들은 첨부된 청구 범위에서 규정하는 본 발명의 범위에 포함되는 것으로 해석되어야 할 것이다.

부호의 설명

[0142] 10 : 음원 국지화 장치

20 : 다중 마이크로폰

100 : 혼합 신호 입력 모듈

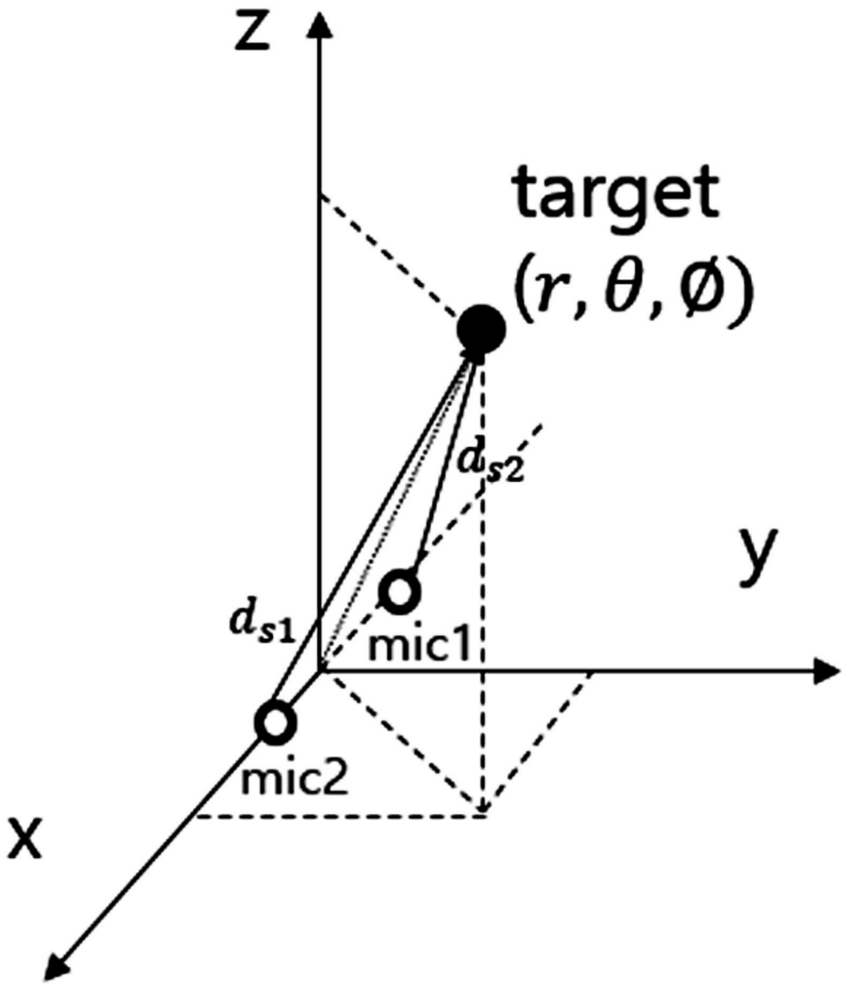
110 : 마스크 생성 모듈

120 : 입력 신호 전처리 모듈

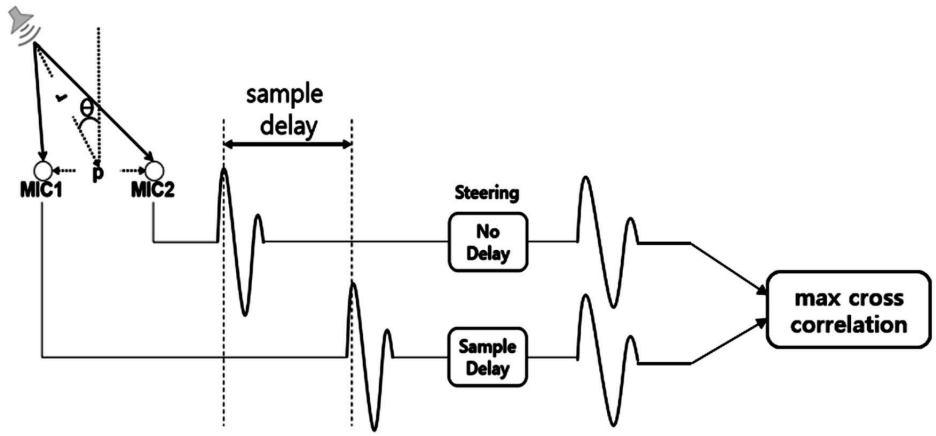
130 : 타겟 방향 추정 모듈

도면

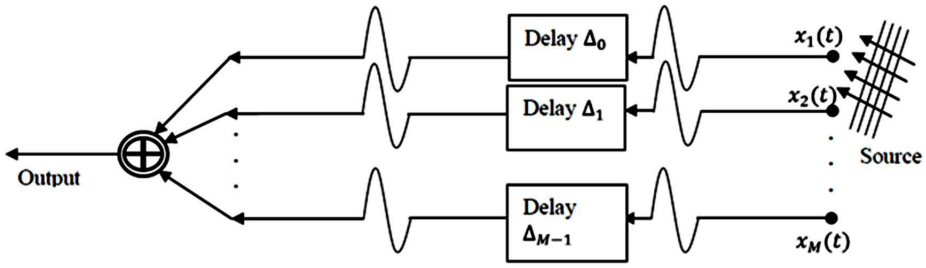
도면1



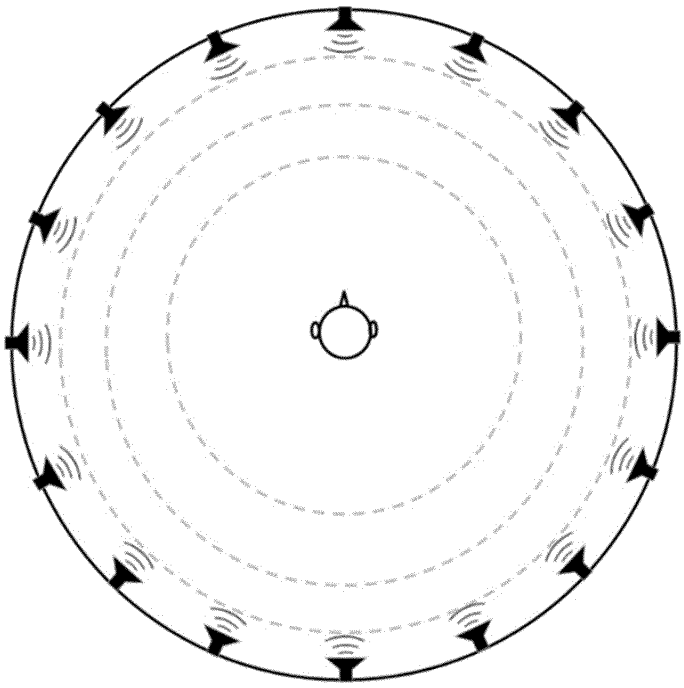
도면2


sample
delay
MIC1
MIC2
Steering
No
Delay
Sample
Delay
max cross
correlation

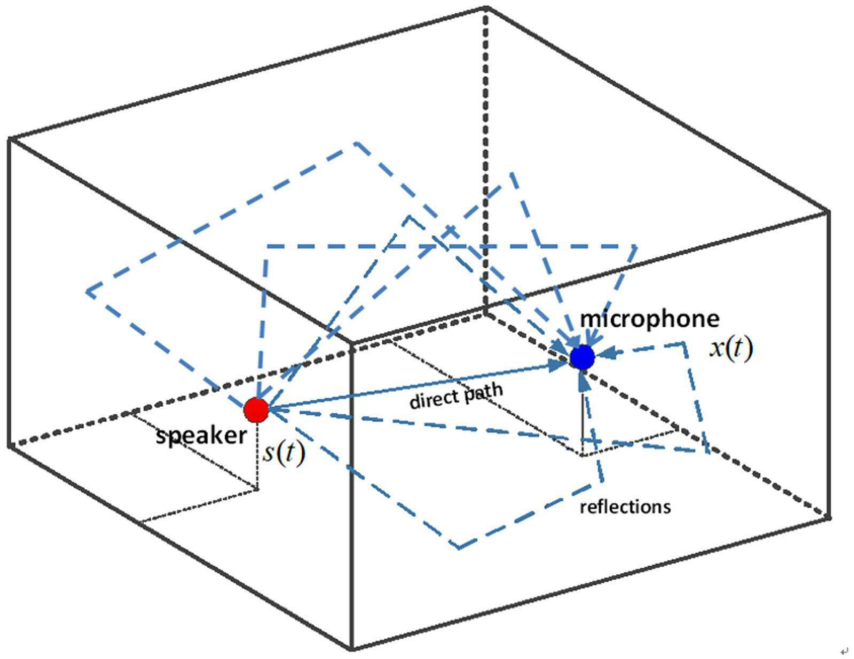
도면3


Delay Δ0
Delay Δ1
Delay
ΔM−1
x1(t)
x2(t)
xM(t)
Source
Output

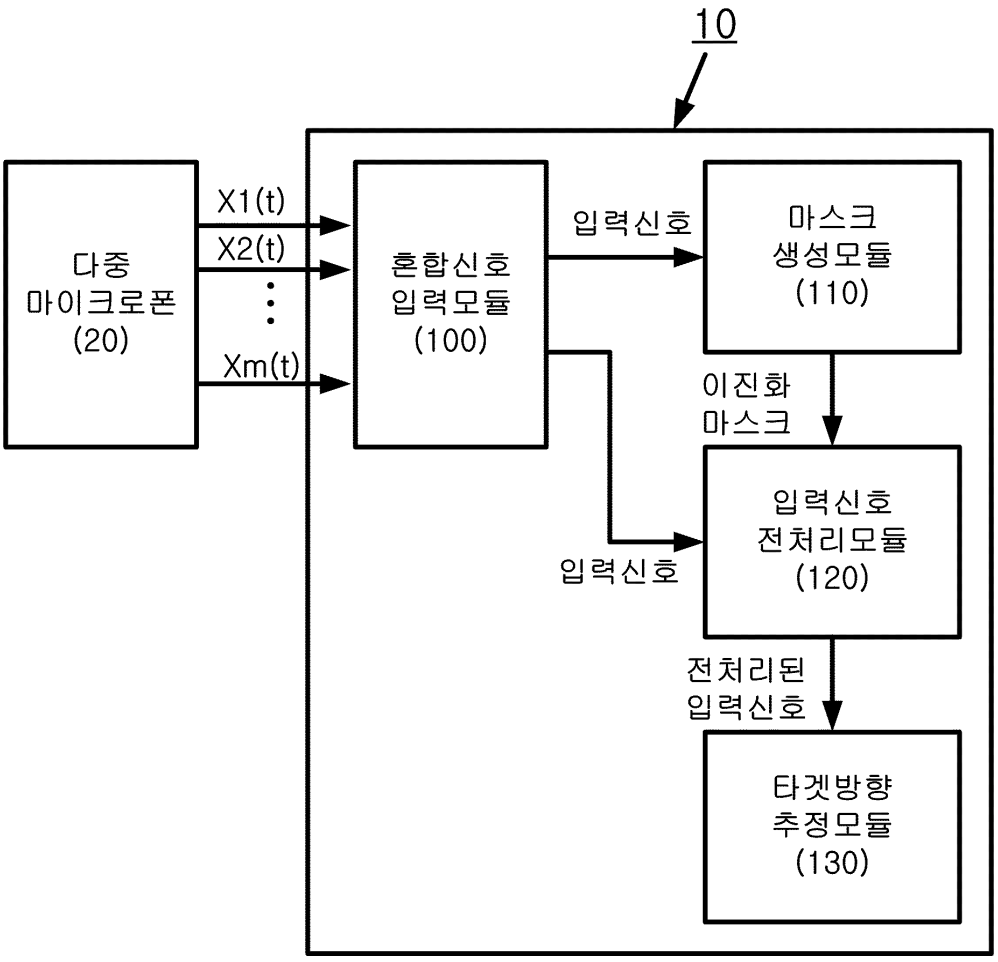
도면4



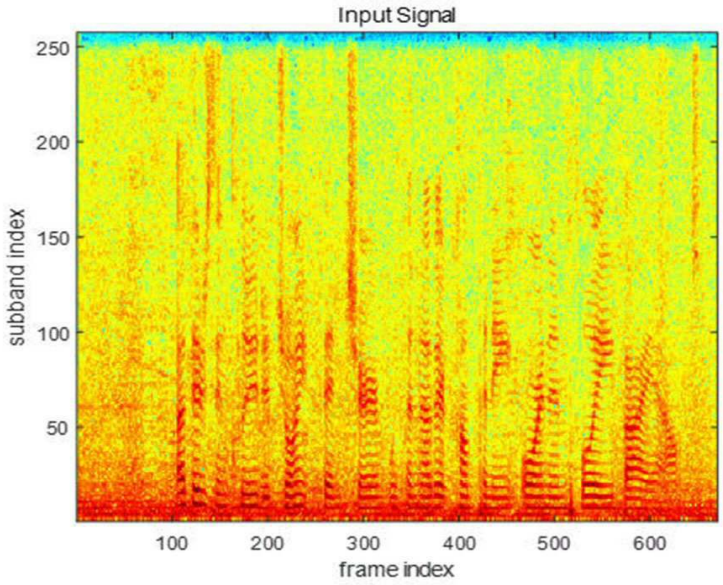
도면5


microphone
$x(t)$
direct path
speaker
$s(t)$
reflections

도면6

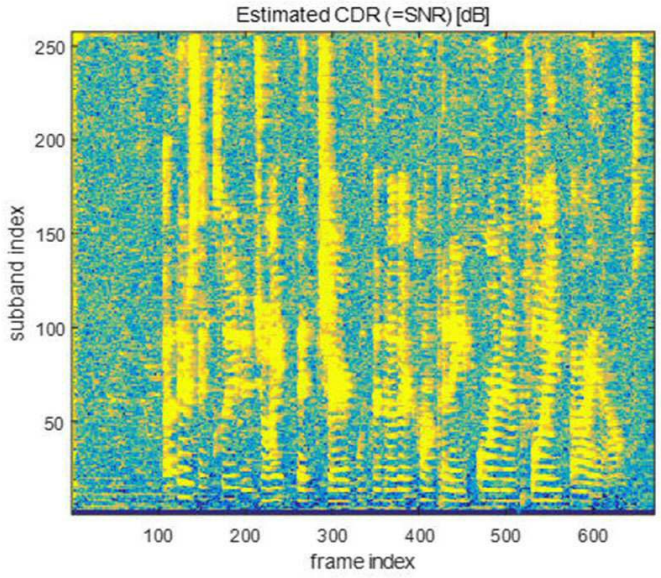

10
다중
마이크로폰
(20)
X1(t)
X2(t)
Xm(t)
혼합신호
입력모듈
(100)
입력신호
마스크
생성모듈
(110)
이진화
마스크
입력신호
입력신호
전처리모듈
(120)
전처리된
입력신호
타겟방향
추정모듈
(130)

도면7a


Input Signal
250
200
150
100
50
subband index
100
200
300
400
500
600
frame index

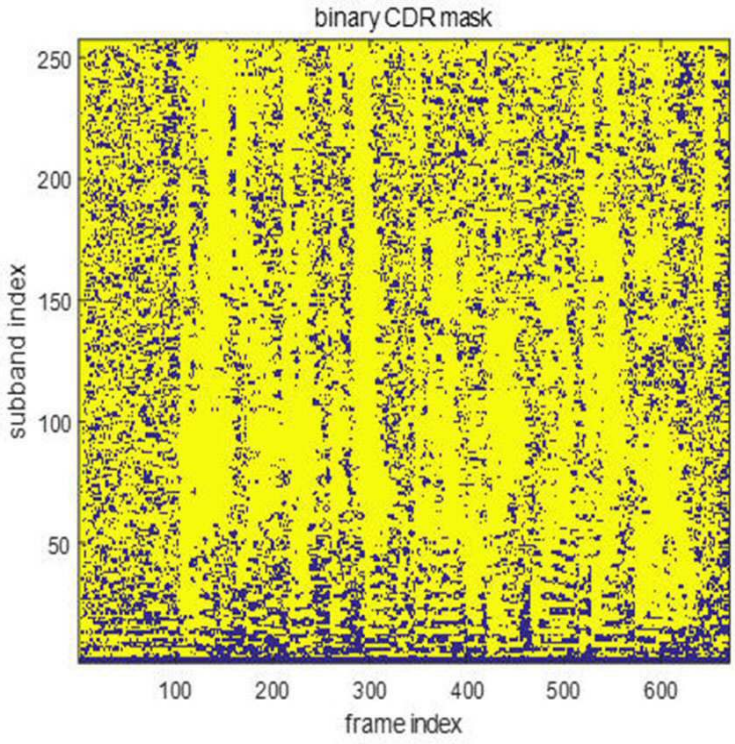
(a)

도면7b


Estimated CDR (=SNR) [dB]
250
200
150
100
50
subband index
100
200
300
400
500
600
frame index

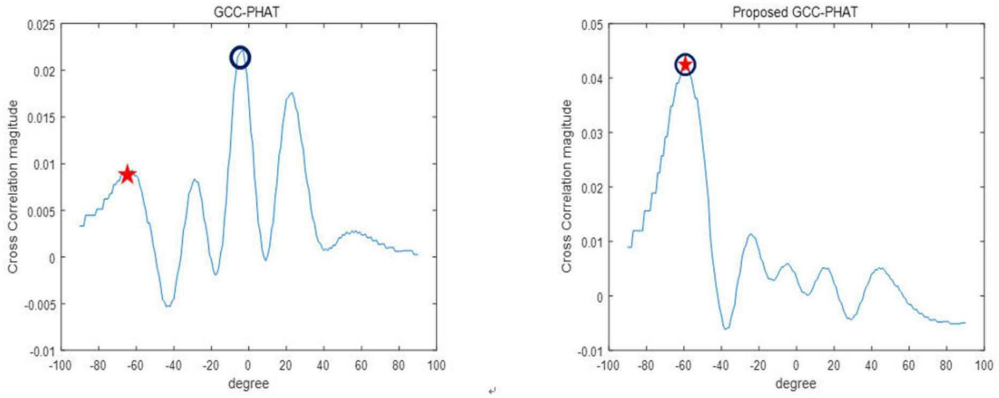
(b)

도면7c


binary CDR mask
subband index
frame index
250
200
150
100
50
100
200
300
400
500
600

(c)

도면8


GCC-PHAT
Cross Correlation magitude
degree
Proposed GCC-PHAT
Cross Correlation magitude
degree

(a) (b)